

# Enforcement Histories and the Violation of Formal Cease-Fire Agreements, 1948–1998

Stephen B. Long  
Kansas State University

sblong@ksu.edu  
[www.ksu.edu/polsci/fac/sblong](http://www.ksu.edu/polsci/fac/sblong) \*

August 10, 2005

## Abstract

When do we expect states to violate formal cease-fires and peace agreements? Does the spectre of tough enforcement prevent states from breaking their cease-fires and peace agreements? I present a theory, consistent with the game theoretic literature on reputation-building, that suggests that states can decrease the probability that their agreements will be violated by investing in a record of tough enforcement. I test these predictions among a field of alternative hypotheses using large-n quantitative analysis on data sets describing formal cease-fires and peace agreements, providing the first quantitative test of predictions of cease-fire failure *and* violations short of failure.

---

\*Data assembly was conducted with the help of EUGene 3.03 (Bennett and Stam, 2000). Thanks to Mark Crescenzi, Tim McKeown, Thomas Oatley, Marco Steenbergen, Georg Vanberg, and Hein Goemans for their advice and comments on this project in its various stages. Thanks also to Page Fortna for providing her data and advice.

# 1 Introduction

This paper examines the factors that drive states to violate their formal cease-fire agreements with each other, focusing specifically on the effect of a state's historical record of enforcement on its cease-fire partners' decisions to violate or honor their agreement.

States have incomplete information about their agreement partners' capabilities and resolve, and yet these are the very traits that may make decisions to violate cease-fires costly or cheap. The cost-benefit calculations that states undertake when deciding to honor or violate their agreements likely considers factors such as the issue or stakes involved, the domestic political benefits or costs of violation, or the likely costs of a response by the target. Still, given these incentives and disincentives, states cannot easily guess the likelihood of a tough response, since this is information that states only know about themselves and have incentives to misrepresent in communication (much like the private information that makes peaceful settlements difficult to reach (Fearon, 1995)). Even the weakest state may be willing to impose punishment that is costly for the violator given sufficient resolve, and even the strongest state may let violations go unpunished given the right circumstances.

So, how can potential violators assess the probability that their agreement partners will impose costly punishment following a violation? The international relations literature suggests that actions, especially costly actions, may be able to signal information about capabilities and resolve in a way that is believable to other states (Fearon, 1994, 1995). States who are deciding whether to violate a cease-fire agreement have every reason to seek reliable information about the probability of a response, since the response directly affects their payoff for violation. I argue that states find this information in other states' historical records of enforcing their cease-fire agreements or failing to enforce those agreements. By observing enforcement and acquiescence over time, potential cease-fire violators can learn about their cease-fire partners, while states wishing to maintain their cease-fires can use enforcement

instrumentally to affect the violation behavior of other states in the future.

I examine the role of two forms of historical information about enforcement. The first is direct, experiential knowledge of enforcement that states gain by being the targets of a specific cease-fire partner's enforcement actions. The second is indirect and reflects the possibility that states learn from what they see happening between others in the international system, specifically how their current cease-fire partner has reaction to violations of cease-fires by all states in the past. This paper examines the theoretical and empirical grounds for believing that states use both forms of historical information to gain information about the likelihood of enforcement and to alter the behavior of potential violators.

Finally, this paper uses the first quantitative data set describing cease-fires and their participants Fortna (2004*b*) not only to explain total failure of these agreements and the subsequent breakdown of the post-war peace, but to explain violations of those cease-fires that fall short of total agreement failure, adding a new dimension to our understanding of behavior surrounding cease-fire agreements. The following sections survey briefly the extant literature on international agreements and the causes of violation, provide a theoretical argument regarding the role of historical information about enforcement in states' decisions to violate cease-fires, and test hypotheses derived from that theoretical framework and existing scholarship using a *logit* analysis and data on cease-fire agreements and militarized interstate disputes.

## 2 Previous Work

Why would states choose to violate an agreement that they found acceptable at the time at which they signed it? Downs, Rocke and Barsoom (1996) suggest that states have been found to be relatively compliant with many agreements because they only sign those that are consistent with their existing interests. Under this logic, if the agreements require real

cooperation that involves doing things they would not otherwise do, states won't sign them. The authors go on to discuss how this result has led scholars to underestimate the importance of enforcement.

There are several factors that may lead states to violate agreements that they once found acceptable, making enforcement a possibility, and perhaps a necessity. Most broadly, as Downs, Rocke and Barsoom suggest, those agreements requiring "deep cooperation," or cooperation that requires states to do what they would not otherwise do, are most likely to have problems with non-compliance (Downs, Rocke and Barsoom, 1996, 383, 397). Within such high-stakes agreements, however, what factors are most likely to lead to violation? Existing work speaking to cease-fires as broader political phenomenon and rather than only on an individual basis is limited to Page Fortna's recent work, but existing scholarship on high-stakes agreements offers several interesting hypotheses that may apply to cease-fire violation as well. The factors that existing work suggests influence decisions to violate agreements can be grouped broadly into the categories of changes after the signing of the agreement, unresolved issues and prior conflict, and agreement composition and design.

## **2.1 Changes After Signing**

### **2.1.1 Power Relationships**

Several scholars have examined the argument, consistent with realist thought, that since agreement reflect the balance of power at the time of signing, changes in the balance of power will lead to violations, renegotiation, or termination of those agreements. Werner (1999), for example, argues and finds support for the idea that changes in relative power after the settlement create incentives to renegotiate that increase the risk that conflict will recur. Measuring changes in relative power in the dyad, she find that as the dyad's balance of power changes from what it was at the time of the peace, the probability of new conflict

increases.

Focusing on another high-stakes agreement type, Leeds argues that alliances are most vulnerable to violation when the expected costs of violation are low or when significant factors have changed since the formation of the alliance (Leeds, 2003, 3). Consistent with several related findings (Bennett, 1997; Siverson and Starr, 1994; Morrow, 1991), Leeds hypothesizes that changes in relative power or the political structure of the member states will influence the probability of alliance failure. Following Werner, she operationalizes power change as any increase or decrease of 10% in the COW capabilities index for a member state, suggesting that such changes increase the probability of alliance failure (1999). Leeds finds support for the hypothesis, suggesting that changes in relative power may powerfully affect the incentives that states have to violate their agreements after they are signed. Even if states expect to benefit from agreements at the time of signing, uncertainty about the future of their power relationships with the other signatories may lead them to violate agreements that were once acceptable to them.

### **2.1.2 Democracy**

The democratic peace literature does not directly speak to the agreement behavior of states, but some scholars have extended the ideas underlying democratic peace scholarship, suggesting that changes in democracy levels may have effects on violation that are not present at the time of signing. Bruce Russett extends Doyle’s “liberal zone of peace” (Doyle, 1986) by suggesting that there may also be a “zone of law.” In this zone, institutionally and ideologically similar legal communities recognize and enforce each other’s law, and mediation and peaceful dispute settlement is more likely (Russett, 1993, 34, 119–120).<sup>1</sup> Anne-Marie Slaughter develops a related argument, suggesting that as democracy spreads, it is more likely that domestic courts, individuals, and groups with a stake in the agreement at is-

---

<sup>1</sup>See also Dixon 1994 and Maoz and Russett 1993.

sue will constrain the governments of liberal agreement partners (Slaughter, 1995, 532–533). Some formal work also supports the proposition that democracies will be more likely to honor their agreements. Fearon, for example, suggests that, “alliance relations between democracies may be less subject to distrust and suspicion if leaders would pay a domestic cost for reneging on the terms of the alliance, so ‘violating the national honor’ in the eyes of domestic critics” (Fearon, 1994, 587). The expected effects of these audience costs on the likelihood of agreement violation occur on the dyadic level, as Fearon expects that “stronger domestic audiences may make democracies better able to commit to courses of action in foreign policy than nondemocracies, features that might help ameliorate the ‘security dilemma’ (Herz, 1950; Jervis, 1978) between democratic states” (Fearon, 1994, 578).

Leeds (1999) provides some empirical evidence supporting the dyadic democratic hypothesis, finding empirical support in the COPDAB data set (Azar, 1993) for the hypothesis that pairs of democracies will be more likely to cooperate with each other, and Gaubatz (1996) employs Doyle’s liberal democracy measure (Doyle, 1983) and the Correlates of War (herein after “COW”) alliance data (Singer and Small, 1966) in a survival analysis that shows that while democracies are no more or less likely to form alliances than other states, alliances between democracies do tend to last substantially longer than mixed or nondemocratic alliances. Other empirical evidence disagrees, however. Gartzke and Gleditsch (2003) argue that democracies will be less reliable alliance partners than non-democracies. Using the occurrence of intervention rather than the duration of the alliance as their outcome of interest, they test the hypothesis that democratic states, subject to electoral cycling and the influence of interest groups, use alliances because they have problems honoring informal agreements, essentially tying themselves together for fear that they would otherwise separate as circumstances change (Gartzke and Gleditsch, 2003, 13–18). The authors find that democracies are in fact less reliable allies than non-democracies.

