The Illusion of Democratic Credibility

Alexander B. Downes and Todd S. Sechser

Abstract   Do democracies make more effective coercive threats? An influential literature in international relations argues that democratic institutions allow leaders to credibly signal their resolve in crises, thereby making their threats more likely to work than threats by nondemocracies. This article revisits the quantitative evidence for this proposition, which we call the "democratic credibility hypothesis," and finds that it is surprisingly weak. Close examination of the data sets most commonly used to test this hypothesis reveals that they contain few successful democratic threats, or indeed threats of any kind. Moreover, these data sets' outcome variables do not properly measure the effectiveness of threats, and therefore yield misleading results. The article then reassesses the democratic credibility hypothesis using the Militarized Compellent Threats data set, a new data set designed specifically to test hypotheses about the effectiveness of coercive threats. The analysis indicates that threats from democracies are no more successful than threats from other states.

It is commonly argued in international relations scholarship that coercive threats issued by democracies are more successful than those issued by nondemocracies.[1] According to this view, democratic leaders who issue threats in crises and then

1. The literature on democratic credibility traces back at least to 1960, when Schelling suggested that democracies could improve their bargaining leverage by marshaling public opinion in support of a particular negotiating stance, thereby constraining their ability to make concessions to adversaries. Numerous scholars have since extended Schelling's logic; see, for example, Fearon 1994; Schultz 1998 and 2001; Smith 1998; Guisinger and Smith 2002; Ramsay 2004; Slantchev 2006; Fang 2008; and Tarar and Leventoğlu 2009.

back down can be punished at the ballot box for tarnishing the nation's honor, revealing their incompetence, or damaging the country's reputation for honesty.[2] Leaders facing such punishment therefore are likely to make threats only when they are willing to escalate and are less likely to back down after issuing a threat. The advantage of a democratic system, the theory argues, lies in its transparency: adversaries can observe a democratic leader's political incentives and realize when the leader is willing to fight.[3] By contrast, the opaque nature of nondemocratic regimes prevents leaders from credibly revealing their resolve.[4] Thus, when democracies issue threats during crises, those threats are more likely to be taken seriously and more likely to persuade an opponent to back down.

This proposition, which we broadly term the "democratic credibility hypothesis,"[5] has proven enormously influential in recent international relations scholarship.[6] According to the Social Sciences Citation Index, books and articles advocating this hypothesis have been cited more than 1,000 times. In part, this is because empirical research appears to confirm the theory's central prediction that threats issued by democracies are more likely to succeed. Partell, Rioux, and Partell and Palmer, for example, all found that democracies tend to prevail more often than nondemocracies in international disputes.[7] Gelpi and Griesdorf showed that democracies tend to win international crises when they demonstrate their resolve through militarized actions, supporting the notion that democratic leaders can credibly tie their hands through public commitments.[8] Perhaps most significantly, Schultz's influential book *Democracy and Coercive Diplomacy* found that when democracies initiate militarized disputes, they are less likely to meet resistance from adversaries.[9] Other scholars continue to find evidence that democracies achieve more favorable outcomes in international disputes.[10]

---

2. Schultz 2001 offers an important caveat to this claim, arguing that democratic leaders will be punished for bluffing only if opposition parties also opposed the threat.

3. Some scholars (for example, Guisinger and Smith 2002) argue that democratic leaders are more vulnerable to punishment than nondemocratic leaders, but others—most notably Fearon 1994 and Schultz 2001—do not make this claim. They argue that democracies are unique not because they punish leaders for backing down, but because their punishment mechanisms are institutionalized and therefore observable to outsiders.

4. An exception is Weeks 2008, who maintains that some autocracies can also signal their resolve through domestic political mechanisms. However, her study nonetheless confirms the finding that, on the whole, democracies make more effective threats than nondemocracies.

5. This notion is sometimes labeled the "audience costs hypothesis," reflecting the central mechanism at work in most versions of the theory. Because our interest is in the outcomes of democratic threats rather than causal mechanisms, however, we use the broader term "democratic credibility hypothesis."

6. Some scholars argue that democracies make more credible commitments in other international arenas as well (for example, Martin 2000; Mansfield, Milner, and Rosendorff 2002; Lipson 2003).

7. See Partell 1997; Rioux 1998; Partell and Palmer 1999.

8. Gelpi and Griesdorf 2001.

9. Schultz 2001. See also Schultz 1999, which reported similar findings.

10. See Prins 2003; Sullivan and Gartner 2006; Clare 2007; and Haynes forthcoming. Some early studies also evaluated the hypothesis that crises involving democracies are less likely to escalate. See Eyerman and Hart 1996; and Partell and Palmer 1999.

In this article we reassess the empirical evidence for the democratic credibility hypothesis and find that it is significantly weaker than the conventional view asserts. First, we investigate the quantitative data sets most commonly used in tests of this proposition—namely, the Militarized Interstate Dispute (MID) and the International Crisis Behavior (ICB) data sets—and find that most apparent democratic "victories" in these data sets are not actually successful threats. Indeed, barely 10 percent of the cases in the MID data set and fewer than 17 percent of the cases in the ICB data set contain coercive threats at all. Instead, most cases in these data sets entail minor military skirmishes, border and airspace violations, fishing boat incidents, and other events in which the participants did not actually make any demands. These cases reveal little about the conditions under which target states are likely to submit to a challenger's coercive threats. Furthermore, the data sets' outcome variables are poor indicators of successful coercive threats, primarily because they do not differentiate crisis victories achieved by brute force from those achieved via coercive diplomacy. This distinction is critical because the democratic credibility hypothesis argues that democracies are better able to prevail in crises without having to resort to decisive force. Schultz, for example, argues that "democratic states have in general been quite successful at using threats to get their way in international disputes and to do so without actually waging war."[11] Equating military victories with successful threats creates the misleading impression that coercive threats by democracies succeed at a high rate, when in fact they do not. In short, the democratic credibility proposition rests on a shaky empirical foundation.

Second, we reassess the hypothesis using the Militarized Compellent Threats (MCT) data set, a new collection of more than 200 compellent threats issued between 1918 and 2001.[12] This data set is a more appropriate set of cases for testing the efficacy of democratic threats, both because its unit of analysis is the coercive threat and because the data set explicitly measures the degree of compliance with coercive demands. This analysis yields no support for the claim that threats by democracies are more effective.

This article contributes to an emerging skepticism about the notion that democracies enjoy special advantages in international crises. Several recent studies have contested the view that democratic institutions confer a unique bargaining advantage on leaders during disputes.[13] Yet while these studies offer compelling reasons to question the logic of the democratic credibility hypothesis, their potency has been limited by the apparent weight of quantitative evidence in favor of the theory. Our analysis contributes to this debate by showing that much of the theory's empir-

---

11. Schultz 2001, 2. See also Fearon 1994; Smith 1998; Partell and Palmer 1999; and Guisinger and Smith 2002.
12. Sechser 2011.
13. See, for example, Finel and Lord 1999; Rosato 2003; Weeks 2008; Snyder and Borghard 2011; Brown and Marcum 2011; and Trachtenberg 2012.

ical support does not, in fact, survive close scrutiny.[14] Our findings also carry broader implications for the study of coercive diplomacy, suggesting that two of the most commonly used data sets in international conflict research are largely inappropriate for studying coercion in international relations.

We proceed as follows. First, we examine the MID and ICB data sets, demonstrating both that they contain few successful democratic threats and that they fail to accurately capture the efficacy of those threats. Second, we employ the MCT data set to conduct a new empirical test of the theory, finding little support for the notion of democratic credibility. We then discuss the limitations of our analysis and suggest several implications for future research.

Existing Evidence for Democratic Credibility

Empirical tests of the democratic credibility proposition have relied overwhelmingly on two data sources: the MID and the ICB data sets.[15] The data sets' broad geographic and temporal scope, large number of observations, and inclusion of conflict episodes that did not escalate to war have made them a prime source of data for scholars studying democratic threats as well as a wide array of other research questions.[16]

We argue, however, that these two data sets are unable to provide empirical support for the democratic credibility hypothesis. The reasons are twofold. First, the majority of cases in these data sets involve episodes in which no coercive threats were actually issued. Second, even when threats appear in the data sets, their outcomes are often coded incorrectly. Consequently, the evidence for the democratic credibility hypothesis in studies using these data sets is considerably weaker than such studies suggest.

A Scarcity of Threats

The first problem with the MID and ICB data sets is that most of their cases do not involve coercive threats. This poses a serious problem for evaluating the dem-

---

14. The importance of "negative" findings for the advancement of scientific knowledge has long been recognized (for example, Popper 1959; Kuhn 1962). King, Keohane, and Verba 1994 emphasize that "a negative result is as useful as a positive one; both can provide just as much information about the world" (105). Indeed, negative results have played a prominent role in recent research about democracies in international relations, prompting debates about democracies' military effectiveness (Reiter and Stam 2002; Desch 2008; and Downes 2009), counterinsurgency capabilities (Merom 2003; and Lyall 2010), and the war-proneness of democratizing states (Mansfield and Snyder 2005; and Narang and Nelson 2009).

15. Among studies of the democratic credibility hypothesis, Partell and Palmer 1999; Schultz 1999 and 2001; Prins 2003; and Sullivan and Gartner 2006 test the theory with the MID data set, whereas Rioux 1998; and Gelpi and Griesdorf 2001 use the ICB archive. Partell 1997 uses both data sets.

16. The MID data set contains information about 2,332 disputes that occurred between 1816 and 2001 (Jones, Bremer, and Singer 1996), and the ICB data set describes 455 interstate crises during the period 1918 to 2007 (Brecher and Wilkenfeld 1997).

ocratic credibility hypothesis, because the theory's chief empirical claim is that democratic governance impacts the effectiveness of military threats. The key question of Schultz's study of democratic credibility, for example, is "whether and when democratic governments can effectively use threats of force to prevail in international crises."[17] Testing this hypothesis requires a data set of coercive threats so that threats issued by democracies and nondemocracies can be compared. We find that the MID and ICB data sets do not meet this condition.

Broadly speaking, coercive threats may take one of two basic forms. *Deterrent threats* aim to dissuade targets from taking a particular course of action, whereas *compellent threats* are designed to persuade a target to change the status quo.[18] In both cases, military force is threatened (often implicitly) as punishment for failing to comply with a substantive demand.

