Settling It on the Field : Battlefield Events and War Termination

Kristopher W. Ramsay

Journal of Conflict Resolution 2008 52: 850
DOI: 10.1177/0022002708324593

Journal of Conflict Resolution
Volume 52 Number 6
December 2008 850-879

10.1177/0022002708324593
http://jcr.sagepub.com
hosted at
http://online.sagepub.com

Settling It on the Field

Battlefield Events and War Termination

Kristopher W. Ramsay
Department of Politics
Princeton University

Using a sample of battlefield data from twentieth-century wars, the author of this article tests a number of previously untested hypotheses linking battle events and the decision to end violent conflict. The author explores how factors like the distribution of power, battlefield casualties, and information that flows from the battlefield influence war termination. The analysis speaks to the validity of competing rational choice theories of war termination.

Keywords: *war termination; battles; learning*

Traditionally, explanations for war have been explanations of why states choose to start wars. Framed in this way, theories of war do not consider waging war as a process for achieving some larger goal; rather, war is treated as an outcome to be explained. But as our understanding of the decision to go to war develops, we are beginning to see that the question of why wars start is intimately related to how and why they end. Since leaders' expectations about the outcome of war affect their decision to initiate conflict, it is argued that understanding war termination is a necessary condition for fully understanding why wars start (Blainey 1988; Wagner 2000).

This growing interest in the process of fighting and ending wars has led to a number of interesting and sometimes contradictory claims about the timing and terms of war termination. While some work has begun to test the hypotheses coming out of this literature (Bennett and Stam 1996; Reiter and Stam 1998b; Slantchev 2004; Werner 1998), there has consistently been a discrepancy between the empirical tests and some of the key hypotheses in the literature. Specifically, many of the hypotheses concerning war termination relate battle outcomes to the decision to continue or end wars. Empirical work, however, has only considered

Author's Note: I would like to thank Stephen Biddle for sharing his data and for helpful conversations that shaped the early versions of this article. Advice and comments were also given by Charles Boehmer, Raymond Duvall, Joanne Gowa, Songying Fang, Mark Fey, Dougle Lemke, John Freeman, Branislav Slantchev, Curt Signorino, Randy Stone, and participants of the Methodology Workshop at the University of Minnesota, Twin Cities. The data used for this study are available at http://jcr.sagepub.com/supplemental/.

data where the unit of observation is the war year or the wars themselves. So if battlefield outcomes influence leaders' decisions to keep fighting or to pursue a peaceful settlement, then the effects of battlefield events can only be known from data that use the battle as the unit of analysis.

This article, then, takes a step in bringing the empirical and theoretical literatures closer together by examining battle-level data and drawing direct connections between battle-level events and the ending of wars. While we know of a growing theoretical literature on the consequences of battles on political settlements (Smith 1998; Slantchev 2003b; Powell 2004) and an empirical literature on how tactics and capabilities influence military outcomes (Reiter and Stam 1998a; Biddle 2001), to our knowledge, there exists no empirical study that investigates battlefield hypotheses related to the war termination literature.

Empirically investigating the battle-level hypotheses of the war termination literature is a difficult task, and the analyst faces two challenges. First, there is limited quality data on battle-level events across numerous wars and over a significant time series. The best existing (and only) data are in the U.S. Army's CDB90 (or HERO) data set on battles.[1] Given that these data are not perfect, their two most prominent limitations deserve some discussion.[2] First, there are issues of coverage. Generally, wars covered by the data set involve Western powers and Israel. There are a good number of wars—the Sino-Indian war, the Soccer War, and the Uganda-Tanzanian War, to name a few—that are classified as twentieth-century wars by the Correlates of War project but for which there are no data in CDB90. Even within the wars covered, there is variance in the level of detail. The Russo-Finish War, for example, has one battle (where there are arguably two), while Okinawa has more than twenty recorded battles. Second, naval battles are not in the data set. For wars like the Russo-Japanese War, in which the naval battle of Tsushima is generally regarded as important, this is a limitation.[3]

With all of these issues, one might wonder why we should look at battles and not other events during war. There are three reasons battles are important. First, battles are the central feature of war. They are the time and place where one army attempts to destroy the other's ability to achieve its goals through coercion. Second, an important function of combat is to convince the opponent to accept a settlement. In the war termination literature, this occurs when the sides have learned enough about their prospects in a war to the finish to decide continuation is not worth the cost. Obviously, battles are a relevant source of information regarding the prospects for total victory, but they are not the only source. Intelligence, media reports, and diplomatic tactics are also sources of information in times of war and may influence war fighting decisions. However, military engagements have received special attention because, unlike diplomacy or intelligence, battlefield success is generally viewed as a nonmanipulable source of information; that is, while strategic incentives may cause decision makers to make outrageous demands at the negotiating table or leak false intelligence to help their cause, battle outcomes reveal information about the quality and strength of the military that is difficult to mask.[4] Third, the relationship between information flows from battle and war termination probability is one of the places where two similar but distinct information arguments, linking the fighting of wars to the acceptance of peaceful settlements, make different predictions. Therefore, the analysis of even noisy imperfect data on battles can make a useful contribution to the debate between two plausible, well-known, and otherwise largely untestable theoretical arguments.

The second challenge to studying battlefield hypotheses is methodological. When considering competing strategic hypotheses, generated from different theories of war, the statistical issues can be complicated. As shown by Signorino and Yilmaz (2003), imposing the wrong functional form for the data generating process on one's analysis can lead to biased estimates of relationships. Moreover, many hypotheses that can be derived from the existing literature are about the functional form or shape of the relationship between independent variables and the probability of war termination. Ideally, we would like to use techniques that reduce the needed number of functional form assumptions and allow the analyst to estimate the possibly nonlinear relationships induced by strategic decision making directly. One way to do this is by using semi- and nonparametric regression techniques, as suggested by Beck and Jackman (1998). Given the methodological complications implied by testing hypotheses from competing strategic models, these flexible estimation techniques seem most appropriate.

This article continues as follows. In the next section, I lay out a diverse set of hypotheses that are implied by various strains of what could be called the war termination literature. I then provide a discussion of the data and the variables they contain. Next, I discuss in more detail the method used to analyze the data and why I believe it is the appropriate method for the task at hand. Section 3 is the core of the article: there, I estimate a semiparametric model of war termination, investigate the conditional effects of my key variables, and check the robustness of my results. Section 4 concludes the article and contains a discussion of what my results imply for future work in this area.

From Fighting to Peace: Some Unanswered Questions

Building on the earlier work of Wittman (1979), Blainey (1988), and Wagner (2000), the recent literature on bargaining and war has focused on a new dimension of the strategic problem faced by states during a conflict. In particular, some scholars have moved their attention from the puzzle of why states fight inefficient wars to ask how fighting resolves the problems that lead to war. This question is particularly relevant for the vast majority of wars that end by negotiated settlements, not military collapse; that is, as these wars were settled by bargaining after costly conflict, one might wonder why the same settlement was not reached before both sides paid the cost of war. A particular aspect of war fighting on which this research has focused is the idea of the battle (Bennett and Stam 1998; Powell 2004; Reiter and Stam 2002; A. Smith 1998; A. Smith and Stam 2004).

In both classic (Dupuy 1979; von Clausewitz 1832/1984) and more recent research (Slantchev 2003b; A. Smith 1998) on war fighting, two general mechanisms are theoretically central to answering the question of how fighting battles leads to the resolution of conflicts. One set of explanations claims that disputes are resolved through violent coercion, the ability to hurt, and the asymmetric distribution of costs from fighting. The second set of explanations focuses on how fighting a war resolves information asymmetries that lead to war in the first place. These differing views of battle lead to a series of conditional hypotheses linking country and battle characteristics to outcomes in war. In this article, I test various war fighting hypotheses on battle-level observations. Given the issue with battle-level data, it is reassuring that I find results consistent with many of the intuitive claims in the literature. Moreover, the analysis also speaks to some competing hypotheses regarding the learning process during war.

The first hypothesis I consider is one that relates the ability to impose costs on the decision to continue fighting:

Hypothesis 1, All but Over Hypothesis: Wars between asymmetric adversaries will be short.

This hypothesis is not directly related to battlefield outcomes or characteristics, but rather to the assumption that when weak states fight stronger states, the greater the difference in their relative strengths, the easier it should be for the strong side to overpower the weak (Dupuy 1979; von Clausewitz 1832/1984; Bennett and Stam 1996). For example, Clausewitz (1832/1984, Book III, chap. 8) discusses how supremacy in numbers is an important factor in the outcome of an engagement, campaign, and war. Similarly, Dupuy (1979, 55–56) writes that power ratios in conflict are important for predicting outcomes of battles and wars. In any case, it is often claimed that the more powerful side has the advantage. As such, wars resolve disputes by allowing the strong to impose its will, through its superior power, on the weak. Such a mechanism, linking war termination and duration to the balance of power, implies that the relationship between the capabilities ratio of the initiator[5] and the termination of conflict should be U-shaped (Bennett and Stam 1996); that is, when two countries have a large difference in their capabilities, we should expect that any given exposure to battle is more likely to lead to victory and the ending of conflict because the stronger state is likely to overpower the weaker.