## 2.2 Unresolved Issues and Prior Conflict

Some issues, when they exist, appear to pose a lasting challenge to cooperation regardless of whether an agreement exists. These forces slowly erode the prospects for continued cooperation in the time that follows the signing of an agreement.

One kind of unresolved issue that can lead to the violation of agreements is indecisive prior conflict, specifically when we are considering post-war agreements. Maoz (1984) compares the “peace by empire” argument (decisive victory destroys a state’s capability and will to renew conflict) first forwarded by Raymond Aron (1966) and the “prudence in victory” argument (symmetrical outcomes minimize grievances about the new status quo) first forwarded by Nissan Oren (1982) in a study of post-dispute stability (Maoz, 1984, 230–232). Maoz finds support for the “peace by empire” argument, suggesting that decisive outcomes lead to better overall post-dispute stability. These findings are supported by Licklider’s (Licklider, 1995, 685) results regarding the resumption of civil wars, and is also consistent with Wagner’s (1993) that decisive victory destroys the defeated state’s organizational structure.

## 2.3 Compliance by Design?

If states sign agreements under one set of circumstances, but are driven by changes that take place after the agreement is signed or by forces created by unresolved issues and prior conflict, why can’t states effectively design agreements to address these challenges? One might argue that states have no incentives to do so, as they might benefit from such changes and wish to violate the agreement themselves. Whether purposeful or not, however, agreement provisions may powerfully affect the incentives that states have to violate agreements after they are signed.

Fortna argues that the provisions of agreements themselves have effects on the durability of cease-fires that are independent from the other factors described above (Fortna, 2004*b*).

Fortna’s results suggest that including certain mechanisms in the terms of cease-fire agreements and implementing those mechanisms strongly improves the prospects for continued peace. Specifically, withdrawing troops beyond pre-war lines, implementing demilitarized zones, and creating joint commissions for dispute resolution increase the duration of peace following wars. In addition, increasing textual “specificity” in the agreements themselves, and implementing permanent peace agreements increase the duration of peace (Fortna, 2004*b*, 176,196).<sup>2</sup> Fortna not only provides the first comprehensive data set of cease-fires, but also demonstrates that the textual content of agreements is not simply the result of contexts already ripe for lasting peace, but has independent effects on the duration of peace separate from these existing conditions.

### 3 A Place for Enforcement

Existing work provides a foundation of knowledge about the forces that drive states to violate high-stakes agreements, but does not follow up on Downs et al’s (1996) suggestion that enforcement should be expected to play a larger role in these agreements than in agreements that do not require deep cooperation. Because of their focus on incentives and disincentives created by unresolved issues, changes in interstate relationships following the signing of an agreement, and the design of the agreement itself, scholars have produced models of violation that expect states to ignore a potentially crucial source of information. This, I argue, could lead to serious errors in empirical prediction.

For example, given changes in relative power, an expectation of highly probable enforcement might prevent even a great and growing power from violating an agreement. A small state that is completely committed to enforcing its agreements might manage to deter the growing agreement partner, despite the incentive that the partner faces to violate the agree-

---

<sup>2</sup>Fortna’s operationalization of the specificity of agreements is the number of paragraphs in the agreement.



ment. In this instance, a prediction based on the various factors described above would be inaccurate because it does not take into account the choices of the potential violator's agreement partner. States that appear to be destined to violate because of a favorable power change and other incentives may not act as expected if they believe that their agreement partner will respond with tough, costly enforcement. States that appear unlikely to violate because of relatively low incentives may act against expectations if they believe that their violation will go completely unpunished.

Consider the case of the 2002 American decision to abrogate the Anti-Ballistic Missile Treaty. While there may have been domestic incentives to begin a full-scale anti-ballistic missile system, the strategic arguments for doing so are dubious, and the costs of potential Russian retaliation high. Yet, despite relatively low incentives to violate the agreement and high potential costs, the United States pushed forward a program that clearly violated the agreement. This would make little sense if a tough Russian reaction was likely, but is not so perplexing if we believe that the expectation of a serious Russian response was low. A low chance of punishment may make even the most minor gains in prestige or security appealing. Leaving information about prior enforcement activity out of our models of agreement violation could lead to inaccurate predictions, essentially omitting what could be a substantively important variable.

## 4 Chain Stores and International Politics

I have argued that existing work on the factors leading states to violate high-stakes agreements of many types provides some insights into the likely causes of cease-fire violation. I have also argued, however, that these models ignore a source of information that both potential violators and enforcers may use in the face of uncertainty.

Uncertainty about states' costs of enforcement, and thus willingness to enforce given a

violation, creates a challenge for both potential violators of cease-fires and their targets. States that hope to maintain their international agreements face a set of multiple potential challengers with incentives to consider challenging their respective agreements with the target state. Challenges by each potential violator can come at any time, and there is little reason for potential violators to believe public statements of resolve regarding the enforcement of the cease-fire, as such statements are likely to be low-cost. Even states wishing to violate the cease-fire themselves have incentives to prevent other states from violating the agreement, if for no other reason than to make a territorial grab or other violation easier and more successful.

States also face challenges when considering a violation of their cease-fires because each potential violator must gauge the likely reaction of its target, given the presence of other incentives and disincentives for violation. A rapid change in the balance of power that favors a state may tempt it to violate the cease-fire, as might some intractable territorial dispute or underlying issue that leads to increasing pressures to violate. However, these incentives must be weighed against the likelihood of a tough enforcement response that is costly the violator.

I argue that, consistent with the general findings of the Chain Store strategic games developed by Kreps and Wilson (1982*a*) and Alt, Calvert, and Humes (1988), potential violators and enforcers can make use of enforcement histories to overcome the informational challenges described above.

Following up on the original game described by Selten (1978), Kreps and Wilson (1982*a*) assume incomplete information about the monopolist's (hereafter "enforcer's") payoffs, and offer a revised model in which the challengers (hereafter "potential violators") revise their assessment of the probability that the enforcer's payoffs reflect short term benefits from being tough (rather than short term losses) (Kreps and Wilson, 1982*a*, 255–256). This model allows for the enforcer's payoffs to be of either type (short-term benefits or losses from being tough),

and the payoff type is known only to the enforcer. Potential violators must begin with an initial assessment of the probability associated with one type and then revise that assessment in their period of play (their chance to violate) based on the enforcer's behavior in all of the previous rounds (256).

The authors begin by assuming that any instance of acquiescence by the enforcer sets the remaining potential violators' assessments of the probability of the monopolist having a strong monopolist's payoffs (those with a short term benefit for fighting market entry) to zero probability. Starting with this assumption and using their concept of sequential equilibrium (Kreps and Wilson, 1982*b*) to solve the game, they find that a truly strong monopolist will always fight market entry. Importantly, however, even a weak monopolist will find it beneficial to suffer the temporary costs of fighting market entry in order to have the chance to get positive payoffs in the future turns unless the challengers' assessment of the probability of it being strong has already been set to zero by prior acquiescence to market entry (260). While other sequential equilibria are possible, these rely on implausible beliefs (with challengers revising their assessment of the probability of strength downward after observing a fighting response) (263). They also expand incomplete information to both sides, so that the monopolist must assess the probability of each challenger being strong or weak in addition to the process described above. Again, the authors again find a role for reputation effects that is consistent with the solution concept of sequential equilibrium. The key insight that their analysis offers is that even in finitely repeated games, it can be worthwhile for players to build reputations, and these reputation effects can be substantial given even small amounts of incomplete information (uncertainty about just one of the players is sufficient) (275).

In their study of hegemonic stability, Alt, Calvert and Humes (1988) modify the chain store game, allowing the monopolist's (the hegemon's, in this case) costs of punishment to vary from case to case (the monopolist is not inherently strong or weak in all rounds). While

the monopolist knows whether the punishment in the current period is costly or cheap, it does not know with certainty whether the punishment will be costly in the next period (although it knows the probability of it being costly on average). With these assumptions, there is some uncertainty in the game for both players. As with the Kreps and Wilson version, the potential market entrants (the allies, in this case) can only estimate the monopolist's probability of having high costs of punishment by observing its actions (450–451).