To evaluate the prevalence of deterrent and compellent threats in the MID and ICB data sets, we consulted the data sets' narrative case descriptions and looked for cases in which coercive demands were coupled with threats of military force. Although the MID data set does not provide summaries for all of its cases, the Correlates of War 2 Project has constructed the MID Narratives document, which contains written synopses of most MID cases between 1992 and 2001.[19] These narratives, while not intended as complete descriptions, nevertheless offer an illuminating look at 294 MIDs—roughly 12 percent of the cases in the MID data set. The ICB data set, by contrast, provides summaries and documentation for all 455 of its crises.[20]

In general, we adopted a lenient standard for identifying threats, acknowledging that threats to use force during crises are often implied—and may generate binding commitments—even when they are not explicitly stated. Thus, military mobilizations or maneuvers were classified as threats to use force according to our criteria as long as the historical record indicated that these actions were designed to support a coercive demand.

Our analysis found surprisingly few cases in either data set containing demands, threats, or ultimata. Instead, we found that the majority of MID and ICB cases fall into roughly five categories, none of which involve coercive threats (see Table 1).

1. **Minor skirmishes and border violations**. One common type of case involves small-scale uses of force that carry no apparent threat or demand. Often, such incidents involve the actions of a few rogue soldiers: for example, MID 4090

---

17. Schultz 2001, 1. Other examples include Fearon 1994, 577; Smith 1998, 633; Guisinger and Smith 2002; and Tomz 2007, 822, all of whom explicitly characterize the democratic credibility hypothesis as a theory about the effectiveness of threats.

18. Schelling 1966 provides a thorough discussion of the distinction between these two types of coercive threats; see also Art and Cronin 2003; and Sechser 2010. We follow Schelling 1966, 71, in using the term "coercive" to encompass both deterrent and compellent objectives.

19. Correlates of War 2 Project 2004.

20. We evaluated the ICB data set at the dyadic level using the actor-level version of the data set, which divides the 455 crises into exactly 1,000 crisis dyads.

represents a 1997 incident in which four Greek soldiers fired on a group of Albanian villagers. Other cases involve shots fired at trespassers, such as a 1999 incident in which Chinese troops shot a Mongolian national who had crossed the border (MID 4178). Airspace violations (real or alleged) also fall into this category: in 1995, for example, Russian helicopters made unauthorized flights over Lithuanian residential areas, prompting a protest from Lithuania's foreign minister (MID 4105). Finally, many cases in this category involve minor raids or surprise attacks in which no coercive diplomacy took place. Israel's 1981 attack against Iraq's Osirak nuclear reactor, for example, is included in both data sets (MID 3101; ICB 324) even though Israel made no demands of the Iraqi government. Although military actions undoubtedly are sometimes used to communicate threats nonverbally, the cases in this category appear to entail no coercive demands. The democratic credibility hypothesis therefore makes no prediction about them because there was no demand for the target to accept.

TABLE 1. *Threats and nonthreats in the MID Narratives and ICB archive*

	MID Narratives		ICB	
Case type	Number of cases	% of total	Number of cases	% of total
Minor skirmishes and border violations	144	49.0	182	18.2
Alerts and exercises	67	22.8	110	11.0
Maritime incidents	37	12.6	2	0.2
Wars and wartime campaigns	4	1.4	192	19.2
Nonmilitarized episodes	12	4.1	349	34.9
Deterrent warnings	17	5.8	49	4.9
Compellent threats	13	4.4	116	11.6
Total	294	100.0	1,000	100.0

2. **Alerts and exercises**. The MID and ICB data sets also record dozens of noncoercive military maneuvers. For example, in August 1999, Japan placed several Aegis cruisers on alert in anticipation of a North Korean ballistic missile test (MID 4322). The vessels, however, were deployed to monitor—not deter—the missile test. Similarly, in 1996, India placed border security forces on alert amid concerns about possible refugee flows from Bangladesh (MID 4006), but the Indian government made no demands. An analogous example from the ICB data set is the 1937 "Postage Stamp Crisis" (ICB 57), in which a Nicaraguan stamp portrayed the country's borders as encompassing a large

chunk of neighboring Honduras. Honduras reportedly reinforced several border garrisons in protest, but it neither threatened war over the stamp nor sought to deter an invasion. It would thus be misleading to conclude that Honduran actions represented a coercive threat.

3. **Maritime incidents**. A third class of episodes involves encounters between coast guard vessels and fishing trawlers, cargo ships, or passenger vessels. For example, between 1993 and 1995 Russian patrol boats chased Japanese fishing vessels, arrested several fishermen, and even sank one boat near the disputed Kurile Islands (MID 4042). Similarly, MID 4019 comprises a brief 1995 episode in which a North Korean patrol boat fired on a Chinese fishing vessel. These incidents, however, are of dubious utility for testing hypotheses about democratic credibility because the primary actors were not political leaders but rather ship captains and private fishermen.

4. **Interstate wars and wartime campaigns**. A fourth type of case includes the onset of interstate wars and military campaigns conducted during ongoing wars. Several disputes in the MID Narratives and ICB data set correspond to wars that were not preceded by coercive demands, such as the war of Bosnian independence (MID 3556 and 3557) and the surprise North Korean invasion of South Korea in June 1950 (ICB 132). The ICB data set also includes several wartime campaigns, including significant World War II battles like Stalingrad (ICB 89), D-Day (ICB 94), and Iwo Jima (ICB 101).[21] Intrawar engagements, however, shed no light on the democratic credibility hypothesis: the dependent variable in the theory is the effectiveness of coercive threats, not the effectiveness of battlefield operations.

5. **Nonmilitarized episodes**. A surprising number of cases in both data sets consist of events that do not qualify as interstate conflict episodes, either because the use of force was never at issue or because one disputant was a nongovernmental actor rather than the state listed. Cases in this category often involve states that found themselves on opposing sides of a crisis but had no direct encounters. For instance, in the 1964 Congolese hostage crisis (ICB 211), U.S. aircraft ferried Belgian paratroopers on a successful mission to rescue hostages held by Soviet-supported rebels. The ICB data set includes U.S.–Soviet and Soviet–U.S. dyads in the crisis, but the two states had little interaction during the episode and did not attempt to coerce one another. Other examples include cases in which a state is listed as a primary disputant, but the actual target was a rebel group. For instance, Rhodesia and South Africa precipitated several ICB crises by launching raids against insurgent groups in neighboring states (ICB 300, 302, 339). Although these cases involved

---

21. To their credit, some studies of democratic credibility (for example, Gelpi and Griesdorf 2001, 646) drop several such intrawar crises from their analyses, although they retain other problematic cases.

direct combat, we classify them as nonmilitarized because the initiators used force against rebel groups rather than opposing states.

These five types of cases cannot support inferences about the credibility of democratic threats for at least three reasons. First, the democratic credibility hypothesis aims to explain the success and failure of coercive threats—it makes no predictions about cases in which coercive diplomacy was not attempted. Second, the target in many disputes was a nongovernmental actor (for example, a rebel group or fishing boat), calling into question these data sets' utility for testing theories about state behavior. Third, the militarized actions in these disputes often were not explicitly authorized by state leaders. Many MIDs, for example, center solely on the actions of individual border guards, soldiers, or fighter pilots, which are not relevant for testing theories of national-level decision making. Statistical correlations within these data therefore reveal little about the ability of democratic institutions to bolster the credibility of coercive threats.

Table 1 classifies the 294 cases described in the MID Narratives document and the 1,000 dyads in ICB into the five categories described above. Additionally, the table reports the number of cases that involve deterrent or compellent threats. Overall, only a small proportion of MID and ICB cases involve coercive threats. Cases with no apparent demands—whether explicit or implied—make up nearly 90 percent of all observations in the MID Narratives archive and 83.5 percent of observations in ICB. Deterrent and compellent threats, in contrast, comprise barely 10 percent of MID observations and less than 17 percent of ICB cases. In other words, roughly one in ten MIDs between 1992 and 2001 and one in six ICB crisis dyads are appropriate for testing the democratic credibility hypothesis.[22]

Coercive threats by democracies are also rare in these data sets. Of the 294 total disputes in the MID Narratives collection, just 15 (5.4 percent) contain threats issued by democracies.[23] Similarly, of the 1,000 crisis dyads in the ICB data set, only 58 (5.8 percent) involve democratic threats. The MID and ICB data sets, in other words, are populated by far fewer democratic threats than the literature on democratic credibility would suggest.

Inappropriate Outcome Variables

A second problem with the MID and ICB data sets is that they were not designed to detect whether coercive threats succeed or fail. Although variables in both data sets contain information about crisis outcomes, a closer look suggests that they are dubious proxies for the effectiveness of coercive threats.

---

22. Because the MID data set does not offer synopses for observations prior to 1992, considerable independent research would be required to determine which of the remaining 2,038 cases contain threats or demands.

23. We adopt the standard Polity definition here, classifying states with POLITY scores of +6 or higher as democracies. See Jaggers and Gurr 1995.

Reciprocation.   Many studies of coercive diplomacy use the MID data set's dispute reciprocation variable (RECIP) as a proxy for failed threats.[24] To reciprocate a dispute, the target need only engage in a militarized action (that is, a threat, display, or use of force) at some point during the dispute.[25] The logic of using RECIP as a dependent variable is that only a noncompliant defender would engage in a militarized action in response to a demand, whereas a cooperative target would be unlikely to do so.[26] Dispute reciprocation is therefore assumed to imply a refusal of the demand, whereas a failure to respond with a militarized action is assumed to signify acquiescence.

Reciprocation, however, is a poor indicator of threat effectiveness. First, a reciprocated threat does not necessarily imply a failed threat. In 1994, for example, the United States demanded the abdication of General Raoul Cédras, who had overthrown Haiti's president in a military coup (MID 4016). The MID data set considers the dispute reciprocated, presumably because Cédras initially resisted the demand. Yet it would be misleading to classify the U.S. threat as a failure because it ultimately succeeded without the use of force: Haitian leaders capitulated once they believed that a U.S. invasion was imminent.

Second, the absence of reciprocation does not necessarily indicate successful coercion. A recalcitrant target need not engage in militarized action to reject a demand—indeed, it may do nothing at all. For example, in 2001, Azerbaijan issued threats of renewed war against Armenia over the disputed territory of Nagorno-Karabakh (MID 4236). The dispute was technically unreciprocated because Armenia took no militarized action in response to the threat, but the threat was a failure—Armenia simply ignored it. Using dispute reciprocation as a proxy for threat outcomes would inaccurately classify this threat as successful.