As Schelling (1966) argues,[6] the power to hurt, and bear costs in return, is another important factor in determining the path and outcome of war. To quote Schelling (1966, 2), “the power to hurt is bargaining power.” Combined with the understanding that war is an attempt to use violence to coerce one’s opponent to give up some territory or make some policy change desired by the war initiator

(Wagner 1994; Hobbs 1979), I expect an increasing relationship between the difference in the rate of cumulative casualties for each side and the probability of war termination; that is, by imposing asymmetric costs on the target for failure to comply with the war initiator's wishes, or on the initiator for pursuing violent coercion, we may expect that the side paying the higher cost will eventually capitulate in the face of increasing scales of defeat[7]:

Hypothesis 2, Brute Force Hypothesis: The probability a war ends is increasing in the imbalance of loss between the two parties.

In contrast to the power-based explanations of war termination, the game theoretic literature has argued that battlefield events may have important effects on the beliefs of belligerents (Schelling 1960, 1966). From a theoretical perspective, battles have certain characteristics that other sources of information do not. One especially important characteristic of battles is that they are viewed as nonmanipulable; that is, on the battlefield, one cannot pretend to be stronger, to have better leadership, or to be better equipped than one actually is. The same is not true at the diplomatic negotiating table. At the same time, independent of what happens on the battlefield, the fact that a country is willing to pay the cost, in treasure and blood, of fighting is also informative about their preferences over changes to the status quo.

Specifically, if war is the result of asymmetric information and battle data are a new source of public information, then the information acquired through warfare may affect warring parties' calculations concerning the value of continuing the conflict. As Fey and Ramsay (2007) point out, there are two information-based mechanisms for war. By one mechanism, war is the result of mutual optimism (e.g., A. Smith and Stam 2004). In such a world, countries have private information about their relative likelihood of success in a drawn out conflict. In each country's case, its private information tells it that the expected value of fighting is greater than the expected value of a settlement. In particular, the country's information regrading the true probability of winning in a drawn out conflict is inconsistent. This inconsistency of beliefs, the argument goes, leads both sides to prefer fighting to war. Fighting battles resolves this uncertainty because battlefield outcomes give new information about the likely outcome of a long conflict. For example, in models like those found in Filson and Werner (2002, 826–28, Results 3, 5, and 6), A. Smith and Stam (2004, 797, Lemma 1), and Werner and Yuen (2005, 265), as more information is provided by fighting battles, countries' estimates of success are revised, eventually leading to the creation of a bargaining space that allows for an ending to hostilities. This theoretical claim leads to the following hypothesis for my battle data[8]:

Hypothesis 3, *Informative Battle Hypothesis:* If battles reveal information about the long-run likelihood of success in war to leaders, then the probability a war ends is increasing in the amount of information coming from the battlefield.

Alternatively, war could be a process of signaling resolve and screening out weak challengers to the status quo; that is, when countries fight wars, they may not learn much about the likelihood of success in the long run but instead may learn about the will or interests of their opponent. Such a mechanism underlies many of the bargaining models of war. It is important to note that this screening (or signaling) mechanism predicts a different pattern in the relationship between the amount of battle information received by countries and the likelihood of war termination. Data in such a world would be generated by different types of countries: countries with revisionist leaders who prefer war to any peaceful settlement, even with complete information, and countries with leaders who bluff their way into war in hopes of a better peaceful settlement. The first type of country fights because it prefers that option, but the second type is in a war because its bluff was called. In such a world, war solves information problems quickly (Slantchev 2003b, 629). Data generated by wars that originate in this way would show an increasing probability of war termination after some short number of battles or battle days, then a decrease in likelihood as the weak or bluffing states will have fallen out of the data. This gives us the following hypothesis:

> *Hypothesis 4, Screening Hypothesis:* If battles reveal information about the resolve of revisionist states, then the probability a war ends is at first increasing, and then decreasing, in the amount of information coming from the battlefield.

To reiterate, the process that generates the pattern of hypothesis 4 is one where weak (or unresolved) types of revisionist take positions diplomatically or politically that lead to called bluffs and then short wars. In these cases, some countries that prefer not to fight are fighting in the data set. These unintended wars should be short because these countries have incentives to avoid paying the additional costs of fighting a long war they did not want in the first place. Therefore, the data generated by such behavior should contain a significant number of short wars. After a while, however, all the bluffing revisionists will be screened out of the data, and the remaining long wars will be contested between states that would fight with complete information; that is, the states are no longer learning about each other's resolve, but rather, they are in the war to win it. As a result, the probability that one country or another sues for peace, *ceteris paribus*, will be decreasing.[9]

Finally, military practitioners and historians also identify another possible battlefield event that is likely to influence war termination: shocking victories (Weigley 1991). As the surprising reversal of fortunes surrounding the Battle of Mount Hermon in the Golan Heights (1973) during the October War demonstrates, unexpected battlefield outcomes appear to have important independent effects on the likelihood that a war terminates; that is, deviations from expectations in combat, as measured by differences between prior trends and current battlefield outcomes, are likely to increase the likelihood of war termination. Thus, I have a corollary to the informative battle hypothesis and my final expected result:

Hypothesis 5, Major Victory (Defeat) Hypothesis: The probability a war ends is increasing in the deviation of its scale of victory (defeat) from the previous trend.

Why would such battles be important? Because we might expect that shocking victories, or relative routes, reveal more information than battles that are representative of the cumulative trend within a war. Such *decisive* or *turning point* battles have received considerable attention from historians and historically minded political scientists alike and are likely to influence the decision to end a war (Davis 2001). This hypothesis is also put forth by Gartner (1997, 50), where some organizational models suggest that a reevaluation of strategies occurs after high-profile events.

Battle Data

The data used in this article is the U.S. Army's CDB90 data set, sometimes referred to as the HERO data set (Dupuy 1995). In particular, I use the observations from the CDB90 for wars in the twentieth century. This is because, as has been noted by others (i.e., Measheimer 1989; Brooks 2003; Desch 2002), the original data set has some flaws. As was already mentioned, the CDB90 data set does not contain naval battles. In the data set, the Russo-Japanese War's last battle is Mukden (February 21–March 10, 1905), not the Battle of Tsushima (May 27, 1905). Similarly, there are occasional omissions from the data set that are surprising. For example, in the Winter War between Russia and Finland, the last battle recorded is Soumussalmi (January 1940), though the war did not end until after the assault on the Mannerheim Line (February 13, 1940).[10]

Desch (2002) points to two additional problems with the data. First, a random sample of eight battles from the original data set—covering the period from 1600 to 1973—deemed the data to be a poor match for the historical record. Second, Desch comments that the coding of many variables in the data set are hard to replicate, like "moral" and "strategy." The first issue is somewhat resolved by looking at only twentieth-century battles. While it is not surprising that there would be little agreement among historians regarding details of battles that occurred in the 1780s, because the documentation of events and the volume of primary sources are limited by these battles' distance in history, the historical record is unarguably clearer for the modern period; that is, there is more detailed documentation available for twentieth-century wars, and as there were hundreds of observations from older battles—dating back to the Netherlands War of Independence in July 1600 and including such a wide variety of historical events as World War II and the American Civil War—it is not much of a surprise that eight randomly selected battles might generate disagreement among military historians. As for Desch's second concern, I agree that the collectors of the CDB90 data set did not leave accessible guidelines for replicating the coding of their more subjective variables. For this reason, I do not focus on these more troublesome variables, instead using them only to probe the robustness of results found under different specifications; that is, the most important data used from CDB90 focus on casualties and other replicable measurements.

Another obvious issue is the potential for selection effects.[11] The CDB90 data set could be criticized for focusing largely on U.S. and European wars. The list of wars in the Correlates of War project not covered by CDB90 is quite long—including the 1948 Indo-Pakistani War, the Cyprus War, and the Sino-Vietnamese War, to name a few. While there is no obvious reason to believe that the mechanisms in the sample of wars in these data are different than others, if there is selection, then the inferences would be inconsistent estimates of the effects of these variables on war termination generally. In the section on robustness, we get a feel for the strength of my results through various robustness exercises.

A final reason I work with the twentieth-century data is the double counting of operations and battles that occurs in the full data set. As noted by Biddle (2001) and Biddle and Long (2004), CDB90 often includes an observation for an operation, like operation GOODWOOD in World War II, as well as the numerous individual engagements that made up the larger operation. As such, the data set tended to overcount and overrepresent major operations. A reason for limiting the analysis to the twentieth century is that there exists a clean version of the data set for that period, in which the double counts have been corrected.[12]

For the purpose of this article, I also have made some changes to the data set's organization, following previous conventions. First, I break a number of multilateral wars into series of bilateral conflicts, or warring dyads, which take place within the framework of the larger war. An attempt was also made to organize the data in a way consistent with the description of twentieth-century wars found in Reiter and Stam (1998b, 2002) and Slantchev (2004). For some conflicts, this was relatively easy, such as World War II. For others, it was not, such as the Arab-Israeli wars. However, the logic that underlies dividing the world wars into a series of bilateral wars would seem to apply equally to any multilateral conflict in which one state is fighting others on different fronts. As Reiter and Stam (2002, 39) argue, "decision makers rarely anticipate or think in terms of larger systems of wars, but instead usually think in terms of sequences of opponents." Moreover, the disaggregation of larger wars allows a clearer description of which side won, on which front, and when. Additionally, from any single participant's perspective, the engagements, that is, battles, that make up a war are almost always (at least as reported in CDB90) of a bilateral nature. So, considering each warring dyad to be a bilateral conflict allows us to more closely track the course of events between any two belligerents, and as a result, I am able to capture the fact that in multilateral wars, the conflict between any two sides is terminated at different times within the larger conflict. A list of the wars in this data set is presented in table 1.