The decision of the second potential challenger (the potential challenger in the second round of the game) is the decision of the most interest here. Given its observation of the behavior of the first potential challenger and the hegemon in the first round, the second potential challenger chooses whether to challenge the hegemon. Its choice is based on a comparison of the value of the payoff from challenging and the potential challenger's estimate of the probability of punishment by the hegemon. This estimate is based on an initial subjective estimate drawn from a beta distribution, with a Bayesian update given the second potential challenger's count of costly and cheap instances of punishment (instances where punishment would have been costly or cheap, regardless of whether punishments occurred). If the first potential challenger doesn't challenge, the second learns nothing more to revise his/her initial subjective estimate. If the first potential challenger does challenge, what the second learns depends on the hegemon's response. Punishment upwardly revises its estimate of the probability of the hegemon having the option of a cheap (and therefore likely) punishment, acquiescence revises the estimate downward. This is, as mentioned before, compared to the value of challenging for the second potential challenger.

The process described above means that, as with Kreps and Wilson, purposeful action by a player facing multiple sequential challenges in an environment of uncertainty about players' payoffs (both sides for Alt, Calvert and Humes; one side for Kreps and Wilson) can sometimes deter future challenges. Specifically, reputation-building behavior by the hegemon in the first round (given a challenge) that runs the risk of bearing costs for the hegemon can

pay off in the long run by upwardly revising a future challenger's estimate of the probability of punishment. If the potential challenger's estimate outweighs the benefits of challenging, reputation-building can work.

Again, incomplete information allows for reputation-building to have some value to players facing potential challengers even though some of the assumptions of the game differ from those of Kreps and Wilson. Whether incomplete information is assumed to be one-sided or two-sided, these games predict a role for reputation. Reputation-building behavior will not always work, of course, since in some cases the value of challenging will be so high that punishment in the first round will not deter the second round's potential challenger. In other cases, the opportunity for reputation-building never arises because no challenge initially takes place. But, where challenges occur initially, the challenged player should be able to take advantage by punishing and thus increasing a future potential challenger's estimate of the probability of punishment, given various levels of benefits from challenging (453).

Consistent with the general findings of these two Chain Store games, I argue that on the whole, when states invest in reputation-building through enforcement, this should decrease the probability of later challenges because of its effects on the group that can be deterred (states that have incentives for violation neither too high to fear punishment nor too low to consider violation at all). Because of the benefits of maintaining agreements without challenges (regardless of states' own plans for violation), states will be compelled to build up histories of enforcing their cease-fires. When they do build these histories, the overall effect should be to decrease the probability of later potential challengers violating cease-fires with the enforcing state. Those who have extremely strong incentives to violate will do so regardless of the reputation of the cease-fire partner. Those who continue to value the cease-fire will not violate it even if there is no effort to build up a history of enforcement. The middle group, however, should be less likely to risk punishment after observing enforcement behavior with prior challengers. This means that the net effect of enforcement history-

building, when it occurs, should be to decrease probability of violation.

The passage of time after the signing of an agreement can in several ways create incentives to violate, and even though these incentives vary across actors and situations, when states build up histories of cease-fire enforcement, these histories should deter at least some of the potential violators of cease-fires.

## 5 Hypotheses

### 5.1 Direct History of Enforcement

Making the minor assumption that potential challengers do not permanently exit the game after challenging, but rather return as potential challengers in the future, the direct history hypothesis becomes clear. If state A enforces its agreements with a particular challenger (k) over time, k adjusts its estimate of A's likely costs of enforcement to reflect its own direct experiences with A in the past. The logic of this updating process is the same as what is described in the Chain Store literature, but this extension allows states to survive challenges, whereas seeing all potential challengers as new and unique does not. Essentially, this assumption simply allows a single challenger, k, to rejoin the ranks of potential challengers after violating and receiving a response. This is simply a broadening of the definition of what actors constitute the set of potential challengers. The predicted effects of enforcement actions on those potential challengers remain the same as predicted in the original games. The direct history hypothesis, then, is as follows:

- H1: States are less likely to violate formal agreements when the other signatory has a greater record of enforcement against them than when the other signatory has a lesser record of enforcement against them.

## 5.2 Indirect History of Enforcement

The second key hypothesis identifies reputation, not just direct history, as an important factor that affects states' decisions to violate or honor their agreements. While an enforcer, A, may signal low enforcement costs to a challenger, k, by enforcing an agreement violated by k, there are also many other potential k's that observe that interaction and update their initial subjective estimates of A's costs of enforcement for their own agreements of the same type with A. Given the presence of this audience, where A builds up a history of enforcement, it can create a reputation that has effects on potential challengers' decisions. The expected effect is embodied in the following hypothesis:

- H2: States are less likely to violate formal agreements when the other signatory has a greater general record of enforcement of agreements of the same type than when the other signatory has a lesser general record.

## 6 Research Design

This paper uses data collected and made available recently by Page Fortna (Fortna, 2004*b,a*) as a basis for a large-n quantitative analysis that tests the two key hypotheses posed above. Using data on militarized interstate disputes to identify violations and responses to those violations, this paper expands on Fortna's work, which focuses solely on the causes of complete cease-fire failure. By incorporating information about militarized disputes, most of which do not reach the level of war, this paper is able to predict non-war violations of cease-fire agreements that have been previously ignored in quantitative analyses and to identify patterns of violation and enforcement that cannot be discovered in an examination limited to total cease-fire failures. Specifically, I argue below that states that war cannot be said to be enforcement of a formal cease-fire or peace agreement, as war represents an abandonment of

the agreement, not enforcement of it. Focusing on the causes of total cease-fire failure, then, makes the analysis of enforcement actions impossible.

## 6.1 Sample

The sample used in this paper is a modified version of Page Fortna’s time-varying data set that describes all cease-fires and peace agreements following interstate wars ending between 1946 and 1994 (Fortna, 2004*a*, 1) and the participants of those cease-fires and peace agreements. The original data set included 876 observations describing all dyad-years in which both members of the dyad are engaged in a cease-fire. Fortna defines cease-fires as breaks or ends in COW wars with or without an agreement. Each dyad in multilateral cease-fires is counted separately, but Fortna includes principal belligerents only. If wars start and stop repeatedly, each break is included as a cease-fire. When follow-up agreements are signed, these are counted as new cease-fires (Fortna, 2004*b*, 45–48). So, for example, in July 1974, the Turco-Cypriot War stopped with the Geneva Declaration, but restarted two weeks later. After another two days of fighting, the war ended again with a unilateral cease-fire. Fortna codes both of these cease-fires as distinct cases (46). A complete list of cases is available in Fortna’s codebook (Fortna, 2004*a*).

Each observation in the time-varying data begins when the formal cease-fire or peace agreement begins or on January 1 (if it is not the first year of the agreement), and ends on January 1 of the following year or when the formal cease-fire or peace agreement ends through war or other means (Fortna, 2004*a*, 2). In other words, most observations are full years, but if the formal cease-fire or peace agreement begins or ends within the year, the observation can last less than a full year. While the presence of partial years in the data is not ideal, most data that are collected and employed in quantitative studies of international conflict are available in an annual format, and the point in the year at which the data are collected is sometimes unknown. This analysis adopts Fortna’s approach of using such



annual data for control variables, even for partial-year observations, for lack of a feasible alternative. The impact is likely to be limited for many variables that change gradually and would not register major changes on a monthly basis.

I use a modified, directed-dyad observation version of Fortna’s data set that contains observations for only formal cease-fires and peace agreements. Cease-fires and peace agreements are considered formal if there is a formal acceptance of the cease-fire proposal or a formal bilateral or multilateral agreement, which excludes unilaterally declared or tacitly accepted cease-fires (Fortna, 2004*a*, 16). Each dyad is represented twice in the modified data set, with each state taking on the role of State 1 and State 2 one time. This is necessary because the behavior that is being predicted is directed action by one member of the dyad against the other. If the outcome of interest were violation by any party, the original data format would be sufficient. Since the outcome of interest here is violation of a cease-fire by one state against the other, directed-dyad data are necessary. Where necessary, variables in the original data set that depend on the identity of State 1 and State 2 are adjusted to reflect the modified format. The full data set, transformed into directed-dyad observation format and reduced to include only formal cease-fires and peace agreements, contains 1,226 observations.

## **6.2 Key Independent Variables: General and Direct Enforcement Histories**

The two key hypotheses that emerge from the earlier theoretical discussion suggest that a state’s history of enforcing its agreements takes on two forms, a direct, dyad-specific history of enforcement and a general reputation for enforcement. This paper employs data on militarized disputes (Maoz, 1999; Ghosn and Bennett, 2003) to identify states’ responses to violations of their cease-fires to create scores measuring states’ historical record of enforcing

cease-fires with a specific interaction partner and of enforcing cease-fires with all states.