Hostility levels.   The MID data set's indicator of hostility levels (HOSTLEV) is sometimes used to represent threat outcomes. Its values fall into five categories, representing increasing levels of crisis escalation for each disputant.[27] Studies using this variable to represent threat outcomes generally adopt one of two approaches. The first considers a threat successful if the challenger escalated to a higher level of hostility than its target during the crisis.[28] A second method simply establishes a threshold for the target's escalation behavior, above which a threat is considered unsuccessful. Some studies, for instance, consider a threat to have failed if the target used force at any time during the dispute.[29]

---

24. See, for instance, Schultz 1999 and 2001; Prins 2003; Weeks 2008; and Haynes forthcoming.
25. Jones, Bremer, and Singer 1996.
26. For example, Schultz 2001, 142–43.
27. The levels are (1) no militarized action, (2) threat to use force, (3) display of force, (4) use of force, and (5) war.
28. For example, Bennett and Stam 2000.
29. In other words, HOSTLEV ≥ 4. See, for instance, Sartori 2005, 83–84.

The escalation behavior of disputants, however, is a questionable proxy of threat efficacy because it says nothing about whether a demand was accepted or rejected. States need not take militarized actions to reject threats; instead, leaders can simply ignore them. The hostility-level approach, however, might mistakenly classify such threats as successes by virtue of the challenger's higher level of hostility. After al-Qaeda militants bombed the U.S. embassies in Kenya and Tanzania in August 1998, for example, the United States demanded that Afghanistan's Taliban regime hand over Osama bin Laden and launched cruise missile attacks when its demands were ignored. The United States is listed as having a higher HOSTLEV value because it was the only participant to take militarized action during the dispute, but its threat failed. Knowing that a state engaged in actions more "hostile" than its opponent conveys little about the target's compliance with the original demand. Indeed, such escalation may actually indicate that the demand was ineffective, requiring the challenger to resort to more forceful measures.

Outcome.   Both the MID and ICB data sets contain a variable called OUTCOME, which indicates whether the encounter constituted a victory, defeat, compromise, or stalemate for each disputant or crisis participant. At first glance, OUTCOME appears to be a better measure of the true effectiveness of threats than either RECIP or HOSTLEV. Indeed, several studies have used this variable to construct indicators of success and failure, with victories for the dispute initiator coded as successful threats.[30]

The OUTCOME variable, however, fails to distinguish successes achieved through coercive diplomacy from those achieved through physical compulsion. Triumphs won by brute military force—such as Germany's invasions of Belgium, the Netherlands, and France in 1940—are conflated with victories achieved by fear—such as Germany's peaceful annexation of Austria, the Sudetenland, and rump Czechoslovakia in 1938 and 1939. Consider, for instance, the U.S. effort to persuade Iraq to withdraw from Kuwait in 1990 and 1991 (MID 3957; ICB 393). Each data set codes the episode as a "victory" for the United States and its allies, because the U.S.-led coalition ultimately achieved its objectives. Yet President George H.W. Bush's original demand that Iraq "withdraw from Kuwait completely and without condition" was rejected, requiring Coalition forces to physically expel the Iraqi military in the 1991 Persian Gulf War. The distinction between coercive diplomacy and physical compulsion is important because the democratic credibility proposition makes predictions about the conditions under which a challenger's threats will be effective, not who will end up with a disputed object after a brawl. Backing down to a coercive threat is not the same as being defeated in battle.

---

30. Rioux 1998; Partell and Palmer 1999; and several others create dichotomous indicators using this variable; Partell 1997 (using MID), and Gelpi and Griesdorf 2001 (using ICB) create trichotomous variables by combining compromises and stalemates into draws.

Illustration: Two Studies of Democratic Credibility

To illustrate how these two problems impede our ability to draw reliable inferences about the effectiveness of democratic threats from the MID and ICB data sets, we turn to a pair of influential studies: Schultz's 2001 book *Democracy and Coercive Diplomacy* and Gelpi and Griesdorf's 2001 article "Winners or Losers? Democracies in International Crisis, 1918–94" in the *American Political Science Review*. These two studies are examples of the best theoretical and empirical work on democratic credibility theory, and they have significantly advanced research about domestic politics and international conflict. They are useful examples both because they reach similar conclusions with different data—Schultz's study employs the MID data set, whereas Gelpi and Griesdorf use data from ICB—and because they adopt different empirical approaches to evaluating the theory. We show, however, that empirical support for democratic credibility theory in these studies rests largely on misclassified or irrelevant observations from the MID and ICB data sets.[31]

Democracy, Coercive Diplomacy, and the MID Data set.   Using a data set of 2,042 MIDs from 1816 to 1992, Schultz's study employed binary logistic regression to estimate the likelihood that a militarized dispute, once initiated, would be reciprocated by its target. The MID reciprocation variable, the study argued, represents a "plausible indicator of how genuine the target believes the challenge to be."[32]

The study's key test of the democratic credibility hypothesis lay with the variable DEMOCRATIC INITIATOR, a dichotomous indicator of whether the initiator of the dispute was a democracy.[33] According to Schultz, the negative coefficient associated with this variable in the regressions implied that MIDs initiated by democracies were less likely to be reciprocated, thus supporting the view that threats from democracies "are more likely to get their targets to back down without a fight." Specifically, the analysis indicated that disputes initiated by democracies were approximately 30 percent less likely to be reciprocated.[34]

31. We emphasize that these problems do not stem from errors in the original studies. Rather, they are a product of the MID and ICB data sets' limitations as testing grounds for theories about coercive threats.

32. Schultz 2001, 142.

33. To identify democracies, Schultz 2001 incorporated data from the Polity III data set, which provides yearly information about the characteristics of national political regimes (Jaggers and Gurr 1995). Democracies were defined as states featuring a competitively elected executive as well as a stable, enduring system of political competition. In the language of the Polity data set, this translates to XRCOMP $\geq$ 2 and PARREG = 2 or 5. Although this is not the conventional method for coding democracies with Polity data, Schultz argued that it more closely reflects the central logic of the democratic credibility hypothesis, which requires a viable political opposition plus a credible threat of electoral punishment for leaders who earn the public's ire.

34. Schultz 2001, 9 and 120; see also Schultz 1999.

To determine which specific cases were most responsible for Schultz's findings, we first replicated the study's results.[35] We were unable to obtain a complete replication data set for the study, but by following the procedures described in *Democracy and Coercive Diplomacy*, we constructed a data set that very nearly reproduces the original regression results.[36]

In the replicated version of Schultz's data set, there are 147 democratic victories—that is, unreciprocated MIDs initiated by democracies. Of these episodes, Table 2 lists the twenty-five most influential cases as measured by the DFBETA statistic, which reports the change in the coefficient of interest when an individual case is excluded from the regression.[37] These twenty-five cases—roughly 2 percent of the overall data set—exert the greatest downward impact on the DEMOCRATIC INITIATOR variable, in effect doing the most work to confirm the predictions of the democratic credibility hypothesis.

The empirical analysis in *Democracy and Coercive Diplomacy* relied on the MID data set because, according to the study, the data set is composed of "cases in which states used threats of force, regardless of how prominent or how severe the ensuing crisis eventually became."[38] If true, the data set would indeed be useful for testing the democratic credibility hypothesis, because the theory aims to explain the success and failure of militarized threats. However, this characterization of the MID data set is largely incorrect. Our research indicates that none of the twenty-five most influential democratic victories in our replication of Schultz's analysis represents an actual threat made by a democracy.[39] In fact, as Table 2 reports, eight of the twenty-five cases appear to involve no militarized dispute at all.[40] The remaining cases entail unilateral raids, skirmishes, or border violations (twelve cases), troop movements or exercises without a coercive demand (two cases), and encounters with fishing boats or other civilian vessels (three cases). Because these cases do not involve coercive threats, they do not belong in an empirical test of democratic credibility theory.

35. In the original study, the regression model most strongly supporting the democratic credibility hypothesis (both statistically and substantively) included bilateral MIDs only and dropped observations from the world war years (Schultz 2001, 146–47). This is the model we replicate here.

36. Schultz 2001 describes these procedures in meticulous detail on pages 125–31, 141–44, and 261–72. Details of the replication can be found in the online Appendix for this article.

37. Pregibon 1981.

38. Schultz 2001, 126.

39. Because the MID data set contains no narratives for these disputes, we conducted independent research using the data set's original sources (where available) as well as the *New York Times*, *Keesing's Contemporary Archives*, and *Facts on File*. Michael Tomz and Jessica Weeks also generously provided access to their MIDipedia database, which contains valuable new research about these and hundreds of other MIDs (Tomz and Weeks 2009).

40. We were unable to locate any evidence of a potential MID in three of these eight cases. In most of the remaining cases, the only militarized activity took place in a nonconflictual context. For example, MID 3382 lists a dispute between Israel and Egypt from 14–19 November 1977, but the only militarized activity occurring during this period was the use of Israeli security forces to guard Egyptian President Anwar Sadat during an official state visit. We found no evidence of an Israeli threat or other militarized action directed against Egypt.

TABLE 2. *The twenty-five most influential democratic victories in the authors' replication of Schultz's (2001) analysis*

Dispute number	Initiator	Target	Year	DFBETA ($\times 10^{-2}$)	Presumed issue	Reclassification
1335	Botswana	Rhodesia	1969	−2.342	Ban on military exports to Rhodesia	No militarized encounter
1739	Greece	Ottoman Empire	1888	−2.265	No dispute identified	No militarized encounter
2089	India	China	1954	−2.158	Alleged Indian incursion into disputed territory near Tibet	Skirmish/border violation
2101	India	China	1979	−2.026	Alert of Indian forces following Chinese invasion of Vietnam	Alert/exercise
1100	Chile	Argentina	1959	−1.922	Incursion into Argentine territory	Skirmish/border violation
2116	United States	Mexico	1842	−1.901	Unauthorized seizure of Mexican fort in Monterey, California	Skirmish/border violation
1462	India	Nepal	1968	−1.859	Arrest of four alleged Nepali trespassers in Indian territory	No militarized encounter
3421	Israel	Egypt	1959	−1.831	Seizure of Egyptian freighter	Maritime incident
3420	Israel	Egypt	1958	−1.813	No dispute identified	No militarized encounter
3376	Israel	Egypt	1962	−1.799	No dispute identified	No militarized encounter
3382	Israel	Egypt	1977	−1.734	Alert of Israeli security forces during visit by Egyptian president	No militarized encounter
3383	Israel	Egypt	1979	−1.722	Alleged Israeli violation of Egyptian airspace	Skirmish/border violation
2087	Argentina	Chile	1984	−1.689	Firing of artillery shells near Beagle Channel	Skirmish/border violation
3414	Israel	Syria	1952	−1.685	Seizure of Syrian cargo ship	Maritime incident
3416	Israel	Syria	1952	−1.676	Seizure of Syrian schooner carrying medical aid	Maritime incident
2841	Burma	Thailand	1959	−1.675	Strafing of Thai village by Burmese aircraft	Skirmish/border violation
3117	Malaysia	Philippines	1980	−1.662	Alleged construction of guardhouse on Commodore Reef	Skirmish/border violation
2100	India	China	1973	−1.636	Indian seizure of facilities taken by demonstrators in Sikkim	No militarized encounter
2099	India	China	1971	−1.629	Alleged Indian reconnaissance activity in Chinese territory	Skirmish/border violation
2878	Austria	Hungary	1960	−1.525	Austrian violation of Hungarian airspace	Skirmish/border violation
1555	United States	Mexico	1836	−1.523	Border violations by U.S. troops during Texas revolution	Skirmish/border violation
2042	Costa Rica	Nicaragua	1957	−1.496	Reinforcement of Costa Rican border defenses	Alert/exercise
1309	India	Pakistan	1961	−1.487	Restatement of commitment to defend Kashmir	No militarized encounter
3187	Greece	Bulgaria	1928	−1.483	Shooting of Bulgarian border guard	Skirmish/border violation
1548	United States	Mexico	1859	−1.470	Attempt to suppress Mexican bandits in Texas	Skirmish/border violation