Table 1
Wars in CDB90 for the Twentieth Century

War Name	Date	Number of Warring Dyads
Russo-Japanese War	1904–1905	1
First Balkan War	1912–1913	1
Second Balkan War	1913	1
World War I	1914–1918	10
Russo-Polish War	1920	1
Manchurian Incident	1938–1939	1
Russo-Finish War	1939–1940	1
World War II	1939–1945	6
Arab-Israeli War	1948	3
Korean War	1950–1953	1
Suez War	1956	1
Six Days' War	1967	3
Israeli-Jordan War	(March) 1968	1
October/Yom Kippur War	1973	3
Israeli-Lebanon War	1982	1
Total number of warring dyads		35
Total number of battles		379

Dependent Variable

In addition to generating a variable for bilateral wars, I also coded *war termination*. This is the central dependent variable of my analysis. I defined a war termination as occurring after the final battle within a warring dyad. Coding war termination in this way gives us the largest number of termination events. Each battle therefore provides a window or marker to assess various combat-related measures and progress in the war. Given the irregularity of battles, cutting the war into a time series of weeks or months seems less natural. Moreover, the battle is the natural point of reassessment for the decision maker and is also a key point for evaluating new information, according to theory. So another way of interpreting my analysis is that I am asking the question, why was it after *this* battle that the war ended?

To check the robustness of results in this article, I also coded conflict-terminating events in alternative ways. For example, there exists a warring dyad with one battle between the United Kingdom and Japan in Malaya in 1941. Such a battle is questionably a war-terminating event. Similarly, one could question whether or not the warring dyad between France and Germany in 1940 ended with a war-terminating event or whether Okinawa was a war-terminating event between the United States and Japan. Therefore, a second series of war terminations was created in which I recoded some conflict-terminating events to be nonterminating. The set of judgment-modified events was very similar to the original set of codings and had no effect on the results. In the end, the analysis here uses all war terminations, given the lack of difference between the judgment-based codings and the initial measure.[13]

Key Independent Variables

I start by defining measures of the wartime balance of military power. The bilateral balance of power is measured using the ratio of military capabilities. I take my measure from Singer, Bremer, and Stucky's (1972) Composite Index of National Capabilities (CINC). To generate the balance of power, I then use the initiator's CINC score and divide it by the total CINC score for the warring dyad. It is an annual measure that can vary within a war but does not vary within a single battle observation.[14] This variable has a mean of 0.40 over all observations and a standard deviation of 0.21.[15]

Because the use of force and its battlefield consequences are also of interest, I employ a measure of combat casualties. Casualties were easily generated using a running tab of each side's losses from CDB90's casualty count for each battle. For the initiator, the cumulative casualties measure had a mean of 138,927.1, while the target's was 211,006.6.

Using the cumulative casualty rates, I then generated a variable that measures each side's relative loss in the bilateral combat; that is, I have the cumulative loss as

$$\text{cumulative initiator loss} = \frac{\text{cumulative casualties (initiator)}}{\text{cumulative no. of troops sent to battle (initiator)}} \tag{1}$$

$$\text{cumulative target loss} = \frac{\text{cumulative casualties (target)}}{\text{cumulative no. of troops sent to battle (target)}}. \tag{2}$$

My measure of coercion is equal to the absolute value of the difference in the cumulative losses. Therefore, a high value on the coercion variable implies that one side is losing troops at a higher rate than the other. This is thus a measure of the coercive power one side is able to bring to bear on its opponent.

Next, I come to my measures of information flows from the battlefield. The theoretical literature often argues that the battle is the key source of information during war. One possible measure of the information flow from the battlefield is the cumulative number of battles. In reality, however, battles are hardly standard units. Some battles can be very long, while others are very short. The issue then becomes how to generate a scale that takes the unfortunate fact that battles may differ greatly into account. The way I do this is to employ a standard measure of battlefield experience, the battle day. Since I have good data on the length of individual battles, I measure the flow of battlefield information in terms of the number of battles days the belligerents are engaged; that is, when troops are engaged in

Table 2
Means and Variances for Key Variables of Interest

Variable	Mean	σ	Minimum	Maximum
Battles per war	56.39	49.50	0	173
Battles per warring dyad	15.27	17.90	0	79
Battle days	63.83	80.01	.02	389.20
Battle duration	5.19	10.72	.014	130
Capabilities ratio	.40	.21	.01	.99
Military trend (initiator)	−3.51	5.68	−17	9
Casualties (initiator)	138,927.1	310,367.2	10	1,731,002
Casualties (target)	211,006.6	457,311.5	10	1,919,962
Coercion	.2732	.2242	0	.9139
Shocking victory	−.0099	.2271	−.6412	.7716

combat, the countries are learning about the other's ability and resolve. Since I use the latter in my analysis, I would note that the mean of the number of battle days for a warring dyad is 63.83, with a standard deviation of 80.01.

Finally, it is conventional wisdom that the information from the battlefield has another dimension; that is, it is not just the amount of information the belligerents may receive, but also its nature. In particular, it is sometimes argued that battles with outcomes that significantly deviate from expectations are important. These are sometimes called *decisive battles*. Therefore, I construct a variable for *shocking victories*. For the first observation in each warring dyad, I measure the difference between the two opponents' capabilities ratio and the current battle's rate of loss for the initiator. The rate of loss for the initiator is the measure of its casualties divided by the number of troops sent to battle, which is then divided by the sum of this ratio for both the initiator and the target. Throughout the war, I proxy a country's expected rate of loss in the current battle by the cumulative rate of loss; that is, if the cumulative rates of loss are as in equations (1) and (2), then

$$\text{cumulative rate of loss (initiator)} = \frac{\text{initiator loss}}{\text{initiator loss} + \text{target loss}}.$$

For every observation, but the first, within a warring dyad, the shocking victory variable is generated by taking the difference between the initiator's cumulative rate of loss through the previous battle $[t-1]$ and the rate of loss in the most recent engagement. So if we imagine that the capabilities ratio, and then the previous proportions, constitute the trend in the rate of loss for each subsequent battle, then our shocking victory measure captures the difference between the expectation and the realization in the current battle. For convenience, Table 2 contains the summary statistics for many of these key variables.

Control Variables

We can also imagine that some other variables, not related to our hypotheses, might be relevant for explaining war termination. First, there are variables the literature has hypothesized that are political and not directly related to the battle itself. For example, I control for what Slantchev (2004) calls prearmistice negotiations, which might be an important factor leading to peace. As such, I code a dummy variable that equals 1 for every battle in which the two sides have been involved in prior negotiations or where there has been international mediation or proposed intervention. Last, like many previous studies of war, I include measures of regime type (Bennett and Stam 1996). The variable liberal war is a dummy variable that indicates that at least one side in the conflict is a liberal democracy, as listed by Doyle (1986). I also include an indicator variable that marks all warring dyads initiated by a liberal state (Clark and Reed 2003; Bennett and Stam 1998; Reiter and Stam 1998b; Bueno de Mesquita et al. 1999).[16]

Second, there are also battle characteristics that may be important for conditioning my estimates. For example, as J. Smith (1995) argues, the military trend may be an important factor in determining the likelihood of war termination. As such, I construct a measure of the military trend that starts at zero and increases by 1 for every victory by the initiator, decreases by 1 for every initiator loss, and remains unchanged if the battle outcome is a draw.[17] CDB90 also has many qualitative variables that are not as reliable as the other data but are worth investigating. Later, in the robustness section, I consider the effect of troop, leadership, and intelligence quality as well as physical advantages such as logistical and terrain concerns.

Method

To analyze the CDB90 battlefield data, I use a variant of semiparametric spline regression that is flexible with respect to the structure of the response variable, Y. The primary reason for choosing this approach is that while the linear class of models is very rich and useful in a number of settings, some practical issues repeatedly arise in the analysis of data generated by strategic interactions. First, among them is model specification. For inferences to be valid, the functional form of the relationship linking our dependent variable (Y_i) and the explanatory xs must be correctly specified. In many applications in the social sciences, however, a linear specification is not appropriate.[18] If we had a particular strategic model in mind, which implied a specific structure to the data generating process, then one avenue to pursue would be a structural equation estimation.[19] Here, we have no such model, but rather, we draw competing hypotheses from a variety of different theoretical frameworks. Therefore, we may want the data to tell us the functional relationship between the response and predictor variables. Such a method would allow us to see if the relationship suggested by competing theories of war termination exist in the data. In practice, the preferred model would be

$$Y_i = \beta_0 + f(\mathbf{x}_i) + \varepsilon, \tag{3}$$

where f is some unknown function mapping our predictors to the response variable. The question then becomes one of estimating f, rather than coefficients.

