

# "The Costs of Reneging: Reputation and Alliance Formation"

Douglas M. Gibling  
University of Alabama

**ABSTRACT:** Reputations are supposed to matter. Decision-makers consistently refer to reputations for resolve, and international relations theories confirm the value of being able to credibly signal intentions during times of crisis. However, empirical support for the effects of reputation has been lacking. Problems of strategic selection have hampered previous quantitative tests, and the qualitative literature provides scant support for the concept in individual crises. In this paper I shift the focus from crisis behavior to alliance commitments and examine the effects that opportunities to uphold previous commitments have on future alliance commitments and conflicts. My results demonstrate that alliance reputations do affect both alliance formation and dispute behavior.

Paper presented at the 2006 Shambaugh Conference "Building Synergies: Institutions and Cooperation in World Politics," University of Iowa, 13 October 2006.

President Bush recently argued that the United States—fighting an insurgency in Iraq that was perhaps on the verge of civil war—could not now abandon its policy of state building. Responding to independent assessments that the war had the perverse effect of increasing the number of terrorists in al Qaeda and other terrorist organizations abroad, President Bush argued:

The greatest danger is not that America's presence in the war in Iraq is drawing new recruits to the terrorist cause. The greatest danger is that *an American withdrawal from Iraq would embolden the terrorists* and help them find new recruits to carry out even more destructive attacks (emphasis added).<sup>1</sup>

The linkage between current actions and future dilemmas is not new to this president nor is it unique to American leaders. Decision-makers often cite future reputations as rationales supporting the maintenance of resolve in the face of crisis, and most deterrence theorists seem to agree with these prescriptions. Reputations for resolve give added leverage to leaders who wish to send credible signals during times of intense hostility. Threats made by previously resolute leaders are not dismissed as bluffs, making opponents more likely to back down.

The problem with this theory is that quantitative tests of such an intangible concept like reputation, in an environment plagued by problems of strategic selection, have not consistently demonstrated that opponents take reputations of resolve seriously. Further complicating the issue is that most of the qualitative literature cannot even demonstrate mixed support for the concept; instead, reputations of resolve appear to be inconsequential when compared to variables like state interests and power, which both vary drastically from crisis to crisis, even for the same rivals.

---

<sup>1</sup>“Bush Speech Links Politics, Terror Fight,” by James Gerstenzang, Los Angeles Times, September 29, 2006.

I use this paper to slightly alter the question of resolve during crisis and test the effects of reputation somewhat differently, by focusing on the effects of reputations to honor or violate alliance commitments. As I argue in the text that follows, alliances hold several advantages over crises when testing the effects of reputation. Alliances are written public promises, and the meanings of alliance commitments vary little from one situation to the next. These two facts establish an interdependence of cases that is often lacking for examinations of crises. Moreover, alliances are associated with conflict but are also better insulated from the intense strategic selection found in crisis behavior. Nevertheless, despite these differences, the expectations for alliance reputations still follow the logic of reputations formed during crisis. Honored commitments should build credible reputations, increasing the likelihood that other leaders expect future commitments will be honored as well.

### **REPUTATIONS: RICH THEORY FINDING ONLY MIXED SUPPORT**

Leaders do seem to be concerned with their and their state's reputations. Mercer (1996: 2) quotes former US President Ronald Reagan as arguing that, if the US lost in Central America, "our credibility would collapse and our alliances would crumble" – a statement that echoes Nixon's concern over abandonment of Vietnam, "the cause of peace might not survive the damage that would be done to other nations' confidence in our reliability," and Truman's argument that failure in Korea "would be an open invitation to new acts of aggression elsewhere."

These types of statements are not limited to American leaders. Germany's Baron von Holstein remarked in 1905, "But if we allow our feet to be stepped on in Morocco

without a protest we simply encourage others to do the same elsewhere,” and a senior British official observed of the Falklands crisis, “If we can’t get the Argentineans out of the Falklands, how long do you think it will be before the Spaniards take a crack at Gibraltar?” (quoted in Mercer, 1996: 12 and 21).

The timing of these pronouncements, often made during the darkest days of a crisis or just prior to state interventions in difficult situations, suggests reputations may simply be used to rally attentive publics rather than representing actual factors in the decision calculus of leaders. But while a rallying effect does seem to follow these public addresses, there also seems to be ample evidence that leaders use reputations in their private decision-making. Press (2005: 13), for example, recounts the discussion of the Cuban crisis between US Secretary of State Rusk, US Secretary of Defense McNamara, and President Kennedy. Rusk argued:

Now suppose that they were to consider this a major back down, then this would free their hands for almost any kind of intervention that they might want to try in other parts of the world. If we are unable to face up to the situation in Cuba against this kind of threat, I think that they would be critically encouraged to go ahead and eventually feel like they’ve got it made as far as intimidating the United States is concerned.