One of the data set's most influential democratic successes—a MID initiated by India against China in 1954 (MID 2089)—illustrates the disjuncture between the study's inferences about the credibility of threats and the MID cases that supposedly support those inferences. In this episode, China protested what it claimed was an incursion of troops near Barahoti, a tiny Indian village north of the Pindari glacier that sits at an elevation of nearly 17,000 feet. A letter dated 17 July 1954, from the Chinese embassy in Delhi complained that a small group of Indian troops had crossed into Tibet via the Niti pass, just northwest of Barahoti, on 29 June. India's reply, delivered on 27 August, denied that any incursion took place and instead accused Tibetan officials of crossing into Indian territory without proper documentation.[41] While it is probably impossible to adjudicate these competing claims, there is no evidence that either side made a threat to use military force over the issue during the span of the MID (29 June—19 September 1954). Moreover, because the episode consisted only of private notes exchanged between diplomats, key domestic audiences—specifically, the Indian electorate and opposition parties—did not even know about the event and could not have bolstered India's bargaining position by threatening political punishment for backing down.

The Barahoti case also illustrates why the data set's dispute reciprocation variable is a poor indicator of the effectiveness of threats. The MID data set codes the case as a victory for India, apparently because China did not immediately retaliate. However, repeated border violations by Chinese troops in the years following the dispute—culminating in India's defeat in the Sino-Indian war of 1962—suggest that India's incursion, whether real or perceived, was hardly unreciprocated. While China may not have reciprocated this particular dispute according to MID coding rules, it certainly did not acquiesce to any Indian demand.

In short, when we closely examine the cases in Schultz's analysis that provide the greatest statistical support for the democratic credibility hypothesis, they do not appear to substantiate the inference that "threats from democratic governments are less likely to be resisted than threats made by nondemocratic governments."[42]

Winners, Losers, and the ICB Data set.   Perhaps the leading study finding support for the democratic credibility hypothesis with ICB data is Gelpi and Griesdorf's article "Winners or Losers."[43] Three features of their empirical approach distinguish it from Schultz's method. First, Gelpi and Griesdorf posited that democratic challengers enjoy an advantage only when they have a comparatively greater ability to generate audience costs. The rationale, following Fearon's logic,[44] was that a democratic target might be able to neutralize a challenger's advantage by generating audience costs of its own. A key independent variable in their empirical analysis was therefore "relative audience costs"—the difference between the

---

41. Indian Ministry of External Affairs 1959.
42. Schultz 2001, 234.
43. Gelpi and Griesdorf 2001.
44. Fearon 1994.

disputants' levels of democracy. Second, Gelpi and Griesdorf included military escalation as an independent variable in their analysis, arguing that military actions can publicly commit democratic leaders to hard-line bargaining stances. Third, Gelpi and Griesdorf used a trichotomous measure of threat outcomes, contrasting with Schultz's dichotomous approach.

Using a data set of 422 ICB dyads (representing 283 crises), Gelpi and Griesdorf employed ordered probit regressions to evaluate the factors influencing a challenger's likelihood of achieving a victory, draw, or loss in a crisis. In keeping with their argument about the interactive effects of relative audience costs and resolve, they tested democratic credibility theory using an interaction term combining these two factors. Their analysis showed that highly resolved, democratic challengers facing poorly resolved, autocratic opponents enjoyed a greater likelihood of victory than autocratic challengers under the same circumstances. Their statistical model estimated that an autocratic challenger with the maximum advantage in relative resolve would stand only about a 13 percent chance of achieving victory against an autocratic opponent, whereas its odds of winning would rise to roughly 80 percent if it were democratic.[45] Gelpi and Griesdorf argued that these results provide "striking and dramatic support" for the argument that political institutions allow democracies to make more credible threats than nondemocracies.[46]

Democratic victories in Gelpi and Griesdorf's data set must meet four criteria to validate the study's argument about the interaction of audience costs and resolve. First, the challenger must issue a demand backed by the threat or demonstration of military force, in an attempt to "coerce an opponent into backing down."[47] Second, the defender must comply with the demand: in other words, the democracy must "persuade their opponent to yield" voluntarily.[48] Importantly, this does not mean that the crisis must avoid violence altogether—but it does imply that the target must eventually give up the disputed item of its own volition rather than simply rejecting the threat or resisting to the point where the challenger seizes it by force. Third, the democratic challenger must issue a more forceful public demonstration of resolve than the target, thereby signaling its willingness to fight and tying the leader's hands. Finally, the challenger must be a democracy with a higher POLITY score than the defender, thus giving the challenger a greater ability to generate audience costs.

Table 3 lists all twenty-three cases in Gelpi and Griesdorf's analysis in which a democratic challenger prevailed. Most striking is how few of these democratic victories actually entail successful coercive threats. Just five cases involve a successful deterrent or compellent demand backed by a threat of military force—and only four of these meet the criteria enumerated above. Of the remaining nineteen

45. We obtained these estimates using *Clarify* (King, Tomz, and Wittenberg 2000).
46. Gelpi and Griesdorf 2001, 644.
47. Ibid, 636.
48. Ibid, 637.

cases, fifteen consist of cross-border attacks in pursuit of insurgents (for example, Operation Tangent), surprise raids (for example, Raid on Entebbe and Iraq Nuclear Reactor), minor clashes involving no demands (for example, Ecuador-Peru Border III), and interstate war dyads in which no threats were issued (for example, the U.S.-Germany and U.S.-Italy dyads in the Pearl Harbor crisis).[49] Three cases (Mid-East Campaign, Junagadh, and Invasion of Panama) contain coercive threats but do not support democratic credibility theory because these threats did not succeed: the challenger prevailed only by militarily defeating its opponent after coercive diplomacy failed. The final case—the Cuban Missile Crisis—does not support the formulation of democratic credibility theory described by Gelpi and Griesdorf because, according to the ICB data set, the United States failed to demonstrate higher resolve than the Soviet Union, thus nullifying its audience cost advantage.

Just four democratic victories in Gelpi and Griesdorf's data set meet the necessary conditions to affirm the logic of democratic credibility: the Hungarian War, German Reparations, Central America-Cuba II, and Goa II crises. Careful examination of these cases, moreover, reveals several problems that undercut their support for the hypothesis that democracies make more effective threats in international crises.

First, two of these democratic victories actually involve several unsuccessful threats that do not appear in the data set. In the Hungarian War crisis, for example, France and its allies threatened Hungary over its possession of Czech and Romanian territory four times in 1919, succeeding only once.[50] Likewise, in the German Reparations crisis, France issued two threats in the spring of 1921 to compel German payment of reparations for World War I; Germany rejected the first and complied (temporarily) with the second.[51] In both instances, however, the ICB data set combines these threats into a single crisis and codes only the final outcome, thus excluding the failed threats and painting a misleading picture of democratic credibility.

A second problem is that one of the four democratic threats "succeeded" only after the challenger launched an invasion to seize the item it had demanded. In 1961, India demanded Portugal's withdrawal from its colonial enclave of Goa. The Indian public strongly backed Prime Minister Jawaharlal Nehru's threat to retake the territory by force, which—according to democratic credibility theory—ought to have communicated the genuineness of the threat to Portugal.[52] But Por-

---

49. The United States is coded as challenging Germany and Italy during the Pearl Harbor crisis because it declared war against Japan—an ally of the two Axis powers—on 8 December 1941. Because the United States is judged to have won this crisis with its victory at the Battle of Midway in 1942, it is coded in ICB as defeating Germany and Italy in the crisis as well.

50. See Swanson 2001, 79–81, 130–31, 139–40; and Juhász 1979, 24.

51. German compliance lasted only a short time; France occupied the Ruhr in January 1923 after Germany defaulted on its reparations payments. See Trachtenberg 1980, 204–5; and McDougall 1978, 155.

52. As the *New York Times* reported on 10 December 1961, "It is clear … that prevailing public opinion favors military action. The troop movements and anti-Portuguese speeches in Parliament this

TABLE 3. *Victories by democratic challengers in Gelpi and Griesdorf's (2001) data set*

Crisis number	Year	ICB name	Challenger	Defender	Audience cost advantage?	Relative resolve advantage?	Threat?	Effective threat?
6	1919	Hungarian War	France	Hungary	✓	✓	✓ (×4)	✓ (×1)
20	1921	German Reparations	France	Germany	✓	✓	✓ (×2)	✓ (×1)
27	1923	Ruhr I	France	Germany	✓			
84	1941	Mid-East Campaign	United Kingdom	Iraq	✓		✓	
88	1941	Pearl Harbor	United States	Germany	✓			
88	1941	Pearl Harbor	United States	Italy	✓			
118	1947	Junagadh	India	Pakistan	✓			
144	1953	Guatemala	United States	Guatemala	✓		✓	
178	1960	Central America-Cuba II	United States	Cuba	✓	✓	✓	✓
190	1961	Goa II	India	Portugal	✓		✓	✓
196	1962	Cuban Missile Crisis	United States	Soviet Union	✓	✓	✓	✓
222	1967	Six-Day War	Israel	Soviet Union	✓			
242	1971	Bangladesh	India	Pakistan	✓			
257	1974	Cyprus III	Turkey	Cyprus	✓			
270	1976	Raid on Entebbe	Israel	Uganda	✓			
276	1976	Operation Tangent	Rhodesia	Botswana	✓			
294	1978	Nicaragua Civil War II	Costa Rica	Nicaragua	✓			
307	1979	Rhodesia Settlement	Rhodesia	Botswana	✓			
311	1980	Raid on Gafsa	France	Libya	✓			
322	1981	Ecuador-Peru Border III	Ecuador	Peru	✓	✓		
324	1981	Iraq Nuclear Reactor	Israel	Iraq	✓	✓		
391	1989	Invasion of Panama	United States	Panama	✓	✓	✓	
406	1992	Iraq No-Fly Zone	United States	Iraq	✓			

tuguese leaders resisted, conceding only after a full-scale invasion by 30,000 Indian troops. Although India prevailed in the dispute, there is little evidence that democratic hand-tying mechanisms had any bearing on the outcome. The deciding factor in India's victory was not Nehru's public commitment but a well-planned military assault.