The process of creating these scores begins with identifying, for each directed-dyad observation in the data set (all of which describe dyads engaged in cease-fires), whether a violation occurred in the year.<sup>3</sup> For the directed-dyad observations in which State 1 violated against State 2, I code the combination of State 1's hostility level in the violation and State 2's hostility level in any dispute that it initiated later in the same time period against State 1. This creates seven count variables because each time a violation takes place, one of seven possible scenarios can take place. If the violator violates with a hostility level of four out of five (violence short of war), the target can respond by doing nothing, by initiating a new militarized dispute and reaching a lower level of hostility in it, by initiating a new militarized dispute and reaching the same level of hostility in it as the violator did in the violation, or initiating a new militarized dispute with a higher level and reaching a higher level of hostility (five, or war). If the violator violated with a hostility level of five (war), the target can respond by doing nothing, by initiating a new dispute with a lower level of hostility, or by initiating a war. Each violation by State 1 can be matched to only one response by State 2, so if State 1 violated against State 2 at the start of a time period, and State 2 initiated two new disputes against State 1 later that time period, only one of those disputes is counted as a response. These seven count variables describe State 2's responses in a given time period to violations by S1 in the same year. If no violation occurs, all seven variables are zero.<sup>4</sup>

I also code, for each directed-dyad observation, how State 2 responded to violations by State 1 from the previous time period. For violations by State 1 in the previous time period

---

<sup>3</sup>Note that, as described above, not all observations represent full years, such as when a cease-fire begins or ends in the middle of a year. Throughout, when I refer to "directed-dyad observation," I am referring to the unit of time for each observation for a specific directed pair of states, which is usually a full year, but is sometimes less. When I refer to "time period," I am also speaking about period of a year or less than a year for the directed-dyad observation.

<sup>4</sup>All responses must occur after State 1's violation, must be initiated by State 2, and must have State 2 as an originator (involved on the first day). In observations in which the time period is less than a full year, responses are coded only if they take place in that shorter time period.

that had no response by State 2 in the previous time period, I code the number of times each of the seven violation-response combinations occurs in the current time period. Actions by State 2 that are already coded as responses to violations in the current time period are not counted as responses to violations in the prior time period. In other words, if a violation by State 1 occurred in the last time period that had no response from State 2 in the last time period, any response by State 2 this time period is coded as a response to last time period's violation, until a new violation by State 1 takes place. Once a new violation by State 1 takes place, any response by State 2 is coded as a response to that violation rather than the violation in the prior time period.

So, for each directed-dyad observation, these variables record how, if at all, State 2 responded to violations by State 1 earlier that time period, and also how State 2 responded to previously unanswered violations by State 1 in the prior time period. The measures avoid counting any violation as having two responses or counting any single response as matching two violations.<sup>5</sup> While my approach counts responses by State 2 in which State 2 reaches a hostility level of 4 or 5 as violations as well as responses, this is not problematic because the decision to respond to a violation with force is also a decision to violate the basic premise of that agreement, a prohibition on violence. This technique allows states to respond to each other's response histories both when they violate an agreement first, and when they choose to respond to violations. Each action in a sequence of action-reaction is equally affected by each state's history of prior behavior.

These measures provide the basis for the direct enforcement history score and the general enforcement history score used in my analysis. For any given year, I measure enforcement actions as those in which a state responds to a prior violation (earlier in the time period, or in the prior time period) by initiating a new militarized dispute in which it reaches the same

---

<sup>5</sup>Note that State 2 can be coded as not responding to the same violation twice, once in the year of the violation and once in the following year, but that this is accounted for when the composite enforcement scores are created.

hostility level as did the violator in the earlier violation. So, if State 1 initiates a violation against State 2 in 1959 and reaches a hostility level of four out of five (violence short of war), I would code a response in the same or next time period by State 2 as enforcement if that response is a new militarized dispute initiated by 2 in which State 2 reaches a hostility level of four and is involved on the first day. If State 2 does not respond, or if it responds with a lower-level dispute, these actions are not counted as enforcement, but rather as acquiescence. If it escalates the dispute, this is also not counted as enforcement, nor is any war-level response (even if it responds to a war-level violation).

The logic behind this coding scheme is based on my argument that war cannot be considered enforcement of a cease-fire agreement, nor can a state's immediate response to a militarized dispute. I argue that war cannot be seen as enforcement because the step to war is a crucial one that permanently alters the political landscape, sometimes changing the territorial borders that are the subject of the cease-fire, significantly altering bargaining leverage, and engaging states in a process that is increasingly difficult to end as domestic constituencies become involved (Long, 2003). The decision to go to war shatters all hope of returning to the broken cease-fire as it was and represents an abandonment of the cease-fire, not enforcement of it. When states fail to respond to violations, or when they respond with lower hostility, such responses are signals of acquiescence, not enforcement. It is responses of equal hostility that represent enforcement, imposing costs on the enforcer without necessarily making the continuation of the cease-fire impossible. I focus on new disputes rather than states' responses immediately following a violation (within the same dispute) because I do not want to confuse enforcement and self-defense. Measuring states' responses to violations within the same militarized dispute as enforcement would dramatically increase the number of observed instances of enforcement, but the decision to react immediately to a border incursion, for example, may not reflect the purposeful use of military action for long-term goals that is implicit in the concept of agreement enforcement. The final versions of the key

independent variables used in this analysis consist of two enforcement scores.

The first score is a directed-dyad observation measure of State 2's direct enforcement history with State 1 in the prior time periods in which they are present as a directed-dyad in the data set. For any given directed dyad in a specific time period, the following is calculated

$$S2_t = S2_{t-1} + E_t - A_t \quad (1)$$

where  $S2_t$  is State 2's enforcement behavior at time  $t$  and  $S2_{t-1}$  is State 2's enforcement behavior in the prior time period.  $E_t$  is the number of times in which State 2 responds to a violation in the current or prior time period by initiating a new, non-war MID with a hostility level equal to that of State 1 in the corresponding violation.  $A_t$  is the number of times in which State 2 does not respond to a violation by State 1 in the current or prior time period or responds by initiating a new MID with a lower hostility level than that of State 1 in the corresponding violation.<sup>6</sup> This measure results in a count, for each directed-dyad observation, of State 2's enforcement responses to violations by State 1 in the same time period or the prior time period, reduced by the number of acts of acquiescence in the same period. When there is neither a violation by State 1 nor a response by State 2, the measure retains its value from the prior year.

The second enforcement score measures the enforcement history of State 2 against all other states in the data set. This measure is the result, for any directed-dyad observation, of the sum of  $S2_t$  in all prior time periods in which State 2 has the same identity as in the current observation. This measures State 2's responses to all other State 1's up until, but not including, time  $t$ .<sup>7</sup>

---

<sup>6</sup>The coding scheme used to collect data on State 2's responses to current-period and past-period violations makes it impossible for State 2 to be coded as responding twice to the same violation or responding once to two violations. Over-counting acquiescence is also avoided through steps described in the procedure for creating the scores, available upon request from the author.

<sup>7</sup>These measures were compiled to reflect responses up until and including time  $t$  and were then lagged. The lagging procedure occurs by observation, not actual calendar year, since some observations describe

Both of these scores are normalized by dividing the raw scores by the total number of violations over the same time periods. For the direct enforcement history score, *direct enforcement history*, the raw score is divided by the total number of violations by State 1 against State 2 in all prior cease-fire time periods. For the general enforcement history score, *general enforcement history*, the raw score is divided by the total number of violations by all State 1s against State 2 in the cease-fire time periods prior to time  $t$ . This normalization process ensures that states are not described as frequent enforcers simply because they have more opportunities for enforcement (more violations to which they can respond) because they face persistent violators or because they appear in the data for long periods of time.<sup>8</sup>

### 6.3 Dependent Variable: Violation

Violation of cease-fires is measured here using data describing militarized interstate disputes (Maoz, 1999; Ghosn and Bennett, 2003). The sample employed here includes all directed-dyad observations in which both members of the dyad are participants in a cease-fire with each other. In the case of cease-fires in which more than two states are involved, each possible pairing of participating states is included separately for each time period in which the cease-fire exists. By comparing these cease-fire dyad observations with the record of dyadic militarized disputes, it is possible to identify the events that are most likely to be unambiguous violations of the most basic element of a cease-fire: the prohibition of directed violent action between participants.

---

partial years, as noted above. Thanks to Mark Crescenzi for his advice and assistance in creating these enforcement scores.

<sup>8</sup>While the independent variables described above describe responses to violation, and thus can only be moved in a positive or negative direction when violations have occurred, this is not a standard selection effect. First, the issue relates to effects on the independent variables, not the dependent variable. Second, instead of being unable to observe the independent variables when the dependent variable is zero, the coding of the enforcement scores allows them to distinguish between a non-violation, non-response observation and a violation, non-response outcome. When there is no violation and no response,  $E_t$  and  $A_t$  are both zero, meaning that the observation does not move the enforcement score in one direction or the other. When there is a violation and no enforcement,  $A_t$  is 1, moving the enforcement score in the negative direction.