To estimate f, we need to make some assumptions. First, we assume that f is a smooth function of the variables, requiring it to have a continuous second derivative. Second, we assume that there are no wild changes in the relationship between our dependent variable and our independent variables, for small changes in the xs. These assumptions lead us to characterize the trade-off between fit and summary within our model with a measure of volatility, or "wiggliness," that penalizes a candidate function $\hat{f}$ if it too closely tracts the data, while maintaining the idea that the best candidate function should minimize the distance between the predicted value of y, given x, and the observed value of the response variable. Equation (4) is an objective functional, like least squares, that characterizes such a trade-off:

$$\min_{f(\cdot)} D = \sum_{i=1}^{n} (Y_i - [\beta_0 + f(\mathbf{x}_i)])^2 + \lambda \int_a^b f''(\mathbf{x}_i)^2 dx. \tag{4}$$

Looking at equation (4), we see that the first part of the functional is minimized by choosing an f that minimizes the distance between the predicted and actual values of the response variable. Therefore, a function is rewarded for fitting the data well. The functional penalizes a candidate function by the integral of the second derivative of f squared. This portion of the functional rewards functions for their smoothness and punishes functions that are too volatile.[20] Given this setup, there exist a number of techniques for estimating the unknown nonlinear functions of interest (Wood 2006). Subsequently, I use thin plate regression splines as doing so is computationally efficient and overcomes some limitations of other smoothing methods.[21]

Semiparametric Model and Results

With my method described, I now turn to the analysis of my data.[22] Given my hypotheses, outlined above, and the list of variables taken from the theoretical literature, we have a clear list of variables that we may wish to include in our model. Moreover, if we take our hypotheses as being *ceteris paribus*, the most natural functional form is a generalized additive model.

Table 3 presents the results from two semiparametric specifications of the war termination model. The table reports regression coefficients for linear variables and the degrees of freedom and χ^2 statistics for the smoothed terms.[23] Furthermore, the estimated functions for the nonlinear variables can be found in figure 1. Model 1

Table 3
Semiparametric Estimates for Probability of War Termination

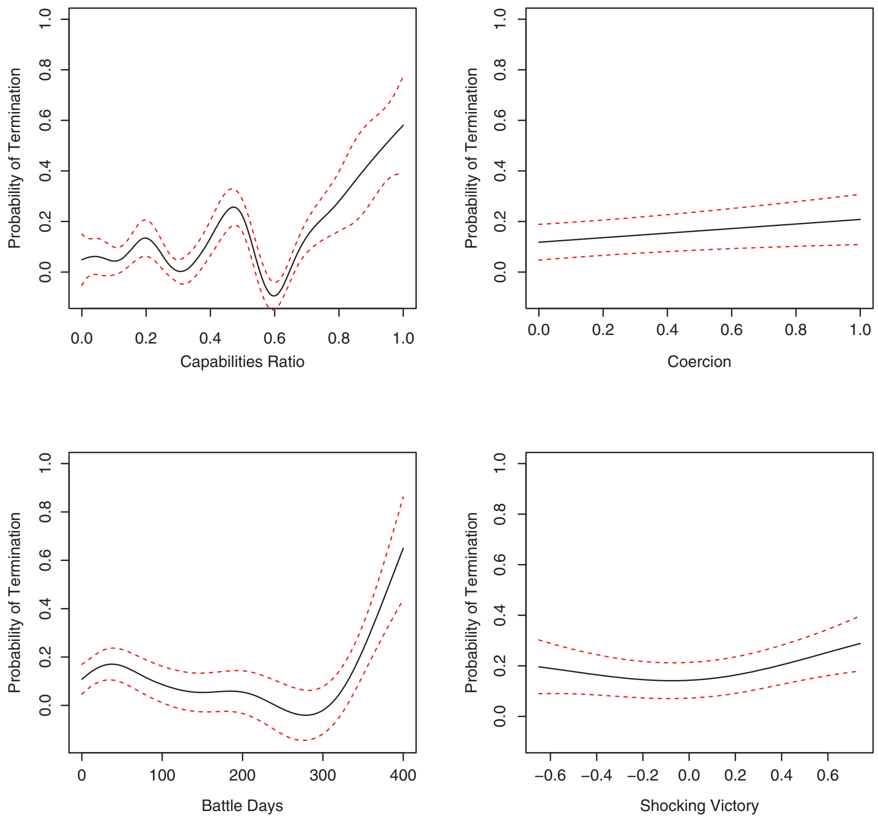
	Model 1	Model 2
Intercept	.24 (.038)**	.19 (.05)**
S (capabilities ratio)[a]	9.14 (50.85)**	9.14 (50.85)**
S (battle days)[a]	5.17 (19.99)*	5.17 (19.99)*
S (shocking victory)[a]	1.77 (10.01)	1.77 (10.01)
S (coercion)[a]	1.0 (1.32)	–
S (military trend)[a]	1.0 (5.15)*	–
Coercion	–	.09 (.07)
Military trend	–	−.01 (.004)*
Democratic initiator	.193 (.07)**	.193 (.075)**
Liberal war	−.189 (.04)**	−.189 (.04)**
International mediation	−.03 (.05)	−.03 (.051)
Adjusted R^2	.153	.153
Deviance explained (%)	20.0	20.0
GCV score	.08	.08
n	379	379

Note: Pooled results reported. The numbers reported for splines are the estimated (or effective) degrees of freedom. The estimated degree of freedom is the trace of the hat matrix and captures the effective degree of the spline. Reported after the spline's estimated degrees of freedom is the χ^2_e statistic. The χ^2_e statistic is for the significance of the smooth compared to the null of no effect for the variable. It is distributed with degrees of freedom $= e$, the estimated degrees of freedom of the spline. It should be noted, however, that the generalized cross-validation fitting procedure (discussed in the appendix) used to select the smoothing parameter makes the χ^2 test only approximate and of lower power.
a. Estimated degrees of freedom and χ^2 statistics reported for smooth terms.
$*p < .05$. $**p < .01$.

reports the estimates of the semiparametric specification where all nonbinary independent variables are smoothed. Since the estimation finds that the *coercion* and *military trend* variables are linear, model 2 presents the linear coefficients for those variables.[24]

From the control variables, we find, all else equal, that wars in which at least one belligerent is a liberal state are less likely to terminate, implying that wars with democracies tend to be longer, that is, take more battle days. Note, however, that that result is controlling for the fact that the initiator of the war may have been a democracy. In fact, the democratic initiator variable is significant and positive, implying that when democracies start a war, those wars tend to be shorter. Together, these results are consistent with the findings of Clark and Reed (2003), Reiter and Stam (1998b), and Bueno de Mesquita et al. (1999) that democracies tend to fight harder when attacked, but pick their fights carefully. However, unlike Slantchev (2004), the effect of regime type remains even when controlling for other factors such as combat duration.

Figure 1
Semiparametric Estimates for Probability of War Termination



Note: The dark lines are the estimated function of the *x*-axis variable. Predicted effect of coercion (right) each with ± 1 standard error of the prediction. Smooth estimates with pointwise credible intervals available from the Web appendix.

We also find that the military trend of the initiator has an effect on the likelihood of termination. In particular, as the initiator's net wins increase, the likelihood of termination decreases. This result demonstrates that in the CDB90 data, if the initiator is winning more battles than it is losing, then conflict is less likely to end in a given battle, all else equal. However, when the initiator loses more battles than it wins, the probability of the war ending increases. Such a result would be consistent with a view of war in which that initiator is trying to test the strength of the target. If the initiator is winning often, then it continues its military push because the trend reinforces its prewar belief that a war to the finish will go in its favor. If, on the other hand, the opponent is winning more often, then the belligerents are more likely to be able to agree to stop fighting. In fact, a simple cross-tabulation of war trends to war outcomes reveals that as the military trend goes in favor of the initiator, the initiator is more likely to win the war, while in cases where the trend is against the initiator, the initiator is more likely to concede. It is also worth noting that this result is conditional on the information flows; that is, a specification of the model where the trend is dropped produces the same results for other variables. If, however, we choose to keep the trend variable and drop the battle days flow variable, the trend effect is not close to significant.

Somewhat surprisingly, we find that the existence of prearmistice negotiations and/or international intervention has no effect on the likelihood that a war ends. While beyond the scope of this article, this result provides evidence for the claim that once a war has begun, battlefield events are at least as, if not more, important in determining war outcomes as the actions of mediators and third-party institutions. On the other hand, the prearmistice negotiation variable is a pretty poor measure of the effect of diplomacy. Ideally, future work will collect data on various diplomatic events during war and better analyze the claims regarding the relative importance of battlefield and nonbattlefield information.

Turning to my key variables of interest, I start by discussing the two variables associated with the use of force: the belligerents' capability ratios and the coercion measure. From table 3, and the predicted smooths in figure 1, we see that the relationship between the capabilities ratio (i.e., the dyadic balance of power) and the likelihood that a war ends on a given battle is highly nonlinear. In particular, we see that there is a distinct spike in the probability that a given war ends, all else equal, when power is distributed equally. Were all of our results this way, we might question whether my model was fitting noise in the data. Given the reasonable nature of the other results, this nonlinear fit is worth exploring.[25]

This result contradicts previous findings reported by Bennett and Stam (1996). One possible explanation for this finding is that the distribution of power is likely to be a good, if not perfect, forecasting tool for both the length of a war and prospective outcome of continued conflict.[26] As such, a possible explanation for this peak at .5 is that when the distribution of power is even, neither side has much hope for a long-term success conditional on the other variables. Under these conditions, both sides would be more willing to reach some settlement to end the war, rather than continue fighting. Notice how this incentive differs from the way the capabilities ratio may influence the decision to go to war. For example, in the power transition literature, it is argued that small shifts from parity make conflict less likely, while power parity makes war more likely (Organski and Kugler 1980). Before war breaks out, the power transition claim has seen mixed results (Bennett and Stam 2004). However, given that a war is ongoing, the logic may be turned on it head. To see this, consider the shape of the predicted effect and the potential difference in the two countries' cost-benefit analyses. On one hand, when there is a slight imbalance in capabilities, the stronger side may have an incentive to continue fighting as its relative military power predicts likely success in the long term. Similarly, when the difference in the balance of power is relatively small, the weaker side may have an incentive to keep fighting as the short-term cost of capitulation is large and the long-term likelihood of success is not too bleak.