Finally, the fourth democratic victory involves a deterrent threat whose effect is probably impossible to ascertain. In the Central America-Cuba II case, U.S. President Dwight D. Eisenhower issued a threat in November 1960 to deter Cuban-backed insurgents from mounting attacks against U.S. allies in Central America. Eisenhower stated that U.S. forces in the area would "act to prevent invasions by communist-directed elements,"[53] but it is not clear that any such invasions were planned, nor that Cuba (ostensibly the target of the threat) played any role in dissuading them. The case represents an ambiguous success at best.[54]

Thus, among the twenty-three democratic victories in Gelpi and Griesdorf's quantitative data set, only four actually confirm their formulation of the democratic credibility hypothesis. These four cases contain a total of eight threats, of which just two were clearly successful—four others failed outright, and two outcomes were ambiguous. The democratic victories in Gelpi and Griesdorf's data set therefore furnish scant support for the logic of democratic credibility.

Overall, this analysis suggests that the frequency of successful democratic threats in the MID and ICB data sets is considerably lower than previously believed. In both Schultz's analysis of MID data[55] and Gelpi and Griesdorf's study of ICB crises,[56] many apparent democratic victories either do not involve coercive threats at all or are coded incorrectly. These findings imply that existing empirical support for the democratic credibility theory should be viewed with skepticism.

Re-evaluating Democratic Credibility Theory

In this section, we conduct a more comprehensive multivariate reassessment of democratic credibility theory using a new data set constructed explicitly for studying coercive threats: the Militarized Compellent Threats (MCT) data set.[57] The MCT data set contains information about 210 interstate compellent threats issued between 1918 and 2001, comprising 242 crisis dyads.

---

week by Mr. Nehru, Mr. Krishna Menon and others have strengthened this sentiment, not weakened it. Foreign observers question whether public sentiment would permit the Army to withdraw without taking some action."

53. *New York Times*, 18 November 1960.

54. Indeed, the case is not included in either of the extended deterrence datasets compiled by Huth and Russett 1984; and Danilovic 2002.

55. Schultz 2001.

56. Gelpi and Griesdorf 2001.

57. Sechser 2011.

The MCT data set is well suited for testing democratic credibility theory because it meets the criteria necessary for an appropriate empirical test of the democratic credibility proposition. First, it contains the proper unit of analysis: a coercive demand accompanied by a threat to use force. Specifically, the unit of analysis in the MCT data set is the compellent threat, defined as "an explicit demand by one state (the challenger) that another state (the target) alter the status quo in some material way, backed by a threat of military force if the target does not comply."[58] In other words, episodes in the MCT data set have two components: a coercive *demand* and a *threat* to use military force. With respect to coercive demands, the data set requires that demands be made verbally to mitigate the possibility that the target did not understand what was being asked of it.[59] Threats to use force, however, may be communicated explicitly through verbal messages or implicitly through militarized actions (such as troop maneuvers or exercises) that coincide with a verbal demand. The MCT data set's focus on coercive threats contrasts sharply with the MID and ICB data sets: whereas these data sets include many episodes of conflict that do not contain threats or demands, the MCT data set excludes such cases.[60]

Second, the episodes in the MCT data set are coded according to whether the target complied voluntarily with the threat, irrespective of who eventually ended up in possession of the disputed item. Specifically, the COMPLIANCE variable reports whether the target complied with the challenger's original compellent demands in full, in part, or not at all.[61] In coding this variable, only coercive diplomatic outcomes—rather than military outcomes—are considered. For example, whereas the MID and ICB data sets both code the 2001 war in Afghanistan as a victory for the U.S.-led coalition on the grounds that it achieved its objective of removing the Taliban from power, the MCT data set codes it as a failed compellent threat because the U.S. ultimatum demanding the extradition of al-Qaeda leaders was clearly rejected by the Afghan government. Although the U.S. coalition achieved its objectives, it did so only through battlefield victories, not successful coercive threats. This distinction is important because democratic credibility theory asserts that democratic threats are more believable and therefore more likely to achieve their objectives with minimal military force. Consequently, if a challenger attains its objectives only by employing large-scale military force, the threat must be considered a failure. As Schelling notes, "successful threats are those that do not have to be carried out."[62]

---

58. Ibid, 380.
59. The requirement that demands be made verbally has the advantage of providing an objective gauge for measuring threat outcomes. In contrast, if a challenger leaves its demands unstated, one cannot be sure whether the target complied with the challenger's wishes or not.
60. Importantly, the data set does not exclude crises simply because military escalation took place; rather, it excludes cases only if militarized actions occurred without a demand being made.
61. Sechser 2011.
62. Schelling 1966, 10.

The cases in the MCT data set also dovetail nicely with the mechanisms through which democratic institutions are thought to enhance the credibility of threats. The MCT data set demands are public and articulated verbally to the target, thus increasing the likelihood that domestic audiences (such as voters or opposition parties) can observe them.[63] This is essential because theories of democratic credibility maintain that democratic threats tie a leader's hands only if they are issued before a public audience. Moreover, verbal demands commit leaders to specific negotiating positions that are difficult to retract. In contrast, tacit or unspoken demands inherently carry a degree of plausible deniability: if no demand was made, then leaders can always claim that they never actually asked for a revision of the status quo. This has the advantage of providing leaders with face-saving exits from commitments they do not wish to fulfill, but at the same time, the logic of democratic credibility theory also expects such demands to be less convincing because they expose leaders to less political risk. Explicitly enumerated demands, by contrast, are harder to disavow, and therefore ought to be more persuasive. The MCT data set thus offers a set of cases in which the logic of democratic credibility is most likely to operate.[64]

Below we use the MCT data set to reassess the democratic credibility hypothesis. We emulate the operationalization and model specification procedures used by Schultz and Gelpi and Griesdorf in order to provide a favorable testing ground for their hypotheses about democracies and threats. Our findings, however, do not support the hypothesis that democratic threats are more effective.

Binary Logit Analysis: Democracy and Threat Effectiveness

The first analysis implements Schultz's method of evaluating democratic credibility, using binary logistic regressions to estimate the effect of democracy on crisis outcomes.[65] Specifically, the dependent variable is the success or failure of a compellent threat. Compellent threats are considered successful here if they meet one of two conditions. First, a threat succeeds if it is coded in the MCT data set as achieving full compliance from the target (COMPLIANCE = 2) without the use of force. Second, threats that achieve their objectives only after force is used are nevertheless considered successful if the military engagement entails fewer than 100 military fatalities on the target's side. This standard acknowledges the possibility that small-scale uses of force can reveal a challenger's resolve, helping the challenger to achieve its coercive aims without resorting to outright physical com-

---

63. More than 90 percent of MCT episodes contain documentation from a contemporary news source, suggesting a high degree of public visibility. See Sechser 2011, 392.

64. It is important to distinguish between *demands* and *threats*. To be included in the MCT data set, leaders must issue verbal demands ("do *x*"). However, threats to use force ("I will attack unless you comply") need not be verbal—in the MCT data set, threats to use force may be implied by troop movements or other militarized actions. See Sechser 2011.

65. Schultz 2001.

pulsion. Successful threats are coded 0 (and 1 otherwise) so that factors that improve the effectiveness of threats take on a negative coefficient, consistent with Schultz's study.

We include five groups of control variables, in keeping with Schultz's analysis. Four were generated from the same data as the original study: (1) major power status, (2) material capabilities, (3) geographic contiguity, and (4) foreign policy portfolio similarity.[66] For the fifth set of control variables, which indicate the issue at stake in each crisis, we utilized the MCT data set's issue classification variables, which code for the same issue types as the MID data set. The analysis thus includes all fourteen control variables present in Schultz's original regressions.

As in *Democracy and Coercive Diplomacy*, the main test of democratic credibility theory is provided by the independent variable DEMOCRATIC INITIATOR, which is coded 1 if the initiator of the threat meets Schultz's criteria for being a democracy, and 0 otherwise.[67] If democratic threats are indeed more credible, as the theory expects, the variable's estimated coefficient should be negative and statistically significant.

Results.   Schultz originally presented four binary logit regression models (reproduced in the top half of Table 4), which variously include or exclude multilateral disputes and observations from the world war years (1914–18 and 1939–45).[68] We repeated each of these regressions using the 242 compellent threat dyads in the MCT data set; the results are reported in the bottom half of Table 4.[69] In all four regressions using the MCT data set, the coefficient for DEMOCRATIC INITIATOR is statistically insignificant even at the 90 percent level—indeed, the coefficient is positive, contrary to the predictions of democratic credibility theory. These results indicate that we cannot reject the null hypothesis that democracies are no more likely to make successful compellent threats. In other words, the impact of democracy on the effectiveness of compellent threats is statistically indistinguishable from zero.

We repeated these regressions using several different dichotomous measures of democracy, including the standard Polity definition of democracy[70] as well as alternative indicators constructed by Cheibub, Gandhi, and Vreeland,[71] Przeworski and

66. The control variables are discussed further in the online Appendix.
67. Schultz 2001, 127–30.
68. Ibid, 146–47.
69. Owing to space considerations, Table 4 excludes fourteen control variables. Complete regression results can be found in the online Appendix.
70. As noted earlier, Schultz's study adopted a somewhat idiosyncratic method for defining democracies with the Polity data set, using some component indicators of regime type and discarding others. A more widely used method involves subtracting a state's 10-point AUTOCRACY score from its 10-point DEMOCRACY score, resulting in an overall POLITY score ranging from −10 to +10. States with scores of +6 or higher are then classified as democracies.
71. Cheibub, Gandhi, and Vreeland 2010.