A cease-fire violation by State 1 against State 2 is coded as having occurred when a militarized dispute occurs between cease-fire partners that meets the following qualifications.<sup>9</sup> First, the MID must have been initiated by State 1.<sup>10</sup> Second, the MID must have started during the time period of the cease-fire.<sup>11</sup> Third, State 1 must be an originator in the MID (involved on the first day). Finally, State 1's hostility level must have reached a level of 4 (use of force) or 5 (war). This measure does not record decisions by states to join disputes after they begin as violations and ignores states that are technically on the side of the initiator, but that do not have direct interactions with the target themselves.

Focusing on the initiators of violent disputes has positive and negative consequences. One negative consequence of using data on militarized disputes, specifically those that are violent in nature, is that not all actions that violate the letter of cease-fire agreements are violent in nature. Moving troops into a demilitarized zone, for example, may not result directly in violence, but may still be a violation of the specific terms of a cease-fire agreement. This

---

<sup>9</sup>EUGene 3.04 (Bennett and Stam, 2000) was used to generate the militarized disputes data used here. Maoz's dyadic disputes data were used for cases through 1991, while the COW dyadic MIDs data were used for cases in 1992 and after (while Maoz codes cases from 1992, EUGene allows COW-MID coding for those cases to override Maoz's coding). Because EUGene is capable of reporting values for only one militarized dispute per year, I relied on the Maoz and COW-MID data to code the additional cases in years in which multiple MIDs occurred.

<sup>10</sup>In some rare cases, dispute data in the the original Maoz and COW sources appear to disagree with the EUGene-generated output. Where such disagreements appear, I accept EUGene's coding practices, as it is a frequently-updated and widely accepted source for commonly used conflict data sets.

<sup>11</sup>While Fortna identifies which cease-fires ended in war and the date of that war, she relies on COW war dates, then makes some adjustments based on her own research. These decisions lead to some differences between the dates on which Fortna records cease-fires ending in war and the dates on which the MID data record wars. In some cases, COW wars and MID wars do not appear to match, which could be due to differences in coding participants, start dates, or hostility levels of the participants. More often, Fortna codes a cease-fire as occurring in the midst of what the MID data see as a continuing dispute. When Fortna codes the cease-fire as ending, she also codes the start of a new war, while the MID data do not code the dispute as ever having stopped (so no new war is coded). In order to remain consistent with my use of MID dates for non-war disputes throughout the rest of the data, I use the MID dates for wars, as well, even though there are some likely errors that result from this choice. Specifically, some militarized disputes may be coded as violations or not coded as violations because a small discrepancy in dates makes them appear to be within or outside of a cease-fire period. Generally, the errors appear to under-count, rather than over-count violations, and a rough tally identifies a net loss of 11 likely violations. The effect of this loss may be amplified through the procedure that creates the enforcement scores. Since many of these date discrepancies occur due to Fortna's adjustments for Arab-Israeli wars and cease-fires, future case work may be able to remedy this problem.

means that some valuable information is being lost. The most obvious alternative would be a measure that codes violations based on an interpretation of the specific obligations of each peace agreement. This approach would do a better job of capturing non-violent violations of cease-fires, but may reduce the sample size because of its need for the text of the agreement. Coding based on the individual terms of the agreements may also create new opportunities for coding errors and make replication more difficult. Using a well-documented and widely-used data source on militarized disputes, on the other hand, minimizes false-positives by adopting a stricter interpretation of cease-fire obligations and makes the measure more transparent and more easily replicated than a manually coded measure of violation. Another negative consequence of using violent disputes to identify cease-fire violations is that some of the initiators of these disputes are actually responding to a lower-level incident initiated by the other side. These true first incidents may not reach a hostility level sufficient to count as a militarized dispute, so in some instances, responders are being coded as initiators because of data collection techniques. While this is not ideal, the actions that are being coded as violations still qualify as such, and there is little reason to believe that the theoretical framework provided here would apply any differently to these violations than to violations that are both violent and first chronologically.

## **6.4 Secondary Independent Variables**

Testing the key hypotheses in isolation runs the risk of either failing to identify causal relationships that speak to the validity of the hypotheses or identifying apparent, but spurious causal relationships. There are several hypotheses that can be derived from the existing literature that are likely to help explain states' decisions to violate or honor their cease-fires with other states. Fortna summarizes these in her model of the "baseline prospects for peace", which she uses as a foundation for tests of the independent effects of agreement provisions on the durability of peace (Fortna, 2004*b*, 76). I apply a subset of her controls



Table 1: Secondary Independent Variables

Variable	Operationalization	Changes
<i>War Ended in Tie</i>	1 if prior war ended in tie	Dir-Dyad Format
<i>Costs of Prior War</i>	S1's deaths in war	S1 instead of total
<i>Multiple States in Prior War</i>	1 if war was multilateral	None
<i>History of Disputes</i>	prewar MIDs/years S1&S2 in system	None
<i>Issue is Territory</i>	1 if MID revision type=territorial	None
<i>Existence at Stake</i>	1 if ICB value=existence	None
<i>Contiguous</i>	1 if states contiguous	None
<i>Monadic Democracy</i>	S1 POLITY score	Continuous
<i>Both States Democratic</i>	1 if S1 and S2 scores >5	None
<i>Change in Relative Power</i>	1-year change in S1-S2 CINC ratio	IDs beneficiary
<i>Balance of Power at Cease-Fire</i>	$ cap1 - cap2 /(cap1 + cap2)$	None
<i>Agreement Strength</i>	Fortna's <i>index</i>	None

here with the same purpose: to control for the effects of a representative group of variables from the existing literature so as to be able to more accurately identify the independent effects of direct histories of enforcement and general reputations for enforcement. The table below describes briefly each of the controls used here, referring the reader to Fortna's data manual for details on the sources of the data employed in the creation of the controls (Fortna, 2004a). I also describe any necessary adjustments to Fortna's variables resulting from the use of directed-dyad data in this study. Taking the two key hypotheses from above as the first and second of interest, the control hypotheses below begin with the third to be tested.

Two general hypotheses offered by Fortna are omitted here. The first suggests that cease-fires and peace agreements between states with histories of conflict are more likely to break down. As a robustness check, I provide results that include and exclude this variable, but it is not a part of the primary model because it shares common conceptual ground with the key independent variable *direct enforcement history*, which also describes prior conflict behavior in a dyad. The second excluded hypothesis is that the expected utility of war Bueno de Mesquita and Lalman (1992) affects the likelihood that conflict will recur. I exclude this

here because the predicted outcome is clearly dyadic and speaks to war specifically, but has unclear implications for violation.

## 6.5 Model Choice

The statistical model employed in this paper is *logit*. While I collected data describing the number of total violations by State 1 against State 2 for each observation, the resulting variable has very few observations with values other than zero or one, as shown in Table 1 below. This makes it inappropriate for use in an Ordinary Least Squares regression because OLS assumes a continuous dependent variable. Coding the same data in dummy format makes more sense statistically, while retaining its connection with the key concept of interest, states' decisions to violate their formal cease-fires or peace agreements. Given the format of the dependent variable, *logit* is the appropriate statistical model.<sup>12</sup> The results presented below also adjust for the effects of likely clustering in the data, specifically by conflict. Since there is likely to be a connection between observations describing the same pair of states in the same conflict period, the assumption of independence of observations is violated in these data, making it necessary to use adjusted standard errors that account for this characteristic of the data.

## 7 Analysis

Table 2 presents the results of three models predicting State 1's choice to violate its formal cease-fire or peace agreement with State 2 in a directed-dyad observation format (each state in a dyad is represented as State 1 and State 2). Model 1, shown in the first column of results, includes the direct enforcement history variable and the set of controls described

---

<sup>12</sup>Note that while the dummy variable collapses information from multiple violations in a single observation into a dichotomous measure, the enforcement scores still contain information from the handful of violations lost in the dummy variable.

above. Model 2, shown in the second column of results, includes the general enforcement history variable and the set of controls described above. Model 3, shown in the third column of results, includes both enforcement history scores and the controls. The key models of interest here are Model 1 and Model 2, and the discussion of sign, significance, and effects on predicted probabilities below focus on these models. This is because the construction of the enforcement scores is such that the general enforcement score, as part of its calculations of State 2's history of responding to all violations regardless of the identity of the challenger, includes information about past violations and responses for the current State 1. While the two enforcement scores are not perfectly collinear, the construction of the variables and the effects of including both in the analysis make it clear that collinearity is an issue. It may be possible to create an alternate version of the general enforcement history score that ignores prior interactions with the current State 1, but conceptually, there is still likely to be an issue of collinearity, as a pattern of enforcement by State 2 against states in the system in general is still likely to be reflected in the direct enforcement history of State 2 with each individual State 1. This does not mean, however, that useful results cannot be derived from Model 1 and 2.