So, as can be inferred from Figure 1, when there is an even distribution of power, neither side has a reason to expect long-term success. However, when there is only a slight difference in the balance of power, the long-run incentive relevant to the powerful state and the short-run costs faced by the weak state may give both sides an incentive to keep fighting. Finally, as the imbalance of power becomes severe, we find the intuitive result that the probability of war termination increases.[27]

As for the effect of coercion, the conditional effect of the difference in rate of cumulative casualties is increasing and linear. The coefficient, however, is not significant, $p = .25$. One could imagine another specification and a different effect of casualties on war termination. Gartner (1997) suggests that an acceleration in trends is important for decision making. An additional specification of the model where the coercion measure was replaced with a second difference returned a statistically significant quadratic relationship (not reported).[28]

Finally, I turn to the relationship between battlefield information and the likelihood of conflict termination. From the predicted smooth of shocking victory in figure 1, we see the convex relationship we expected. In particular, there is a cluster of observations near zero, and on either side of the origin, the probability of conflict termination increases as the battle's rate of loss deviates further from the experience accumulated from past battles. The substantive effect of shocking victories, however, seems rather small, as at the extreme values of the variable, the probability that a war terminates, all else equal, goes from about .08 to .25. This effect is not statistically significant, and as suggested by an inspection of figure 1, we cannot rule out the possibility of no effect. Furthermore, the result does not support the argument that unexpected victories (losses) can have profound effects on the willingness of belligerents to continue fighting. To some extent, this makes sense. If a single battle's outcome is determined in part by luck or chance, then even if the effect of luck is small on average, it is not clear why a single outlier observation should have a large effect on a leader's beliefs, especially when there may be a large amount of battlefield information to consider.

Possibly the most interesting result is the relationship found between the flow in information from the battlefield, that is, the number of days the warring parties are engaged, and the probability that a war ends. From the predicted smooth in figure 1, we see that over a large portion of the range, the probability that a war ends is *decreasing* in the amount of battlefield information. What does this imply? Well, if we expected that battlefield events relayed information about the underlying probability of success in a total war, then the plot should be generally increasing (with possible flat or small decreasing region). It definitely should not be the case that after 250 to 300 battle days, *ceteris paribus*, the two sides are less likely to end the war than they were at the first battle.

The interesting characteristic of this estimated curve is that, at first, it is increasing, then, for a large portion of the range of the measure, it is decreasing. Finally, it increases rapidly. This empirical relationship cast doubt on theories of war termination based on players' divergent prior beliefs regarding long-term success. The picture is consistent, however, with a view of war as a screening process. If some wars are the result of weak states risking wars (which sometimes materialize) to change the status quo—or if some status quo states bluff and are eventually challenged by resolved revisionists—then we would expect there to be an increase in the likelihood that wars end after just a few days of combat (Filson and Werner 2002; Powell 2006; Slantchev 2003b). On the other hand, longer wars may be produced by a different mechanism. In these cases, two truly resolved states find themselves fighting. For these wars, the duration is likely to be long. As a result, a data set generated by such a screening process, where some wars start by called bluffs and other wars are the result of deep conflicts of interest, would look much like the result I have presented here. In the region of zero to 30 or 40 battle days, the probability of a war terminating increases. After that, resolved belligerents are the only observations left in the data set, and the probability of termination decreases. Finally, at the extreme high end of the distribution of battle days, there is a large probability of termination since one side or the other is likely to be exhausted.

Robustness of Results

The validity of the inferences from my semiparametric analysis depends on the assumption that my model is well specified and that the findings are robust.[29] Although the specification hypothesis has face validity, here, I substantiate it further by considering some alternative specifications. First, I show that if we interact the *liberal war* variable with the *capabilities ratio* measure, some of the unusual results of my initial analysis became less pronounced, and the functional relationship between the capabilities measure and the probability of war termination can be partially explained. Second, I consider controlling for two other variables also found in the war termination literature, the *rates of loss* for attackers and defenders in a war. When I include these variables, our estimates are stable and very similar to our previous analysis. Next, I run our analysis and include both decade fixed effects and decade and war fixed effects. This analysis attempts to account for unobserved technological change in warfare and war-specific unobserved heterogeneity. Finally, I examine the quality of the semiparametric results by running the analysis on subsamples of the population of wars, excluding, in turn, World War II and the Arab-Israeli wars. Overall, we find that my results change remarkably little.

One of the most surprising results of the analysis of the CDB90 data is that there appears to exist a complex relationship between the relative power within dyads and the probability of war termination. Column 1 of table 4 explores this relationship. One possibility is that there is a difference between wars fought by different regime types (Clark and Reed 2003; Bennett and Stam 1998; Reiter and Stam 1998b; Bueno de Mesquita et al. 1999). To investigate this possibility, column 1 estimates the model with an interaction term; that is, I interact the indicator that a warring dyad contains a liberal democracy with the capabilities measure. As can be inferred from table 4, the relationship between the capabilities ratio for warring dyads with liberal countries is linear. In particular, as the dyad capabilities increasingly favor the initiator, the probability that the war terminates at a given battle increases with respect to the baseline relationship. For warring dyads without a liberal democracy, the relationship is closer to the *U*-shaped association expected by the hypothesis[30]; that is, battles are more likely to be terminating, all else equal, when there are asymmetries in the distribution of dyadic power. Yet, even in this analysis, there is a spike in the probability of a conflict-terminating event when the dyadic distribution of power is equal, which is consistent with the results in table 3. This relationship between the dyadic distribution of capabilities, in fact, holds in every specification, except with fixed effects for decades and wars. In this last case, the dyadic capabilities play no role until there is a capabilities advantage for the initiator, where the effect is to increase the probability of war termination.

The war termination literature suggests that a single battle's casualty rate might be relevant for explaining conflict termination (Bennett and Stam 1996; Slantchev 2004). For example, independent of the belligerents' expectations and the current trend in casualties, it may be possible that the attacker's and defender's rates of loss from a single battle could influence the decision to stop fighting. I thus construct a measure of each side's single battle rate of loss, and column 2 of table 4 reports the analysis in which I include the attacker and defender rates of loss.[31] The smooths (not shown) and the coefficient estimates are very close to those of the specifications in table 3, with the only difference being that the standard errors are systematically smaller and the *shocking victory* variable is more quadratic and statistically significant.

Columns 3 and 4 report the estimates, including, first, decade fixed effects, then war and decade fixed effects. Given that the nature and technology of war change over time, and wars are fought for a variety of objectives, including a dummy variable specification is a reasonable first cut at controlling for time and war-specific unobserved heterogeneity; that is, the inclusion of these fixed effects controls for time trends and allows for estimation of the within war effects of our variables. Column 3 of table 4 shows that including the decade fixed effects has no effect on our estimates. Column 4 shows that including both war and decade fixed effects does influence some estimates. In particular, the association between the dyadic balance of capabilities seems to be smoothed out and has no effect on war termination, until there is an approximate balance of power between the initiator and the target in a

Table 4
Additional Specifications for Probability of War Termination

	(1) Capabilities Liberal War Interaction	(2) Rates of Loss	(3) Decade Fixed Effects	(4) Decade and War Fixed Effects
Intercept	.212 (.046)***	.15 (.05)***	.177 (.123)	.106 (.126)
S (capabilities ratio)[a]	8.01 (57.74)***	9.46 (57)***	9.54 (46.92)***	2.92 (36.96)***
S (capabilities, by liberal war)	1 (10.94)***	–	–	–
S (battle days)[a]	5.28 (18.47)**	5.27 (22.41)***	5.17 (19.29)**	4.01 (14.10)
S (shocking victory)[a]	1.82 (12.04)	2.16 (18.93)**	1.70 (9.66)	1.87 (11.24)
Coercion	.125 (.078)	.18 (.08)**	.142 (.092)	.149 (.088)*
Military trend	−.007 (.003)**	−.01 (.004)***	−.008 (.004)*	−.005 (.003)
Democratic initiator	.235 (.072)***	.182 (.075)**	.226 (.11)**	.114 (.112)
Liberal war	−.214 (.045)***	−.184 (.044)***	−.165 (.050)***	−.074 (.050)
International mediation	−.045 (.050)	−.051 (.050)	−.066 (.12)	.124 (.52)
S (rate of loss A)	–	1.70 (20.2)**	–	–
S (rate of loss D)	–	1.70 (20.92)**	–	–
Decade dummies	No	No	Yes	Yes
War dummies	No	No	No	Yes
Adjusted R^2	.167	.186	.17	.181
Deviance explained (%)	21.3	24.0	23.5	24.3
GCV score	.08	.08	.09	.084
n	379	379	379	379

Note: The numbers reported for splines are the estimated (or effective) degrees of freedom. The estimated degree of freedom is the trace of the hat matrix and captures the effective degree of the spline. Reported next to the spline's estimated degrees of freedom is the χ^2_e statistic. The χ^2_e statistic is for the significance of the smooth compared to the null of no effect for the variable. It is distributed with degrees of freedom = e, the estimated degrees of freedom of the spline. It should be noted, however, that the generalized cross-validation fitting procedure (discussed in the appendix) used to select the smoothing parameter makes the χ^2 test only approximate and of lower power.
a. Estimated degrees of freedom and χ^2 statistics reported for smooth terms.
*$p < .1$. **$p < .05$. ***$p < .01$.

warring dyad. However, as the balance shifts in favor of the initiator, the probability of war termination, all else equal, increases.