TABLE 4. *Logit analyses of* MID *reciprocation and compellent threat failure*

	Logit analysis of MID reciprocation in Schultz 2001			
	All crises		*Bilateral crises only*	
	1 *World wars included*	*2* *World wars excluded*	*3* *World wars included*	*4* *World wars excluded*
DEMOCRATIC INITIATOR	−0.22	−0.49**	−0.43*	−0.53**
	(0.16)	(0.17)	(0.18)	(0.19)
N	1,768	1,559	1,329	1,184
Correctly predicted	68.2%	67.0%	68.0%	68.0%
Modal outcome	51.1%	51.3%	51.2%	50.8%

	Logit analysis of compellent threat failure using the MCT data set			
	All crises		*Bilateral crises only*	
	1 *World wars included*	*2* *World wars excluded*	*3* *World wars included*	*4* *World wars excluded*
DEMOCRATIC INITIATOR	0.31	0.22	0.40	0.32
	(0.55)	(0.56)	(0.63)	(0.64)
N	230	198	177	149
Correctly predicted	74.8%	78.3%	71.8%	71.8%
Modal outcome	60.9%	64.1%	56.5%	61.1%

Notes: Standard errors in parentheses. Models 1 and 2 employ robust standard errors clustered on crisis. † $p < .10$; * $p < .05$; ** $p < .01$. Dummy variables for world war years are included in Models 1 and 3 (both versions) but not reported here; coefficients and standard errors for fourteen other control variables are also not reported.

colleagues,[72] and Boix and Rosato.[73] Moreover, we repeated the regressions with more lenient—as well as more stringent—indicators of success and failure.[74] In all cases, the results remained substantively unchanged.[75] These findings undermine the contention that democracies make more effective coercive threats than nondemocracies.

72. Przeworski et al. 2000.
73. Boix and Rosato 2001.
74. We redefined successful threats in two ways: first, by restricting them to threats that achieved their objectives without any military fatalities (as opposed to permitting 100 or fewer fatalities), and second, by expanding them to include threats that achieved any of their objectives (as opposed to all of their objectives).
75. These results are available in the online Appendix.

One might object that these results are simply driven by the smaller size of the MCT data set: in other words, it could be the case that democracies actually make more effective compellent threats, but the smaller *n* obscures this relationship by inflating the standard errors.[76] As we have demonstrated, however, the MID data set contains a large number of cases that are inappropriate for evaluating the credibility of democratic threats. This implies that the standard errors in Schultz's study were artificially low, owing to an inappropriately large data set. The higher standard errors reported in Table 4, therefore, are a more valid assessment of the uncertainty surrounding the coefficient estimates.

Furthermore, if the smaller *n* of the MCT data set is responsible for our null findings, then we should be able to obtain similarly insignificant results by shrinking the MID data set down to MCT size. To test this possibility, we re-estimated Models 3 and 4 against 10,000 random samples of bilateral MIDs containing the same number of cases as the MCT versions of these models (177 and 149, respectively). The probability of obtaining standard errors as high (or higher) as those in our regressions was just 2 percent. The probability of getting both high standard errors and coefficients of equivalent direction and magnitude was even lower: 0.3 percent. We therefore can reject the hypothesis that the smaller size of the MCT data set is responsible for our findings.

Ordered Probit Analysis: Relative Audience Costs and Threat Effectiveness

Gelpi and Griesdorf tested a slightly different variant of the democratic credibility hypothesis, using ordered probit regressions to evaluate the claim that democracies enjoy a credibility advantage in crises only if they engage in military escalation—and only if they are more democratic than their opponent.[77] The dependent variable in our analysis, in keeping with Gelpi and Griesdorf's measurement of crisis outcomes, is a trichotomous indicator of the success of compellent threats. Compellent threats are coded as "fully successful" (SUCCESS = 2) if the challenger fully achieved its goals with 100 or fewer target fatalities; "partially successful" (SUCCESS = 1) if the challenger achieved some but not all of its objectives with 100 or fewer target fatalities; and 0 otherwise.[78] This indicator enables us to estimate ordered probit regressions, as in the original study.

---

76. Importantly, our results are not due to a shortage of democratic threats in the MCT data set. In fact, democracies are the most frequent challengers, comprising 36.4 percent of challengers, compared with 36.0 percent for autocracies and 27.6 percent for anocracies. Yet democracies accounted for only about 32 percent of all states during the period covered by the data set, suggesting that they are somewhat more likely than other regimes to issue compellent threats. See Sechser 2011, 388.

77. Gelpi and Griesdorf 2001.

78. Gelpi and Griesdorf 2001 argue that threats become more credible as the challenger demonstrates greater resolve, measured by the challenger's relative reliance on military force. To fairly test their argument, it is therefore important to allow the possibility that successful threats may involve some degree of force.

Gelpi and Griesdorf employed six main classes of independent variables, with several interactions among them. We were able to include four of these six sets—including the study's measure of relative audience costs—using the same procedures and data: (1) democracy score, (2) military capabilities, (3) nuclear weapons possession, and (4) shared alliance ties.[79] Of the remaining two independent variables, RELATIVE RESOLVE depended heavily on data from the ICB data set, so we used an equivalent indicator from the MCT data set instead.[80] The variable RELATIVE INTERESTS AT STAKE, which originally relied on the ICB team's judgment about the value of the issue to each side, could not be included because the MCT data set contains no equivalent assessment of interests. This variable, however, was not central to the study's test of democratic credibility theory.

TABLE 5. *Ordered probit analysis of challenger victory and compellent threat success*

	1 *Gelpi and Griesdorf* *original model*	*2* *Constituent* *terms included*	*3* *Reassessment using* *the MCT data set*
RELATIVE AUDIENCE COSTS	0.0011*	0.0013*	−0.0006
× RELATIVE RESOLVE	(0.0006)	(0.0006)	(0.0012)
RELATIVE AUDIENCE COSTS		−0.0018	0.0009
		(0.0015)	(0.0030)
RELATIVE RESOLVE	−0.009	−0.017	−0.310*
	(0.097)	(0.098)	(0.170)
N	409	409	242
Correctly predicted	47.7%	47.4%	57.4%
Modal outcome	41.1%	41.1%	56.6%

Notes: Robust standard errors in parentheses, clustered on crisis. † $p < .10$; * $p < .05$; ** $p < .01$, in one-tailed tests. Coefficients and standard errors for ten control variables not reported.

Results. Table 5 reports the results of three regressions, displaying only those variables relevant to Gelpi and Griesdorf's original test, which emphasized the

79. We calculated RELATIVE AUDIENCE COSTS by subtracting the defender's score on the POLITY index (ranging from 1 for least democratic to 21 for most democratic) from the challenger's score and squaring the difference, following Gelpi and Griesdorf's procedure (2001, 638).

80. In the original analysis, RELATIVE RESOLVE drew from the ICB data set's measurement of the role of violence in each state's bargaining strategy. The MCT data set does not measure escalation in the same way, so we coded RELATIVE RESOLVE as 1 if the challenger conducted military maneuvers or demonstrations in connection with the threat, 2 if the challenger used military force during the crisis, and 0 otherwise. Although our procedure is different, the central principle is the same: to identify a public signal of resolve that might engage domestic audiences and enhance the credibility of democratic threats (Fearon 1994, and 1997).

interaction between RELATIVE AUDIENCE COSTS and RELATIVE RESOLVE.[81] This interaction term permitted the study to evaluate the hypothesis that states with a comparatively high ability to generate audience costs are more likely to win crises—but only if they also transmit signals of resolve.

The first model in Table 5 reproduces Gelpi and Griesdorf's main test of the theory.[82] Because the original study omitted the constituent variables of several interaction terms in the model, we included these variables in a revised regression in column (2).[83] Column (3) displays the results yielded by estimating an equivalent model using the MCT data set.

Because the key variables in these regressions are interacted with each other, it is difficult to evaluate their independent effects by examining coefficients and significance levels in isolation. Instead, we derive overall predicted probabilities for the outcome of interest. The graphs in Figure 1 depict the likelihood of challenger victory as a crisis challenger becomes increasingly democratic. The hypothetical scenario used to generate these two graphs assumes conditions that are most favorable for a democratic challenger: a highly resolved challenger and a highly autocratic defender.[84] The first graph displays predicted probabilities and confidence intervals using the estimates from Gelpi and Griesdorf's original analysis (Model 2 in Table 5). The second graph uses the estimates from an equivalent model estimated with MCT data (Model 3).[85] To provide a relatively lenient test for democratic credibility theory, the shaded areas in each chart represent 90 percent confidence intervals surrounding the point estimates.[86]

Comparing the two graphs, the differences are immediately apparent. Democratic credibility theory expects the probability of compellence success to increase in tandem with the POLITY score of a highly resolved challenger. Indeed, in Gelpi and Griesdorf's original analysis (chart *a*), it is possible to reject the null hypothesis that democracy does not increase a challenger's likelihood of prevailing in a crisis. In graphical terms, the null hypothesis posits that one can draw a horizontal line across the width of the graph while remaining within the confines of the 90 percent confidence interval. Because this is impossible in chart *a*, the predictions of democratic credibility theory appear to be confirmed.

However, matters change considerably when the MCT data set is used to derive predicted probability estimates (chart *b*). In the second chart, the predicted likeli-

81. Gelpi and Griesdorf 2001. See the online Appendix for complete results.

82. Ibid., 640.

83. Recent scholarship stresses that all subordinate components of interaction terms should be included in regression models to avoid introducing bias into the results (Braumoeller 2004; Brambor, Clark, and Golder 2006). Adding the constituent terms does not alter Gelpi and Griesdorf's (2001) substantive results.

84. In other words, for Gelpi and Griesdorf 2001, RELATIVE RESOLVE = 3 and DEFENDER DEMOCRACY = 1. For the MCT analysis, RELATIVE RESOLVE = 2 and DEFENDER DEMOCRACY = 1.