## **7.1 Model 1: Direct Enforcement History**

Looking first at sign and significance, Model 1 yields several interesting results that speak to the hypotheses. The key independent variable in Model 1, the direct enforcement history of State 2, appears to have the effect predicted by the theory described above. Specifically, the stronger a record of enforcement maintained by State 2 up to time  $t - 1$ , the less likely is State 1 to violate its formal cease-fire or peace agreement with State 2 at time  $t$ . The effect is highly statistically significant and does not appear to be driven by Arab-Israeli dyads or

Table 2: Logit Model of Formal Cease-Fire and Peace Agreement Violation

<b>Variable</b>	<b>1</b>	<b>2</b>	<b>3</b>
	Direct Only	General Only	Direct and General
<i>Direct Enforcement History</i>	-0.78*** (0.30)		-0.78 (0.67)
<i>General Enforcement History</i>		-0.61* (0.34)	-0.02 (0.82)
<i>War Ended in Tie</i>	1.01*** (0.29)	0.92*** (0.34)	1.02*** (0.27)
<i>Cost of Prior War</i>	0.01* (0.01)	0.01** (0.01)	0.01 (0.01)
<i>Multiple States in Prior War</i>	-1.10** (0.53)	-1.12** (0.54)	-1.09** (0.47)
<i>History of Disputes</i>	1.40*** (1.31)	1.22*** (0.19)	1.41*** (0.17)
<i>Issue is Territory</i>	1.00* (0.60)	1.13 (0.73)	1.02* (0.60)
<i>Existence at Stake</i>	0.60*** (0.23)	0.61*** (0.23)	0.58** (0.27)
<i>Contiguous</i>	-0.72 (0.58)	-0.46 (0.68)	-0.72 (0.50)
<i>Balance of Power at CF</i>	0.57** (0.25)	0.48* (0.25)	0.56*** (0.20)
<i>Change in Relative Power</i>	-0.05 (0.07)	-0.06 (0.07)	-0.05 (0.06)
<i>Monadic Democracy</i>	-0.05*** (0.01)	-0.04*** (0.01)	-0.05*** (0.01)
<i>Both States Democratic</i>	0.73* (0.40)	0.69 (0.51)	0.71* (0.39)
<i>Agreement Strength</i>	0.04 (0.15)	0.03 (0.13)	0.03 (0.14)
N	938	956	936
Log likelihood	-281.4	-292.9	-280.8

Standard errors adjusted for clustering by conflict in ( ).

\*\*\*=significant at the .01 level, \*\*=.05, \*=.1.

by the presence of multiple dyads representing single cease-fire agreements.<sup>13</sup>

The control variables yield interesting results, as well. The hypothesis that State 1 is more likely to violate its formal cease-fires with State 2 when the war preceding the cease-fire resulted in an indecisive outcome is supported by the statistical evidence. The variable representing this hypothesis, *war ended in tie* is positive and highly statistically significant, meaning that states whose prior war ended in a tie are more likely to violate a formal cease-fire or peace agreement with each other afterwards. The hypothesis that states with histories of militarized disputes will be more likely to violate their formal cease-fires or peace agreements, is also strongly supported in the results.<sup>14</sup> The evidence suggests that when one of the states' existence is at stake, states are more likely to violate their formal cease-fires or peace agreements with each other, and the more democratic a state is, the less likely it is to violate against other states regardless of their regime type. Using an alternate, ordinal measure of the stakes that ranks the stakes from political influence to existence, the sign

---

<sup>13</sup>Including a control dummy indicating cases in which Israel is an actor leads to only a minor decrease in *direct enforcement history*'s significance ( $p=0.025$ ). Some of the control variables exhibit greater sensitivity to the inclusion of an Israel control. *War ended in tie*, *cost of prior war*, and *multiple states in prior war* become insignificant, while *issue is territory* ( $p=0.009$ ), *contiguity* ( $p=0.048$ ), and *both states democratic* ( $p=0.029$ ) improve in traditional levels of significance. The remaining six control variables retain their signs and traditional levels of significance. Dropping all but the cases for one dyad in multiple dyad cease-fires only slightly affects *direct enforcement history*'s significance ( $p=0.20$ ), while the controls are more sensitive. *Cost of prior war* and *issue is territory* become insignificant, while *change in relative power* changes sign, and *multiple states in prior war* ( $p=0.000$ ) improves in its traditional level of significance and *balance of power at cf* changes signs and loses in its traditional level of significance. Some of these changes are likely to be the result of *contiguity* being dropped in the robustness check because it perfectly predicts failure in the analysis, but the controls are clearly more sensitive to dropping cases for all but one dyad in multiple dyad cease-fires.

<sup>14</sup>The history of militarized disputes variable very powerfully affects the results for the other control variables. Dropping the measure from the analysis makes the other controls in almost all cases insignificant (with the occasional exception of monadic democracy, depending on control variable measurement), although the key independent variables in Models 1 and 2 retain their sign and significance levels. This is a surprising result, since dispute history shares some common conceptual ground, although it is not highly correlated, with the enforcement scores and one would expect it to affect them more than the controls. The dispute history, however, does include militarized disputes not accounted for in the enforcement scores (MIDs of lower hostility). Also, my argument for measuring violation and enforcement in the way that I do also hinges on the assertion that there is a difference between a MID initiated with or without a formal cease-fire in place at any hostility level and a MID initiated at a violent hostility level during a formal cease-fire. For these empirical and theoretical reasons, I keep the conflict history in Models 1, 2 and 3.

and traditional significance level of the stakes variable remains the same.<sup>15</sup> The evidence also suggests that balanced power at the time of the cease-fire decreases the chances of violation, so regardless of the effects of changes in power on the probability of violation, the configuration of power at the start of the cease-fire appears to have independent effects.

There is marginal statistical support for the hypothesis that states are more likely to violate cease-fires following high-cost wars. This suggests that the underlying issues indicated by a high-cost conflict lead states to violate their cease-fires in such cases rather than being deterred by the losses that they experienced in the prior war. Measuring this concept with the natural log of total dyadic deaths in the prior war, total deaths for all participants in the prior war, or duration of the prior war yields results that generally support this finding.<sup>16</sup> The hypothesis that, when the issue under dispute is territory, states are more likely to violate formal cease-fires or peace agreements with each other is also marginally supported, a finding that reflects the consensus in the field regarding the role of territory in international conflict.

No support is found in the results the idea that contiguity increases the probability that a state will violate, or for the idea that a favorable change in the balance of power increases the probability that a state will violate. While Werner's hypothesis about changes in the balance of power has been applied to total cease-fire failure and to alliance reliability, changes in the balance of power that are favorable to a state do not appear to lead the state to violate or

---

<sup>15</sup>Using the ordinal measure of stakes results in the key independent variable's traditional level of significance move from 0.01 to 0.05, but it is extremely close to the 0.01 level ( $p=0.11$ ), and it improves the traditional significance level of territory ( $p=0.023$ ) and joint democracy ( $p=0.006$ ). The costs of war variable ( $p=0.524$ ) becomes insignificant.

<sup>16</sup>The natural log of dyadic deaths and the total deaths for all participants measures yield results of the same sign and better statistical significance, while duration is of the same sign but insignificant. Using these alternate measures leaves most of the other controls' and the key independent variable's sign and traditional level of significance unaffected, with the exception of some minor changes. Using the natural log of dyadic war deaths improves the traditional level of significance of territory ( $p=0.03$ ) and joint democracy ( $p=0.025$ ) and worsens the traditional level of significance of the tied war outcome variable ( $p=0.024$ ). Using the total deaths for all participants in the prior war has the same effect on the traditional levels of significance of territory ( $p=0.047$ ) and joint democracy ( $p=0.036$ ). Using war duration makes the territory variable insignificant ( $p=0.113$ ). On balance, these are minor changes in a small set of control variables.

totally abandon its cease-fire with the weakened opponent after controlling for the effects of that state's enforcement history against the potential violator and the other factors described above. Surprisingly, Fortna's own key hypothesis that the strength of agreement terms has an independent effect on the probability of total cease-fire failure does not seem to help explain violations more broadly (including those not leading to total failure). The results for agreement strength remain insignificant when we consider violations short of war (MID hostility level of 4) alone as the dependent variable.