The other relevant change in the analysis is that while the relationship between the number of battle days and the probability of termination stays the same, our estimate of the effect is less certain; that is, while I am able to reject the hypothesis that the true relationship between the number of battle days and the probability of war termination is zero at the .05 level in the baseline specification, the *p* value when the decade and war dummies are included is .114. This is not completely unexpected as the increased number of variables in the estimation is likely to increase the uncertainty surrounding our estimates, given a fixed number of observations.

Looking at our analysis, one might reasonably expect that many battle-specific variables may be correlated with the information flows—as I have measured them—and therefore may explain the patterns in the data; that is, whether a battle is informative about the distribution of relative power and the likelihood of success in a longer war may depend on the quality of leadership, whether best or second-rate forces were employed, or what terrain and logistical challenges may have been present. To address these questions, table 5 presents three additional specifications of my model with variables that control for these concerns.

Column 1 of table 5 adds variables related to troop quality, leadership, and available intelligence. The army coded troop quality in the CDB90 data set, and I use it here. Quality included the relative combat capabilities of the forces engaged, leadership quality, and intelligence quality. In each instance, these data were coded in a relative manner; that is, the army data coders tried to establish which side in a given battle had the advantage on each of these dimensions. The data produced a nine-point scale ranging from 0 (*no one favored*) to 4 (*very strongly favor the attacker*) for the defenders. In the model, there was no change in the variables of interest when including these measures of troop and leadership quality.

Column 2 of table 5 considers a model in which I directly control for the troop strength of each side in the battle, again with no significant change in the results. Here, relative troop strength is the ratio of the attacker's troops to those of the defender. Column 3 of table 5 displays the estimates from a model with controls for advantages with respect to terrain and logistics as well as troop strength. These estimates, like those of a fourth specification including all the battle variables and decade and war fixed effects (omitted from the table), also produced the same results as our original specification.

Finally, noting that the CDB90 data set has a large number of observations that are either warring dyads from World War II or warring dyads in the frequent Arab-Israeli conflicts, one might wonder if the reported results are driven by these two sets of conflicts. As a final check on the robustness of our baseline estimates, table 6 reports the results from the analysis on two subsets of the data. The first excludes any observations from World War II. The second excludes any observations from an Arab-Israeli war. The results are remarkably similar to those from the baseline model. All the smooths are of the same shape (not shown), and all smooths are at least as significant as in the baseline estimates. One thing that does change, however, is that the *democratic initiator* variable becomes insignificant, with a standard error larger than the estimated effect in the subsample without Arab-Israeli wars. This, too, is not surprising because a large proportion of the democratically initiated wars in the data are started by Israel.

In the end, the results from the baseline model appear to be generally robust to alternative specifications and to analysis on subsamples of the data. This is encouraging as it makes the previous findings more believable.

Conclusion

In this article, I set out to empirically investigate a number of war termination hypotheses using data from battles. Unlike previous work, where the focus has

Table 5
Estimates for Probability of War Termination with Battle Conditions

	(1) Troop, Leadership, and Intelligence Quality	(2) Troop Strength	(3) Strength, Logistic, and Terrain Advantages
Intercept	.180 (.046)***	.178 (.05)***	.132 (.051.)***
S (capabilities ratio)[a]	10.06 (62.44)***	9.21 (51.6)***	10.4 (67.4)***
S (battle days)[a]	5.55 (22.1)**	5.16 (20.6)**	5.4 (24.4)**
S (shocking victory)[a]	1.48 (11.1)	1.78 (9.55)	1 (.20)
Coercion	.121 (.094)	.09 (.07)	.148 (.094)
Military trend	−.011 (.003)***	−.010 (.003)**	−.011 (.004)***
Democratic initiator	.103 (.081)	.21 (.08)***	.145 (.083)*
Liberal war	−.205 (.045)***	−.193 (.045)***	−.20 (.044)***
International mediation	−.070 (.05)	−.045 (.054)	−.085 (.057)
Quality advantage (A)	.007 (.025)	–	–
Leadership advantage (A)	.036 (.027)	–	–
Intelligence advantage (A)	−.054 (.028)*	–	–
Relative troop strength	–	.01 (.015)	.012 (.014)
Logistic advantage (A)	–	–	.043 (.031)
Terrain advantage (A)	–	–	−.049 (.027)*
Adjusted R^2	.194	.152	.193
Deviance explained (%)	25.2	20.2	25
GCV score	.073	.084	.073
n	350	350	350

Note: The numbers reported for splines are the estimated (or effective) degrees of freedom. The estimated degree of freedom is the trace of the hat matrix and captures the effective degree of the spline. Reported next to the spline's estimated degrees of freedom is the χ^2_e statistic. The χ^2_e statistic is for the significance of the smooth compared to the null of no effect for the variable. It is distributed with degrees of freedom = *e*, the estimated degrees of freedom of the spline. It should be noted, however, that the generalized cross-validation fitting procedure (discussed in the appendix) used to select the smoothing parameter makes the χ^2 test only approximate and of lower power.
a. Estimated degrees of freedom and χ^2 statistics reported for smooth terms.
$*p < .1$. $**p < .05$. $***p < .01$.

been on wars or war years, using battle-level characteristics allows for the first direct exploration of the many conjectures and hypotheses surrounding war termination. I find support for some of my five hypotheses and less support for others. In particular, within my sample, I find little evidence that asymmetric wars tend to be more likely to end. So while there is little evidence that a preponderance of power systematically leads to higher probability of war termination, further investigation of the power variable did find that expected results, that is, that war termination is more likely when there are large asymmetries of power, hold when we do not consider wars with a liberal state. This is not too surprising, given the theoretical and empirical literature linking liberal regime characteristics and war outcomes

Table 6
Estimates for Probability of War Termination on Important Subsamples

	Without World War II	Without Arab-Israeli Wars
Intercept	.264 (.063)***	.127 (.051)**
S (capabilities ratio)[a]	7.54 (30.60)***	13.05 (80.92)***
S (battle days)[a]	4.65 (18.49)**	6.21 (26.45)***
S (shocking victory)[a]	1.95 (9.86)	1 (.926)
Liberal war	–.211 (.067)***	–.204 (.010)***
Military trend initiator	–.025 (.008)***	–.016 (.004)***
Democratic initiator	.179 (.096)*	–.117 (.215)
Coercion	.005 (.123)	.191 (.095)**
International mediation	.043 (.061)	–.057 (.215)
Adjusted R^2	.189	.252
Deviance explained (%)	28.5	31.4
GCV score	.112	.061
n	206	309

Note: The numbers reported for splines are the estimated (or effective) degrees of freedom. The estimated degree of freedom is the trace of the hat matrix and captures the effective degree of the spline. Reported next to the spline's estimated degrees of freedom is the χ^2_e statistic. The χ^2_e statistic is for the significance of the smooth compared to the null of no effect for the variable. It is distributed with degrees of freedom $= e$, the estimated degrees of freedom of the spline. It should be noted, however, that the generalized cross-validation fitting procedure (discussed in the appendix) used to select the smoothing parameter makes the χ^2 test only approximate and of lower power.
a. Estimated degrees of freedom reported for smooth terms.
$*p < .1$. $**p < .05$. $***p < .01$.

(Bueno de Mesquita et al. 1999; Lake 1992; Reiter and Stam 1998b).[32] In particular, the tendency of democracies to use their resources more efficiently in times of war and avoid economic distortions generated by authoritarian rule (Lake 1992) may be generating these results, but a direct test of this claim would require different data than contained in this analysis.

As for the coercion hypothesis, the results are mixed. The estimate recovered the expected effect, that large asymmetry in costs means the war is more likely to end, but the results were only marginally significant and have only small substantive effects. Similar results are found for the decisive battle or shocking victory hypotheses. The functional form of the relationship is in line with expectations; large departures from expectations increase the probability that war will end, but the estimate does not produce a statistical or substantively significant effect.

The informative battle hypothesis, however, finds less support. This is in contrast to some reasonably robust support for the screening model of war. This empirical finding reinforces observations by others that while incomplete information may explain early experiences in war, long conflicts have different explanations (Powell 2006; Fearon 2007). The empirical result presented previously provides further evidence for this claim and suggests that different theoretical approaches to the study of war may be valuable.

Finally, I have made an argument for the utility of semiparametric regression. The flexibility of this method, and the ease with which it is used, demonstrates that it may deserve attention by political scientists interested in doing empirical work. Also, given the flexibility and sorts of inference that can be made because of its few functional form assumptions, semiparametric methods can increase the utility of the conversation between theorists and empiricists.