85. Other variables in the two models are held at their mean or modal values.

86. Confidence intervals and point estimates were plotted using *Clarify* (King, Tomz, and Wittenberg 2000).
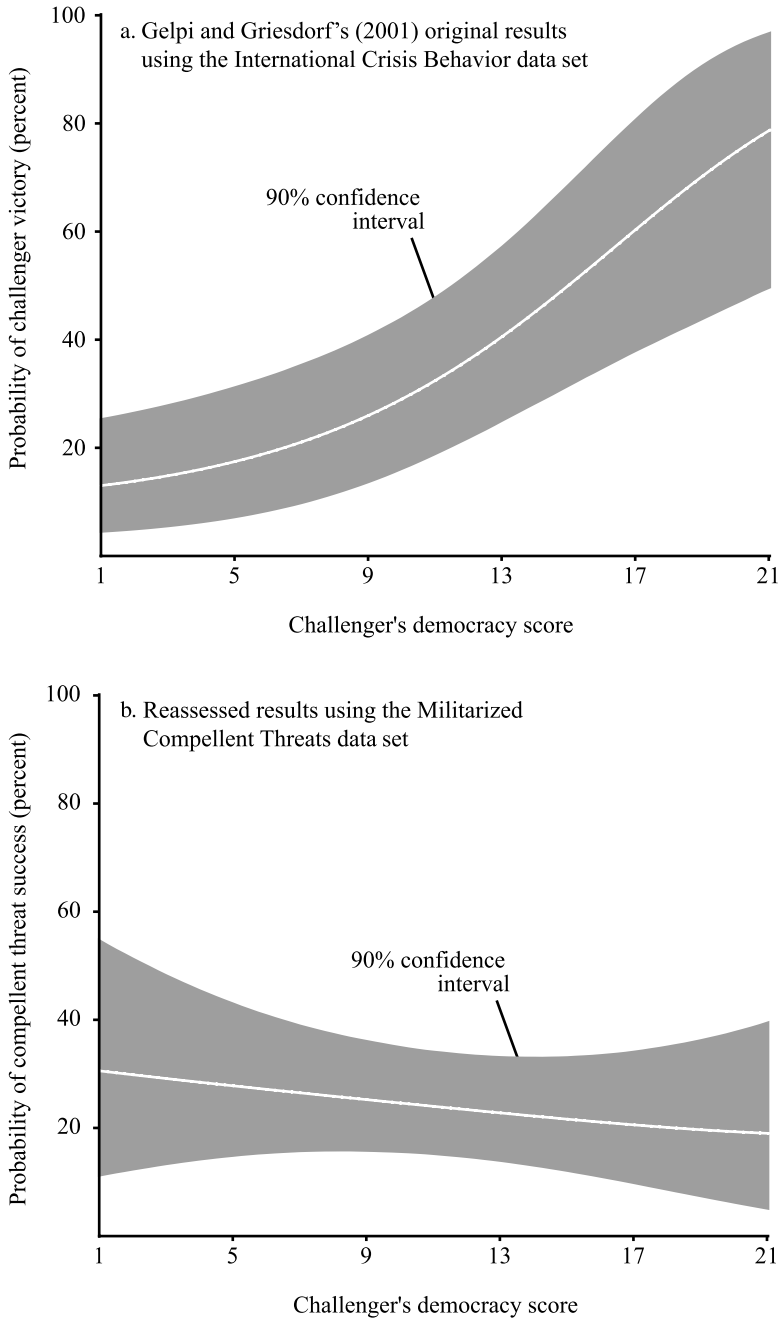
FIGURE 1. *Predicted probabilities of challenger victory and compellant threat success*

hood of democratic victory remains essentially unchanged as the challenger's democracy score increases. Indeed, the likelihood of compellence success declines slightly as a highly resolved challenger becomes more democratic. Moreover, these null results are robust to alternative continuous measures of democracy, including those provided by the Polyarchy data set[87] and the Unified Democracy Score data set,[88] thus casting further doubt on the democratic credibility hypothesis.

Limitations of the Analysis

These results raise considerable concern about the validity of democratic credibility theory, but it would be a mistake to interpret them too broadly. Because the MCT data set does not include deterrence episodes, our analysis cannot yield any specific insights regarding the credibility of democratic deterrent threats. This is an important limitation, but it does not invalidate the tests conducted here. As described by Fearon, Schultz, and others, the logic of democratic credibility theory appears to encompass both deterrence and compellence.[89] Indeed, to our knowledge, no study has argued that democracies enjoy a credibility advantage only when making deterrent threats. Thus, while our results cannot speak directly to democratic credibility in deterrence crises, they do provide an important and valid test of the theory's implications. At the very least, the results suggest that the theory requires revision to account for the apparent lack of a democratic advantage in compellence crises.

A related concern is that the MCT data might be biased against democratic credibility theory because compellence is thought to be harder than deterrence.[90] However, the inherent difficulty of compellence should not prejudice the results against the theory because all challengers in this sample—not just democracies—presumably are burdened by this problem. If compellent threats are inherently less likely to succeed than deterrent threats, this should be true for all compellence challengers irrespective of regime type. Thus, the MCT data set offers a level playing field for democracies and nondemocracies alike.

In fact, the exclusion of deterrent threats in the MCT data set may enhance the data's reliability. Compared with deterrent threats, compellent threats are likely to be less susceptible to the problem of false positives—that is, successes in which the challenger demanded behavior that would have taken place anyway. This is a potentially serious problem when attempting to identify successful deterrent threats, because a deterrent threat might appear to work simply because the recipient never

87. Vanhanen 2000.
88. Pemstein, Meserve, and Melton 2010.
89. See Fearon 1994; and Schultz 1998 and 2001.
90. See, for instance, Schelling 1966.

intended to act in the first place.[91] For instance, the movement of two U.S. aircraft carriers to the Taiwan Strait in response to Chinese military exercises in 1996 could be considered a successful deterrent threat because China did not invade Taiwan. However, this mis-states the effect of the threat because China appears never to have harbored any intention to invade Taiwan during the crisis.[92] By contrast, because compellent threats can succeed only if the target actively alters the status quo, it is more likely that a target's compliant behavior in such cases is actually attributable to the threat.

A final concern is that the MCT data set overlooks cases in which democracies prevailed without ever making a threat. It could be the case that the democratic advantage manifests itself early in disputes, before the challenger has actually made a coercive threat. Because the MCT data set excludes such cases, it might therefore under-report the rate of democratic success. While plausible, this explanation is inconsistent with the logic of democratic credibility, which expects clear and public statements of commitment—like those in the MCT data set—to be more credible than demands that are merely insinuated. As Fearon and others have argued, the quiet diplomatic conversations that often precede military crises may lack credibility because leaders have few disincentives for bluffing, whereas only resolved leaders would risk political punishment by making a public threat.[93] The MCT data set therefore contains cases in which the theory expects democracies to have the greatest advantage. If democracies are instead more likely to succeed with quiet diplomacy, this would be strong evidence against the theory.

Conclusion

This article demonstrates that empirical evidence in favor of the democratic credibility hypothesis is considerably weaker than previously thought. Quantitative support for the proposition that democracies make more effective threats in international crises rests largely on data sets that are almost entirely unsuitable for testing the theory. Just 10 percent of the disputes in the MID data set and 17 percent of the crises in the ICB collection represent coercive threats, making these data sets dubious choices for testing a theory about the conditions under which threats work. Furthermore, the outcomes of the few threats contained in the data sets are often coded incorrectly. When we replace these data sets with one containing only coercive military threats—the MCT data set—support for the democratic credibility

---

91. For example, Lebow and Stein argued that thirty-four of fifty-four threats in Huth and Russett's deterrence data set were improperly identified as successes because the target of the threat never planned to act anyway. Huth and Russett replied that their data collection rules required evidence of a state's initial intent to act, although such counterfactual claims are necessarily speculative. See Huth and Russett 1984; Lebow and Stein 1990; and Huth and Russett 1990, 481–82.
92. Ross 2003.
93. Fearon 1994, 578; 1995.

hypothesis vanishes. Moreover, these results are not idiosyncratic: the foregoing analyses tested democratic credibility theory using a variety of different methods for measuring democracy, resolve, and successful threats, with little effect. Taken together, these results cast significant doubt on the idea that political constraints confer unique advantages on democratic leaders when they make threats in interstate crises.

The implications of these findings are potentially wide-ranging. While the analysis in this article focused primarily on democratic threats, our critique implies that any study of coercive threats conducted through the lens of the MID or ICB data sets should be re-evaluated. The rarity of threats and incorrect coding of threat outcomes in these data sets make them inappropriate for testing any theory about the conditions under which coercive threats are likely to be effective.

It is important to emphasize that our analysis does not necessarily imply that "audience costs" do not exist. Indeed, several studies have found that voters frown upon democratic leaders who renege on threats.[94] Our results do not contest these findings. Yet, even if audience costs exist, one can envision several reasons why they might not bolster the effectiveness of democratic threats. First, democratic credibility theory assumes that democratic leaders can do little to shape their political costs for backing down. It could be, however, that leaders can minimize audience costs by manipulating public opinion, making concessions in secret, or shifting blame, thus mitigating the hand-tying effects of public threats.[95] Second, the political costs of withdrawing a threat may not be severe enough to deter democratic leaders from backing down. Backing down might be a minor concern for voters when weighed against other political issues. Moreover, if voters disapprove of a threat in the first place, their preference for avoiding war might outweigh their distaste for backing down.[96] Thus, even if reneging on a threat tarnishes a leader's image in the abstract, the actual electoral consequences may be slight. Third, democratic credibility theory presumes that target states understand the internal politics of democracies. This assumption, however, might be too heroic: even if democratic leaders stand to pay a political price for failing to fulfill threats, the noise of democratic politics may prevent their adversaries from accurately perceiving this.[97] Finally, audience cost dynamics simply may not be unique to democracies: if autocratic leaders can also generate and signal audience costs, then we would observe little difference between the overall effectiveness of democratic and autocratic threats.[98] Any of these explanations could account for our findings. Further research is needed to help adjudicate among them.

Our empirical findings also highlight the need for continued research about the differences between deterrence and compellence in international relations. While

---

94. See, for instance, Tomz 2007; and Trager and Vavreck 2011.
95. Brown and Marcum 2011.
96. Snyder and Borghard 2011.
97. See Finel and Lord 1999; and Rosato 2003.
98. See Weeks 2008; and Weiss forthcoming.

our analysis suggests that democracies do not, on average, make more effective compellent threats, democracies might nevertheless enjoy a credibility advantage when making deterrent threats.[99] Because the MCT data set includes only compellent threats, the analysis above cannot reject this possibility. At the same time, however, in its current formulation the democratic credibility proposition encompasses both deterrent and compellent threats. The theory therefore may require significant revision to explain why democracies might make more effective deterrent threats but not compellent threats.

Overall, our conclusions call into question one of the most widely accepted theoretical propositions in international relations scholarship: that the political institutions of democratic states render their international threats systematically more credible. Our analysis also questions theories that depend on this mechanism to explain the democratic peace, military effectiveness, and a variety of other empirical phenomena in security studies and international political economy. While democratic credibility theory may begin from empirically valid premises, the implications of those premises deserve serious reconsideration.

References

Art, Robert J., and Patrick M. Cronin, eds. 2003. *The United States and Coercive Diplomacy*. Washington, D.C.: U.S. Institute of Peace.

Bennett, D. Scott, and Allan C. Stam. 2000. A Universal Test of an Expected Utility Theory of War. *International Studies Quarterly* 44 (3):451–80.

Boix, Carles, and Sebastian Rosato. 2001. A Complete Data Set of Political Regimes, 1800–1999. Unpublished manuscript, University of Chicago.

Brambor, Thomas, William Roberts Clark, and Matt Golder. 2006. Understanding Interaction Models: Improving Empirical Analyses. *Political Analysis* 14 (1):63–82.