Surprisingly, there is evidence the idea that State 1 is less, not more, likely to violate a formal cease-fire or peace agreement against State 2 when the prior war involved more than two actors. Since the violation and enforcement interactions described in the data are described on the directed-dyad observation level, the collective action dynamics of multilateral cease-fires may be obscured somewhat by the research design. Based on the existing data, however, agreements ending multilateral wars appear to be less, not more likely to be violated by any member of one side against any member of the other side (given that they engaged each other directly). There is also marginal evidence for the opposite of the argument that two democracies are less likely to violate against each other. There are very few cases in which two states, after signing a formal cease-fire or peace agreement, are eventually both democratic during the period of the agreement, and Fortna questions the coding of one joint democracy, noting that her results hinge on that single pair of states (Fortna, 2004*b*, 110-111). Given the nature of the variable in this data set, it is difficult to come to a confident conclusion about the effects of joint democracy.

Discussing sign and significance sheds light on which predictors are most helpful in predicting variance in the dependent variable, but *logit* coefficients do not lend themselves to direct interpretation as do OLS coefficients. One useful technique is the generation of first differences, which makes it possible to evaluate the comparative impact of individual variables on the probability of the event being predicted (violation), holding all other covariates

Table 3: Model 1 Effects on the Predicted Probability of Violation

Covariate	Min	Predicted p	Max	Predicted p	% $\Delta$
<i>Direct Enforcement History</i>	-1	0.28777	1	0.078346	-20.9%
<i>War Ended in Tie</i>	0	0.081579	1	0.195307	11.4%
<i>Multiple States in Prior War</i>	0	0.422654	1	0.195307	-22.7%
<i>History of Disputes</i>	0	0.075607	2.8696	0.817971	74.2%
<i>Existence at Stake</i>	0	0.117431	1	0.195307	7.8%
<i>Balance of Power at CF</i>	0.0328	0.15042	0.9762	0.232184	8.2%
<i>Monadic Democracy</i>	-10	0.283186	10	0.131849	-15.1%

Variables shown are those with significance at the 0.05 level or better. Predicted probabilities calculated holding continuous variables at their means and dummy variables at 1.

constant. Table 3 shows the percentage change in the predicted probability of violation for each of the variables that qualify as significant by the 0.05 traditional level.

The history of disputes between states has the strongest effect on the probability of violation, such that State 1 is 74.2% more likely to violate a formal cease-fire or peace agreement against State 2 if they have an average of nearly three disputes per year in which they are both members of the international system than if they have no disputes in their history. This is not surprising, as it has long been argued that states with histories of conflict interact in different ways than those with no such history. Rivals are more likely to have wars in the future, more likely to have longer wars, and more likely to escalate conflicts to wars. It does not come as a shock that they are also more likely to violate their cease-fires against one another. Somewhat more surprising is the finding that State 1 is 22.7% less likely to violate a formal cease-fire or peace agreement against State 2 when the prior war involved more than two participants than when it did not.

The results also suggest that State 1 is nearly 21% less likely to violate a formal cease-fire or peace agreement against State 2 when State 2 has a perfect history of responding to violations with enforcement than when it has a perfect record of acquiescence. Given that this effect is strong even when controlling for the large set of alternative explanations



described above, this finding suggests that State 2's direct history of enforcement against State 1 powerfully affects State 1's decision to violate against State 2 in the future. Building up a direct history of enforcement pays off for the enforcers, making challenges to the peace less likely. This effect is highly significant and stronger than the effects of the war outcome, the balance of power at the time of the cease-fire, the stakes, or monadic democracy.

## 7.2 Model 2: General Enforcement History

In Model 1, the variable measuring State 2's direct history of enforcement against State 1 proved to be a highly significant predictor of violation by State 1, reducing the probability of violation as the enforcement history grows, as predicted by the Chain Store argument. Model 2 offers weaker support for the hypothesis that State 2's general reputation for being a tough enforcer of its formal cease-fires and peace agreements. Model 2 shows that when State 2 builds up a reputation for being a tough enforcer of its formal cease-fires and peace agreements, it can reduce the chances of violation by State 1, but the effect is at a marginal level of significance ( $p=0.076$ ). It becomes insignificant ( $p=0.120$ ) when I include a dummy variable representing cases involving Israel or when I drop all but one dyad in multiple-dyad cease-fires.<sup>17</sup>

Most of the control variables mirror the results in Model 1, with some minor exceptions. States are more likely to violate their formal cease-fires and peace agreements if the prior war was a tie, if there is a greater history of conflict between the states, or if one state's

---

<sup>17</sup>In general, Model 2 is more sensitive to the inclusion of an Israel dummy than Model 1. All except *agreement strength* retain their signs, but all but four (*history of disputes*, *existence at stake*, *change in relative power*, and *both states democratic* change in their traditional levels of statistical significance. *War ended in tie* and *multiple states in prior war* become insignificant, while *cost of prior war* ( $p=0.078$ ) and *monadic democracy* ( $p=0.06$ ) lose in levels of traditional significance and *issue is territory* ( $p=0.006$ ), *contiguity* ( $p=0.086$ ), and *balance of power at cf* ( $p=0.005$ ) gain in traditional levels of significance. When I drop cases for all but one dyad in multiple dyad cease-fires, *issue is territory*, *balance of power at cf*, *change in relative power*, and *agreement strength* change signs while retaining their traditional levels of significance. *Cost of prior war* becomes insignificant, while *multiple states in prior war* ( $p=0.000$ ) and *both states democratic* ( $p=0.042$ ) improve in their traditional levels of significance. As in Model 1, this robustness check drops *contiguous*.

existence is at stake,<sup>18</sup> all with the same traditional levels of significance as in Model 1. States are less likely to violate their formal cease-fires and peace agreements if the prior war was multilateral or if they are more democratic (regardless of the other state's regime type), both with the same traditional levels of significance as in Model 1. Contiguity, changes in the balance of power favoring State 1, and agreement strength all remain insignificant as predictors of violation by State 1.

The costs of the prior war variable ( $p=0.034$ ) improves in its traditional level of significance,<sup>19</sup> but the balance of power at the time of the cease-fire loses in its traditional level of significance, while territory and joint democracy both become insignificant. On balance, State 2's direct enforcement history seems to be a better predictor of violation than State 2's general enforcement history, but both still have the effect expected by the key hypotheses. Both become insignificant when included simultaneously, but this is not surprising given the collinearity issue described above, nor does it negate the results of Models 1 and 2. Among the controls, the most consistently and highly significant predictors of violation across all three models are whether the prior war ended in a tie, whether the prior war involved multiple participants, how great a history of conflict exists in the dyad, whether one state's existence is at stake, and the monadic level of democracy.

The examination of first differences for Model 2 yields several interesting results. As in

---

<sup>18</sup>Using the ordinal measure of stakes described above leads to territory ( $p=0.033$ ), balance of power at time of cease-fire ( $p=0.026$ ), and joint democracy ( $p=0.045$ ) gaining in their traditional levels of significance, while the costs of war variable ( $p=0.226$ ) becomes insignificant.

<sup>19</sup>Compared to the results presented in Model 2, using the natural log of dyadic deaths in the prior war leaves many of the controls with the same signs and traditional levels of significance, but there are also several changes. General enforcement history ( $p=0.038$ ), costs of war ( $p=0.001$ ), balance of power at cease-fire ( $p=0.028$ ), joint democracy ( $p=0.070$ ), and territory ( $p=0.032$ ) all gain in traditional levels of significance, while tied prior war (0.135) becomes insignificant. Agreement strength switches signs but stays insignificant. Using total deaths of all participants in the prior war also leaves the majority of controls with the same signs and levels of significance, but general enforcement ( $p=0.035$ ), costs of war ( $p=0.000$ ), territory ( $p=0.059$ ), and balance of power at time of cease-fire ( $p=0.030$ ) gain in traditional levels of significance, while tied prior war ( $p=0.049$ ) loses in traditional levels of significance. Using war duration results in general reputation ( $p=0.036$ ) gaining in its traditional level of significance and multilateral prior war, existence at stake, and balance of power at time of cease-fire losing in traditional levels of significance. Duration itself as a measure of the costs of the prior war is insignificant.

Table 4: Model 2 Effects on the Predicted Probability of Violation

Covariate	Min	Predicted p	Max	Predicted p	%Δ
<i>General Enforcement History</i>	-1	0.275881	1	0.101016	-17.5%
<i>War Ended in Tie</i>	0	0.098049	1	0.215076	11.7%
<i>Costs of Prior War</i>	22	0.19392	750000	0.396572	20.3%
<i>Multiple States in Prior War</i>	0	0.65931	1	0.215076	-44.4%
<i>History of Disputes</i>	0	0.095437	2.8696	0.780104	68.5%
<i>Existence at Stake</i>	0	0.129285	1	0.215076	8.6%
<i>Monadic Democracy</i>	-10	0.282362	10	0.162088	12%

Variables shown are those with significance at the 0.05 level or better, except general enforcement history. Predicted probabilities calculated holding continuous variables at their means and dummy variables at 1.