Appendix: Model Specification and Generalized Cross-Validation Scoring

As is clear from the form of the roughness penalty functional, choosing the value of λ has critical implications for the resulting estimate of f. If the analyst has prior beliefs about the smoothness of the relationship between the response and predictor variables, values of λ could be chosen on those grounds. So, in some cases, the underlying theory may give insight into this choice.

On the other hand, such a method would, to some extent, undermine what we argue to be the advantage of nonparametric regression: the idea that when the functional form is unknown, we can ask the data what relationship exists. For such a method to be meaningful, there is a need for an automatic method whereby λ is chosen by the data.

The idea that underlies the cross-validation method of choosing λ is related to the idea that we, as analysts, are trying to provide the best prediction for the value of the response variable, given the estimated function, the value of x, and the objective function (equation [4])[33]; just as in a linear regression with mean zero errors, the true regression curve f produces the best mean square estimate of Y_i given a value of the predictor x_i.

Of course, in practical analysis, we do not have new data to check the prediction of our candidate function $\hat{f}$. However, since our focus is on the value of the smoothing parameter λ, we can construct a new data set that consists of the original data set minus some arbitrary observation i. We can then estimate the function $\hat{f}^{-i}$, which is the function that minimizes the objective function in equation (4) on the new subset of the original data, and ask, given λ, how well this model predicts Y_i given $\hat{f}^{-i}$ and x_i. In particular, if $f^{-i}(x_i, \lambda)$ is too smooth, then anywhere there should be additional curvature, the difference between the predicted and actual values of Y_i will be large. Since we perform this procedure with any i, a measure of overall predictive success for a given amount of smoothing (i.e., a given λ) is

$$CV(\lambda) = n^{-1} \sum_{i=1}^{n} (Y_i - [\beta_0 + \hat{f}^{-i}(\mathbf{x}_i; \lambda)])^2, \tag{A1}$$

(continued)

Appendix (continued)

where $\hat{f}^{-i}(\mathbf{x}_i;\lambda)$ is the smooth function evaluated on the data without the ith observation.

Now, obviously, a number of issues arise. First is that there are infinitely many values of λ to check. Moreover, there is no guarantee that the minimizer of the cross-validation equation is unique, so for both of these reasons, a grid search method is the most effective way to find the optimal λ.

As may be apparent to the reader, the cross-validation method is computationally burdensome. To deal with the computational burden of the cross-validation procedure, an alternative method of determining the appropriate smoothness for the function $\hat{f}$, called generalized cross-validation (GCV), is used in practice.[34]

The idea behind GCV is the same as the idea that motivates cross-validation. Equation (A2) gives the formula used in the GCV procedure:

$$\mathrm{GCV}(\lambda) = n^{-1}\frac{\sum_{i=1}^{n}(Y_i - [\beta_0 + \hat{f}^{-i}(\mathbf{x}_i;\lambda)])^2}{1 - tr\mathbf{A}(\lambda)\cdot n^{-1}}, \tag{A2}$$

where $tr\mathbf{A}(\lambda)$ is the sum of the diagonal elements of the hat matrix, which generates the equivalent degrees of freedom for the spline given a smoothing parameter λ.

So, as noted by Green and Silverman (1994), the GCV value is equivalent to

$$\mathrm{GCV}(\lambda) = \frac{\text{residual sum of squares}}{(\text{estimated degrees of freedom})^2}. \tag{A3}$$

The technical details showing the relationship between the cross-validation and GCV methods are not presented here, but details can be found in Green and Silverman (1994). For our purposes, it is sufficient to note that the smoothing parameter, λ, is chosen such that it minimizes the GCV score, based on the same logic that drove us to chose the smoothing parameter that minimized the cross-validation score.

It is worth noting here that the GCV score provides a measure of fit and in and of itself does not provide a statistical test, like, for example, a t statistic or F test that allows the analyst to reject hypotheses about linearity. However, when the GCV method is used to select an appropriately smooth curve under the roughness penalty approach, resulting confidence bands and approximate significance statistics for the spline do speak to the likelihood that a linear fit is appropriate. As to the comparison of the fit of a smoothed regression and a linear version of the same model, this can also be easily accomplished with an F test by an analysis of variance procedure because the linear model is a nested specification of the smoothing spline (i.e., a linear specification has $edf = 1$).

While Beck and Jackman (1998) do not promote the use of generalized cross-validation or similar methods, there are good reasons to use this approach. If we consider what would be a desirable criterion for determining the appropriate

(continued)

Appendix (continued)

amount of smoothness, there will be a trade-off between the bias and the variance of our estimate. One natural quantity of interest is the integrated squared prediction error (EPE). To see this, note that

$$\mathrm{EPE}(\hat{f}_\lambda) = E[(Y - \hat{f}_\lambda(X))^2], \tag{A4}$$

$$= \mathrm{Var}(Y) + E[\mathrm{Bias}^2(\hat{f}_\lambda(X)) + \mathrm{Var}(\hat{f}_\lambda(X))], \tag{A5}$$

$$= \sigma^2 + \mathrm{MSE}(\hat{f}_\lambda). \tag{A6}$$

Remarkably, the cross-validation score is an approximately unbiased estimate of the EPE, is minimized at the same value, and provides a natural means for data-driven selection of the amount of smoothness.

Notes

1. Many sources to which one might turn for more battle data to augment CDB90, like the battle dictionaries of Perrett (1996) or Laffin (1995), are of lower quality and contain less detail. So in the absence of a large-scale data collection effort from historical narratives and primary sources, there is little that can be done to fix this data set.

2. See, for example, discussions by Brooks (2003), Desch (2002), and Measheimer (1989).

3. On the other hand, if battles are about learning about the ability of an opponent to take and control territory by force, as our theories suggest, the performance of the army may be more important than naval engagements.

4. For a nice, formal model that explicitly deals with the interaction of strategically manipulable acts and nonmanipulable acts in the context of war, see Slantchev (2003b).

5. The capabilities ratio of the initiator is just (initiator's capabilities)/(total capabilities). Because this is a measure of a dyad characteristic, the choice of the initiator's ratio is without loss of generality.

6. See also Slantchev (2003a).

7. Filson and Werner have begun to do work linking sensitivity to costs (2004) and sensitivity to loss (2007) that predicts important regime effects on the relationship between casualties and outcomes. I leave the detailed exploration of these effects to future work.

8. It is worth noting that A. Smith and Stam's (2004) lemma 1 is an asymptotic result about convergence in distribution of the players' beliefs. In fact, a slightly stronger result is easily proven: assume that countries are playing the game described by A. Smith and Stam (2004). Then the difference in the players' expected values *converge monotonically* to zero; that is, in their model, two countries' beliefs are never further apart than at the start of a war. If we then believe that across wars, the costs of fighting are randomly distributed, we would get a result for the Smith and Stam article similar to that in Filson and Werner (2002, result 5). See the author for the simple proof.

9. It is this mechanism that Powell (2006) conjectures accounts for the variation in the length of wars, and I test that claim here. For a leading explanation of why war may be preferred, even without asymmetric information, see Goemans (2000).

10. Similarly, Vietnam is recorded as having only one battle because much of the military action was unconventional. Since this is a single observation and has almost no effect on the analysis, I drop the Vietnam War from my data set so as not to imply that information from this ten-year war is central to the inferences made here. We also drop the single observation from the Spanish Civil War.

11. I would note that a different selection effect is *not* relevant for this analysis. It is not the case that because I do not have observations on wars that did not occur as a consequence of countries' expectations about the variables in the model, somehow my estimates are wrong. First, the hypotheses I test are based on the conditional statement that a war has occurred—otherwise, war termination is not an issue. Second, the population equation by which I estimate the battle data coefficients for wars that did and did not occur has no intuitive interpretation and is of no real interest. Selection would be an issue, however, if I were to use the CDB90 data to study decisions about war initiation.

12. I would like to note that the purging and correction of this data set were done by Stephen Biddle and Christopher Hamner, and I thank them for allowing us to use this much more reliable data.

13. The analysis was also run with the following coded as no terminating events: United States–Japan Pacific War (Okinawa), United States–Japan (Malaya), Germany-France (World War II), Suomossalmi (Winter War), and Mukden (Russo-Japanese War).

14. If a battle crosses a calendar year, the CINC score associated with that observation is from the year the battle begins.

15. Histograms of data are available from the Web appendix.

16. For an interesting set of additional hypotheses linking regimes to war termination, see Filson and Werner (2004, 2007).

17. Note that by the result in note 7, independent of the battle outcomes, public data drive players' beliefs toward each other's. Thus, an effect of trends is a distinct mechanism in practice and theory.

18. In particular, the functional form specification problem will be particularly severe if our data are generated by strategic interactions. For a detailed treatment of this issue, see Signorino and Yilmaz (2003).

19. We could also use a duration model, but like ordinary least squares regression, many of these models also make restrictive functional form assumptions (i.e., exponential, Weibull lognormal and loglogistic models) that are inappropriate for strategically generated data. A Cox semiparametric regression, on the other hand, allows for nonparametric estimation of the hazard rate but relies on proportionality assumptions when estimating the conditional effect of the *x*s on the probability of termination.