Braumoeller, Bear F. 2004. Hypothesis Testing and Multiplicative Interaction Terms. *International Organization* 58 (4):807–20.

Brecher, Michael, and Jonathan Wilkenfeld. 1997. *A Study of Crisis*. Ann Arbor: University of Michigan Press.

Brown, Jonathan N., and Anthony Marcum. 2011. Avoiding Audience Costs: Domestic Political Accountability and Concessions in Crisis Diplomacy. *Security Studies* 20 (2):141–70.

Cheibub, José Antonio, Jennifer Gandhi, and James Raymond Vreeland. 2010. Democracy and Dictatorship Revisited. *Public Choice* 143 (1–2):67–101.

Clare, Joe. 2007. Domestic Audiences and Strategic Interests. *Journal of Politics* 69 (3):732–45.

Correlates of War 2 Project. 2004. Dispute Narratives—MID 3.0 Dataset. Unpublished manuscript, Pennsylvania State University, State College.

Danilovic, Vesna. 2002. *When the Stakes Are High: Deterrence and Conflict Among Major Powers*. Ann Arbor: University of Michigan Press.

Desch, Michael C. 2008. *Power and Military Effectiveness: The Fallacy of Democratic Triumphalism*. Baltimore: Johns Hopkins University Press.

Downes, Alexander B. 2009. How Smart and Tough Are Democracies? Reassessing Theories of Democratic Victory in War. *International Security* 33 (4):9–51.

---

99. For example, both Schultz 2001, 173; and Clare 2007 find support for the theory using deterrence data.

Eyerman, Joe, and Robert A. Hart, Jr. 1996. An Empirical Test of the Audience Cost Proposition: Democracy Speaks Louder Than Words. *Journal of Conflict Resolution* 40 (4):597–616.

Fang, Songying. 2008. The Informational Role of International Institutions and Domestic Politics. *American Journal of Political Science* 52 (2):304–21.

Fearon, James D. 1994. Domestic Political Audiences and the Escalation of International Disputes. *American Political Science Review* 88 (3):577–92.

———. 1995. Rationalist Explanations for War. *International Organization* 49 (3):379–414.

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

Finel, Bernard I., and Kristin M. Lord. 1999. The Surprising Logic of Transparency. *International Studies Quarterly* 43 (2):315–39.

Gelpi, Christopher, and Michael Griesdorf. 2001. Winners or Losers? Democracies in International Crisis, 1918–94. *American Political Science Review* 95 (3):633–47.

Guisinger, Alexandra, and Alastair Smith. 2002. Honest Threats: The Interaction of Reputation and Political Institutions in International Crises. *Journal of Conflict Resolution* 46 (2):175–200.

Haynes, Kyle. Forthcoming. Lame Ducks and Coercive Diplomacy: Do Executive Term Limits Reduce the Effectiveness of Democratic Threats? *Journal of Conflict Resolution*.

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

———. 1990. Testing Deterrence Theory: Rigor Makes a Difference. *World Politics* 42 (4):466–501.

Indian Ministry of External Affairs. 1959. *Notes, Memoranda and Letters Exchanged and Agreements Signed by the Governments of India and China, Volume I: 1954–1959*. New Delhi, India: Indian Ministry of External Affairs.

Jaggers, Keith, and Ted Robert Gurr. 1995. Tracking Democracy's Third Wave with the Polity III Data. *Journal of Peace Research* 32 (4):469–82.

Jones, Daniel M., Stuart A. Bremer, and J. David Singer. 1996. Militarized Interstate Disputes, 1816–1992: Rationale, Coding Rules, and Empirical Patterns. *Conflict Management and Peace Science* 15 (2):163–213.

Juhász, Gyula. 1979. *Hungarian Foreign Policy 1919–1945*. Budapest: Akadémiai Kiadó.

King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, N.J.: Princeton University Press.

King, Gary, Michael Tomz, and Jason Wittenberg. 2000. Making the Most of Statistical Analyses: Improving Interpretation and Presentation. *American Journal of Political Science* 44 (2):347–61.

Kuhn, Thomas S. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Lebow, Richard Ned, and Janice Gross Stein. 1990. Deterrence: The Elusive Dependent Variable. *World Politics* 42 (3):336–69.

Lipson, Charles. 2003. *Reliable Partners: How Democracies Have Made a Separate Peace*. Princeton, N.J.: Princeton University Press.

Lyall, Jason. 2010. Do Democracies Make Inferior Counterinsurgents? Reassessing Democracy's Impact on War Outcomes and Duration. *International Organization* 64 (1):167–92.

Mansfield, Edward D., Helen V. Milner, and B. Peter Rosendorff. 2002. Why Democracies Cooperate More: Electoral Control and International Trade Agreements. *International Organization* 56 (3):477–513.

Mansfield, Edward D., and Jack Snyder. 2005. *Electing to Fight: Why Emerging Democracies Go to War*. Cambridge, Mass.: MIT Press.

Martin, Lisa L. 2000. *Democratic Commitments: Legislatures and International Cooperation*. Princeton, N.J.: Princeton University Press.

McDougall, Walter A. 1978. *France's Rhineland Diplomacy, 1914–1924: The Last Bid for a Balance of Power in Europe*. Princeton, N.J.: Princeton University Press.

Merom, Gil. 2003. *How Democracies Lose Small Wars*. Cambridge: Cambridge University Press.

Narang, Vipin, and Rebecca M. Nelson. 2009. Who Are These Belligerent Democratizers? Reassessing the Impact of Democratization on War. *International Organization* 63 (2):357–79.

Partell, Peter J. 1997. Executive Constraints and Success in International Crises. *Political Research Quarterly* 50 (3):503–28.

Partell, Peter J., and Glenn Palmer. 1999. Audience Costs and Interstate Crises: An Empirical Assessment of Fearon's Model of Dispute Outcomes. *International Studies Quarterly* 43 (2):389–406.

Pemstein, Daniel, Stephen A. Meserve, and James Melton. 2010. Democratic Compromise: A Latent Variable Analysis of Ten Measures of Regime Type. *Political Analysis* 18 (4):426–49.

Popper, Karl R. 1959. *The Logic of Scientific Discovery*. London: Hutchinson.

Pregibon, Daryl. 1981. Logistic Regression Diagnostics. *Annals of Statistics* 9 (4):705–24.

Prins, Brandon C. 2003. Institutional Instability and the Credibility of Audience Costs: Political Participation and Interstate Crisis Bargaining, 1816–1992. *Journal of Peace Research* 40 (1): 67–84.

Przeworski, Adam, Michael E. Alvarez, José Antonio Cheibub, and Fernando Limongi. 2000. *Democracy and Development: Political Institutions and Well-Being in the World, 1950–1990*. New York: Cambridge University Press.

Ramsay, Kristopher W. 2004. Politics at the Water's Edge: Crisis Bargaining and Electoral Competition. *Journal of Conflict Resolution* 48 (4):459–86.

Reiter, Dan, and Allan C. Stam. 2002. *Democracies at War*. Princeton, N.J.: Princeton University Press.

Rioux, Jean-Sébastien. 1998. A Crisis-Based Evaluation of the Democratic Peace Proposition. *Canadian Journal of Political Science* 31 (2):263–83.

Rosato, Sebastian. 2003. The Flawed Logic of Democratic Peace Theory. *American Political Science Review* 97 (4):585–602.

Ross, Robert S. 2003. The 1995–96 Taiwan Strait Confrontation: Coercion, Credibility, and Use of Force. In *The United States and Coercive Diplomacy*, edited by Robert J. Art and Patrick M. Cronin, 225–273. Washington, D.C.: U.S. Institute of Peace.

Sartori, Anne E. 2005. *Deterrence by Diplomacy*. Princeton, N.J.: Princeton University Press.

Schelling, Thomas C. 1960. *The Strategy of Conflict*. Cambridge, Mass.: Harvard University Press.

———. 1966. *Arms and Influence*. New Haven, Conn.: Yale University Press.

Schultz, Kenneth A. 1998. Domestic Opposition and Signaling in International Crises. *American Political Science Review* 92 (4):829–44.

———. 1999. Do Democratic Institutions Constrain or Inform? Contrasting Two Institutional Perspectives on Democracy and War. *International Organization* 53 (2):233–66.

———. 2001. *Democracy and Coercive Diplomacy*. New York: Cambridge University Press.

Sechser, Todd S. 2010. Goliath's Curse: Coercive Threats and Asymmetric Power. *International Organization* 64 (4):627–60.

———. 2011. Militarized Compellent Threats, 1918–2001. *Conflict Management and Peace Science* 28 (4):377–401.

Slantchev, Branislav L. 2006. Politicians, the Media, and Domestic Audience Costs. *International Studies Quarterly* 50 (2):445–77.

Smith, Alastair. 1998. International Crises and Domestic Politics. *American Political Science Review* 92 (3):623–38.

Snyder, Jack, and Erica Borghard. 2011. The Cost of Empty Threats: A Penny, Not a Pound. *American Political Science Review* 105 (3):437–56.

Sullivan, Patricia Lynne, and Scott Sigmund Gartner. 2006. Disaggregating Peace: Domestic Politics and Dispute Outcomes. *International Interactions* 32 (1):1–25.

Swanson, John C. 2001. *The Remnants of the Habsburg Monarchy: The Shaping of Modern Austria and Hungary, 1918–1922*. New York: Columbia University Press.

Tarar, Ahmer, and Bahar Leventoğlu. 2009. Public Commitment in Crisis Bargaining. *International Studies Quarterly* 53 (3):817–39.

Tomz, Michael. 2007. Domestic Audience Costs in International Relations: An Experimental Approach. *International Organization* 61 (4):821–40.

Tomz, Michael, and Jessica Weeks. 2009. MIDipedia: An Expanded Database of Militarized Interstate Disputes. Unpublished manuscript, Stanford University, Stanford, Calif./Cornell University, Ithaca, NY.

Trachtenberg, Marc. 1980. *Reparation in World Politics*. New York: Columbia University Press.

———. 2012. Audience Costs: An Historical Analysis. *Security Studies* 21 (1):3–42.

Trager, Robert F., and Lynn Vavreck. 2011. The Political Costs of Crisis Bargaining: Presidential Rhetoric and the Role of Party. *American Journal of Political Science* 55 (3):526–45.

Vanhanen, Tatu. 2000. A New Dataset for Measuring Democracy, 1810–1998. *Journal of Peace Research* 37 (2):251–65.

Weeks, Jessica L. 2008. Autocratic Audience Costs: Regime Type and Signaling Resolve. *International Organization* 62 (1):35–64.

Weiss, Jessica Chen. Forthcoming. Autocratic Signaling and Nationalist Protest in China. *International Organization*.