Model 1, the history of disputes between State 1 and State 2 has the greatest impact on the probability of violation by State 1. State 1 is 68.5% more likely to violate its formal cease-fire or peace agreement with State 2 if they have an average of nearly three disputes per year in which they are both members of the international system than if they have no history of disputes. State 1 is 44.4% less likely to violate against State 2 if their prior war involved more than two actors, and it is 20.3% more likely to violate if the prior war involved three quarters of a million deaths than if it involved nearly zero (22). This last result is a bit surprising, but speaks to the weakness of war costs as a predictor of violation. If a formal cease-fire or peace agreement ending a war involving so many casualties is only 20.3% more likely to be violated by one state against the other, war costs, or at least that for which they are a proxy other than the stakes and issues at risk, must not play a very large role in the calculus behind decisions to violate.

State 1's decision to violate against State 2 is slightly less affected by State 2's general history of enforcing its agreements against all violators than it is by State 2's direct history with State 1 alone. When State 2 has a perfect record of enforcement against all violators, State 1 is only 17.5% less likely to violate against State 2 than if State 2 has a perfect record

of acquiescing to violators. Since the empirical measure of general enforcement includes the information from the State 1-State 2 dyad, it appears that including information about other dyads somehow waters down the visible effect of historical enforcement information on State 1's decision to violate. While it may be possible to sort this out by creating a measure of State 2's history with all states except State 1, there is still likely to be some collinearity (see discussion above). The lesson that can be taken from this as it stands is that the lesson that State 1 learns from State 2's past enforcement actions against it is clearly one that makes deterrence possible, while the lesson that it takes from State 2's history with it and with all others, when combined as one concept, still makes deterrence possible, but is less robust and may still be driven by direct experience. Separating these measures will provide an interesting path for future work.

## 8 Implications

This paper demonstrates that states that build up records of enforcing their formal cease-fires and peace agreements by matching violations with responses of equal hostility are able to deter other states from violating in the future. When states have a history of enforcing these types of agreements with other states, the targets of their actions are most likely to be deterred, but other states that observe those interactions may also learn that the gamble of violation may not pay off. Even though enforcement is costly in the short term, the long-term dividends may make such activity worthwhile.

Ideally, international agreements would be designed so as to be self-enforcing, but the evidence shows that post-war agreements, no matter how strong the textual provisions, face major challenges. Agreements may be acceptable to all sides when they are signed, but changes in the context and the relentless pressures that underlying issues, high stakes, and long histories of conflict place on states can lead to incentives to violate. While states

cannot alter the outcomes of wars or change the stakes so as to improve cease-fire and peace agreement reliability, they can undertake purposeful actions to do so. Specifically, maintaining a record of enforcement appears to be one of the ways in which these forces can be mitigated effectively.

Surprisingly, factors that appear to affect the probability that cease-fires and peace agreements will fail completely may or may not affect the probability that a given state will violate, but not necessarily abandon, these agreements. Agreement strength, for example, appears to delay complete failure, but has no observable effect on violations more broadly. While aspects of the prior war significantly affect both outcomes, changes in relative power, perhaps one of the most intuitively plausible hypotheses, predicts failure, but not violation more broadly. One explanation for this could be that states who have gained an advantage in power do not bother with violations, but rather move directly to complete abrogation of the agreement. While violation more broadly defined might be the more useful path to renegotiation of terms, states may skip this step in favor of the opportunity to impose wholly new terms on a freshly defeated opponent.

While states face many incentives to violate their agreements, it is fruitful to examine violations that do not lead to the end of cease-fires and peace agreements as well as those that do. When we view violation from this perspective, we learn that not all predictors of total failure fare so well, and more importantly, we can examine the effects of violation-enforcement patterns that cannot be observed if we look at total failure alone.

## References

- Alt, James E., Randall L. Calvert and Brian D. Humes. 1988. "Reputation and Hegemonic Stability: A Game-Theoretic Analysis." *American Political Science Review* 82(2):445–466.
- Aron, Raymond. 1966. *Peace and War*. Garden City: Doubleday.

- Azar, Edward E. 1993. *Conflict and Peace Data Bank (COPDAB), 1948–82. Third Release. Computer File*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research. ICPSR 7767.
- 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 Stam. 2000. “EUGene: A Conceptual Manual.” *International Interactions* 26:179–204.
- Bueno de Mesquita, Bruce and David Lalman. 1992. *War and Reason : Domestic and International Imperatives*. New Haven: Yale University Press.
- Dixon, William. 1994. “Democracy and the Peaceful Settlement of International Conflict.” *American Political Science Review* 88:14–32.
- Downs, George W., David M. Rocke and Peter N. Barsoom. 1996. “Is the Good News about Compliance Good News about Cooperation?” *International Organization* 50(3):379–406.
- Doyle, Michael W. 1983. “Kant, Liberal Legacies, and Foreign Affairs. Parts 1 and 2.” *Philosophy and Public Affairs* 12:205–235, 323–353.
- Doyle, Michael W. 1986. “Liberalism and World Politics.” *American Political Science Review* 80(4):1151–1169.
- Fearon, James D. 1994. “Domestic Political Audiences and the Escalation of International Disputes.” *American Political Science Review* 88(3):57–592.
- Fearon, James D. 1995. “Rationalist Explanations for War.” *International Organization* 49(3):379–414.
- Fortna, Victoria Page. 2004a. “Data Notes: The Cease-Fires Data Set.”.

- Fortna, Victoria Page. 2004b. *Peace Time: Cease-Fire Agreements and the Durability of Peace*. Princeton, New Jersey: Princeton University Press.
- Gartzke, Erik and Kristian Skrede Gleditsch. 2003. "Regime Type and Commitment: Why Democracies are Actually Less Reliable Allies." manuscript.
- Gaubatz, Kurt Taylor. 1996. "Democratic States and Commitment in International Relations." *International Organization* 50(1):109–139.
- Ghosn, Faten and Scott Bennett. 2003. "Codebook for the Dyadic Militarized Interstate Incident Data, Version 3.0." Internet.
- Herz, John. 1950. "Idealist Institutionalism and the Security Dilemma." *World Politics* 2(2):157–180.
- Jervis, Robert. 1978. "Cooperation under the Security Dilemma." *World Politics* 30:167–215.
- Kreps, David and Robert Wilson. 1982a. "Reputation and Imperfect Information." *Journal of Economic Theory* 27:253–279.
- Kreps, David and Robert Wilson. 1982b. "Sequential Equilibria." *Econometrica* 50.
- Leeds, Brett Ashley. 1999. "Domestic Political Institutions, Credible Commitments, and International Cooperation." *American Journal of Political Science* 43(4):972–1002.
- Leeds, Brett Ashley. 2003. "Alliance Reliability in Times of War: Explaining State Decisions to Violate Treaties." *International Organization* 57:801–827.
- Licklider, Roy. 1995. "The Consequences of Negotiated Settlements in Civil Wars, 1945–1993." *American Political Science Review* 89(3):681–690.
- Long, Stephen B. 2003. "Time Present and Time Past: Rivalry and the Duration of Interstate Wars, 1846–1985." *International Interactions* 29(3):215–236.

- Maoz, Zeev. 1984. "Peace by Empire? Conflict Outcomes and International Stability, 1816–1976." *Journal of Peace Research* 21(3):227–241.
- Maoz, Zeev. 1999. "Dyadic Militarized Interstate Disputes, version 1.1." Internet.
- Maoz, Zeev and Bruce Russett. 1993. "Normative and Structural Causes of the Democratic Peace." *American Political Science Review* 87(3):624–638.
- Morrow, James D. 1991. "Alliances and Asymmetry: An Alternative to the Capability Aggregation Model of Alliances." *American Journal of Political Science* 35:904–933.
- Oren, Nissan. 1982. Prudence in Victory. In *The Termination of Wars*, ed. Nissan Oren. Jerusalem: The Magness Press.
- Russett, Bruce. 1993. *Grasping the Democratic Peace: Principles for a Post-Cold War World*. Princeton, NJ: Princeton University Press.
- Selten, Reinhard. 1978. "The Chain Store Paradox." *Theory and Decision* 9:127–159.
- 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.
- Wagner, Robert Harrison. 1993. The Causes of Peace. In *Stopping the Killing*, ed. Roy Licklider. New York: New York University Press.



Werner, Suzanne. 1999. "The Precarious Nature of Peace: Resolving the Issues, Enforcing the Settlement, and Renegotiating the Terms." *American Journal of Political Science* 43(3):912–934.