20. The reader may notice that the choice of λ (the smoothing parameter) is a crucial part of the analysis. Exactly how this is done is discussed in the appendix.

21. For detailed discussion of various smoothing spline methods and their strengths and weaknesses, see Wood (2006, chapter 4) or Hastie, Tibshirani, and Friedman (2001).

22. All the analysis in this article is done in the mgcv package in R; see Wood (2001) for details.

23. The χ^2_e statistic is for the significance of the smooth compared to the null of no effect for the variable. It is distributed with degrees of freedom $= e$, the estimated degrees of freedom of the spline. It should be noted that the generalized cross-validation fitting procedure (discussed in the appendix) makes the χ^2 test only approximate and is of lower power.

24. Because the numerical process of estimating a semiparametric model can be complex, we want to be sure that we have specified a well-fitting model. To do this, I checked the properties of the GCV. I was able to verify that the minimization did not occur at a corner solution, as the gradient at minimization was zero. I also confirmed that the Hessian of the GCV/unbiased risk estimator is positive definite so that we have a minimum.

25. Moreover, this highly nonlinear fit is consistent across many specifications. In the robustness section that follows, we see that there may be some interaction with regime type that may be important.

26. This seems to be true given the duration analysis of Bennett and Stam (1996) and Slantchev (2004).

27. Note that there are very few observations at the lowest extreme of the capabilities ratio measure, and we should be skeptical of the dip in the smooth near zero.

28. Results from this specification can be obtained by running the replication file.

29. For discussion of some empirical issues surrounding semiparametric model specification, see McCracken (2004).

30. Smooth figures can be found in the Web appendix.

31. If we think of the cost of battle as being the casualty count, A. Smith and Stam's (2004) corollary to proposition 2 also implies this effect.

32. What needs further investigation, however, is why this liberal democracy effect seems to interact with the effect of capabilities in such a way.

33. This discussion follows that found in chapter 3 of Green and Silverman (1994), which provides an intuitive introduction to the idea of nonparametric regression using the roughness penalty approach.

34. For more details on the properties of the cross-validation procedure and the relationship between cross-validation and GCV, see Green and Silverman (1994) or Loader (1999). For a more technical treatment, see Craven and Wahba (1979).

References

Beck, Nathaniel, and Simon Jackman. 1998. Beyond linearity by default: Generalized additive models. *American Journal of Political Science* 42:596–627.

Bennett, D. Scott, and Allan C. Stam. 1996. The duration of interstate wars, 1816–1985. *American Political Science Review* 90:647–62.

———. 1998. The declining advantage of democracy: A combined model of war outcomes and duration. *Journal of Conflict Resolution* 42:344–66.

———. 2004. *The behavioral origins of war*. Ann Arbor: University of Michigan Press.

Biddle, Stephen. 2001. Rebuilding the foundations of offense-defense theory. *Journal of Politics* 63:741–74.

Biddle, Stephen, and Stephen Long. 2004. Democracy and military effectiveness. *Journal of Conflict Resolution* 48:525–46.

Blainey, Geoffrey. 1988. *The causes of war*. New York: Free Press.

Brooks, Risa A. 2003. Making military might: Why do states fail and succeed? *International Security* 28:149–91.

Bueno de Mesquita, Bruce, James D. Morrow, Randolph M. Siverson, and Alastair Smith. 1999. An institutional explanation of the democratic peace. *American Political Science Review* 93:791–807.

Clark, David H., and William Reed. 2003. A unified model of war onset and outcome. *Journal of Politics* 65:69–91.

Clausewitz, Carl von. 1832/1984. *On war*. Princeton, NJ: Princeton University Press.

Craven, P., and G. Wahba. 1979. Smoothing noisy data with spline functions. *Numerical Mathematics* 31:377–90.

Davis, Paul K. 2001. *Decisive battles: From ancient times to the present*. New York: Oxford University Press.

Desch, Michael C. 2002. Democracy and victory: Why regime type hardly matters. *International Security* 27:5–47.

Doyle, Michael. 1986. Liberalism and world politics. *American Political Science Review* 80:1151–69.

Dupuy, T. N. 1979. *Numbers, predictions and war*. Fairfax, VA: HERO Books.

———. 1995. *Attrition: Forecasting battle casualties and equipment loss in modern war*. Falls Church, VA: NOVA.

Fearon, James D. 2007. Fighting rather than bargaining. Typescript, Stanford University, Stanford, CA.

Fey, Mark, and Kristopher W. Ramsay. 2007. Mutual optimism and war. *American Journal of Political Science* 51:738–54.

Filson, Darren, and Suzanne Werner. 2002. A bargaining model of war and peace: Anticipating the onset, duration and outcome of war. *American Journal of Political Science* 46:819–37.

———. 2004. Bargaining and fighting: The impact of regime type on war onset, duration, and outcomes. *American Journal of Political Science* 48:296–313.

———. 2007. Sensitivity to cost of fighting versus sensitivity to losing the conflict. *Journal Conflict Resolution* 51:691–714.

Gartner, Scott Sigmund. 1997. *Strategic assessment in war*. New Haven, CT: Yale University Press.

Goemans, Hein E. 2000. *War and punishment: The causes of war termination and the First World War*. Princeton, NJ: Princeton University Press.

Green, P. J., and B. W. Silverman. 1994. *Nonparametric regression and generalized linear models: A roughness penalty approach*. New York: Chapman and Hall.

Hastie, Trevor J., Robert J. Tibshirani, and Jerome Friedman. 2001. *The elements of statistical learning: Data mining, inference, and prediction*. New York: Springer.

Hobbs, Richard. 1979. *The myth of victory: What is victory in war?* Boulder, CO: Westview Press.

Laffin, John. 1995. *Brassey's battles: 3,500 years of conflict, campaigns and wars from A–Z.* Washington, DC: Brassey's.

Lake, David A. 1992. Powerful pacifists: Democratic states and war. *American Political Science Review* 86:24–37.

Loader, Clive. 1999. *Local regression and likelihood*. New York: Springer.

McCracken, Marti L. 2004. *Modeling a very rare event to estimate sea turtle bycatch: Lessons learned*. NOAA Tech. Memo. NMFS-PIFSC-3. Honolulu, HI: NOAA.

Measheimer, John J. 1989. Assessing the conventional balance: The 3:1 rule and its critics. *International Security* 13:54–89.

Organski, A. F. K., and Jacek Kugler. 1980. *The war ledger*. Chicago: University of Chicago Press.

Perrett, Bryan. 1996. *The battle book: Crucial conflicts in history from 1469 BC to the present*. New York: Sterling.

Powell, Robert. 2004. Bargaining and learning while fighting. *American Journal Political Science* 48:344–61.

———. 2006. War as a commitment problem. *International Organization* 60:169–203.

Reiter, Dan, and Allan C. Stam. 1998a. Democracy and battlefield military effectiveness. *Journal of Conflict Resolution* 42:259–77.

———. 1998b. Democracy, war initiation, and victory. *American Political Science Review* 92:377–89.

———. 2002. *Democracies at war*. Princeton, NJ: Princeton University Press.

Schelling, Thomas. 1960. *The strategy of conflict*. Cambridge, MA: Harvard University Press.

———. 1966. *Arms and influence*. New Haven, CT: Yale University Press.

Signorino, Curtis, and Kuzey Yilmaz. 2003. Strategic misspecifcation in regression models. *American Journal of Political Science* 47:551–66.

Singer, J. D., S. Bremer, and J. Stucky. 1972. Capability distribution, uncertainty, and major power war, 1820–1965. In *Peace, war, and number*, ed. B. Russett, 19–48. Beverly Hills, CA: Sage.

Slantchev, Branislav L. 2003a. The power to hurt: Costly conflict with completely informed states. *American Political Science Review* 47:123–33.

———. 2003b. The principle of convergence in wartime negotiation. *American Political Science Review* 97:621–32.

———. 2004. Initiators, duration, and the settlement of interstate wars, 1816–1991. *American Journal of Political Science* 48:813–29.

Smith, Alastair. 1998. Fighting battles, winning wars. *Journal of Conflict Resolution* 42:301–20.

Smith, Alastair, and Allen C. Stam. 2004. Bargaining and the nature of war. *Journal of Conflict Resolution* 48:783–813.

Smith, James D. 1995. *Stopping wars: Defining the obstacles to cease-fire*. Boulder, CO: Westview Press.

Wagner, R. Harrison. 1994. Peace, war and the balance of power. *American Political Science Review* 88:593–607.

———. 2000. Bargaining and war. *American Journal of Political Science* 44:469–84.

Weigley, Russell F. 1991. *The age of battles: The quest for decisive warfare from Breitenfeld to Waterloo*. Bloomington: Indiana University Press.

Werner, Suzanne. 1998. Negotiating the terms of settlement: War aims and bargaining leverage. *Journal of Conflict Resolution* 42:321–43.

Werner, Suzanne, and Amy Yuen. 2005. Making and keeping peace. *International Organization* 59:261–92.

Wittman, Donald. 1979. How wars end. *Journal of Conflict Resolution* 23:743–63.

Wood, Simon N. 2001. mgcv:GAMs and generalized ridge regression for R. *R News* 1:20–25.

———. 2006. *Generalized additive models: An introduction with R*. Boca Raton, FL: Chapman and Hall.