And McNamara agreed:

It is not a military problem we’re facing. It’s a political problem. It’s a problem of holding the alliance together. It’s a problem of properly conditioning Khrushchev for our future moves.<sup>2</sup>

The public and private arguments of these decision-makers match well what most deterrence theorists argued throughout the Cold War. Demonstrating resolve is important in each crisis because, as McNamara pointed out, it would force the Soviets to respect as credible the threats of the United States. Consistent backing down leads enemies to

---

<sup>2</sup> Both Rusk and McNamara are quoted in the *Pentagon Papers* (May and Zelikow, 1997).

challenge more forcefully, and in more situations. The citations to this basic theory are numerous (see for example, Snyder, 1984; Jervis 1988; or the review of the literature in Huth, 1997), but Schelling's (1966- 124-25) famous assessment of the Korean War puts the argument most starkly:

We lost thirty thousand dead in Korea to save face for the United States and the United Nations, not to save South Korea for the South Koreans, and it was undoubtedly worth it. Soviet expectations about the behavior of the United States are one of the most valuable assets we possess in world affairs.

More sophisticated treatments of the reputation logic have been produced by formal theorists, both in economics and in political science. In economics, the ability of firm reputation to deter competition has been well analyzed (see Kreps and Wilson, 1982; Wilson, 1989; and Weigelt and Camerer, 1988), and political scientists have adopted these theories as tools in understanding the types of signals leaders can send (see for example, Alt, Calvert, and Humes, 1988; Ordeshook, 1986; and Wagner, 1992). Sartori (2002) and Guisinger and Smith (2002) probably go furthest in arguing that leaders and their envoys have incentives to develop certain types of reputations in order to overcome the uncertainty endemic to crisis diplomacy. In these models, a reputation for honesty allows the sender to credibly give information that would otherwise be "cheap talk", and thus, leaders may concede less important issues, without bluffing, in order to maintain a reputation for honesty when more important issues arise (Sartori, 2002: 122).

The sum argument of these statements and theoretical treatments is clear. Decision-makers argue and act, at least in part, based on reputations. Traditional deterrence theory suggests reputations should be pursued by leaders as important and manipulable tools, which are useful in future crises. Formal theorists agree; reputations provide valuable information when the costs of signaling are low.

But these theoretical arguments often fail when tested empirically, generating, at best only mixed support for the proposition that reputations matter. For example, Snyder and Diesing's (1977) early study found only one instance of decision-making based on past actions. Huth and Russett (1984) demonstrated that past behavior had no effect on future deterrence crises, while Huth's (1988) follow-up work found that reputations may be limited to ongoing rivalries, and even in these cases, the effects of past actions had relatively little effect, mostly limited to the cases prior to World War I (see also Fearon, 1994, for a re-evaluation of the Huth and Russett deterrence cases)..

Fearon (1996) and Smith (1996; 1998) have separately argued that the studies that do find at least some support for reputation (Huth and Russett, 1990; Huth, Gelpi, and Bennett, 1993; and Danilovic, 2001) are often plagued by selection effects. Crises are not random events, and leaders are likely to target those that have already demonstrated a lack of resolve. This makes it exceedingly difficult to grasp the effects of intangibles such as reputation in one crisis and the next. Even the most sophisticated models which take selection effects seriously are still heavily influenced by initial modeling assumptions (see Signorino, 1999, but especially Lewis and Schultz, 2003, and the discussion in Wand, 2005, and Schultz and Lewis, 2006), which makes it difficult for examinations of crisis behavior to give a definitive assessment of the influence of past actions (Press, 2005: 15).<sup>3</sup>

The qualitative literature testing deterrence has been even less charitable to theories of reputation, as few studies demonstrate that reputations matter more than situational variables such as interests and capabilities. Mercer's (1996) work represents

---

<sup>3</sup> A notable exception to the litany of mixed findings is Tomz's (2001) work on sovereign debt lending. Through careful data collection and analysis of lending patterns, Tomz finds that reputations for repayment affect both future interest rates and the availability of lenders.

one of the most comprehensive tests of reputation theory. Analyzing three separate military crises (the First Moroccan Crisis, Bosnia-Herzegovina, and the Agadir Crisis), Mercer argues that dispositional variables like reputation are almost always impossible to build. Most leaders, at least those in the three crises examined, attribute backing down to situational factors, and hence, reputations never form.

Press (2005: 16-17) argues that Mercer does not go far enough in establishing the relative worth of a reputation. Establishing the likelihood of a disposition forming is not the same as determining whether that disposition matters from one crisis to the next. Instead, Press analyzes his own set of “easy” cases for reputation theory (the appeasement of Germany prior to World War II, the crises over control of Berlin, and the Cuban Missile Crisis), and in each analysis, reputations mattered little in determining the credibility of threats:

Neither the senior members of the Eisenhower administration nor those in Kennedy’s government judged Soviet credibility on the basis of Khrushchev’s record for keeping commitments. Khrushchev’s bluffs never damaged his—or the Soviet Union’s—credibility in American eyes. And documents from the 1930s—the crises that have become symbols for the dangers of backing down and losing credibility—show that German assessments of British and French credibility before World War II were not affected much (if at all) by years of British and French appeasement (2005: viii).

What matters in these crises are the realist variables of power and interest, according to Press (2005: 143): “Decisionmakers believe that threats are credible when—and only when—they are backed by sufficient power and serve clear interests.”

Such divergence between rich theory and empirics does happen, but not often in the face of decisionmaker’s statements to the contrary. Given the difficulties of testing reputations using crisis, perhaps it is time to take a step back. So instead of looking for evidence of reputations in past crisis behavior, I examine the effects of failed and

honored alliance commitments. Alliances actually hold many advantages over crises when searching for the effects of past behavior. In fact, there are several reasons to think that international alliances may provide one of the best tools available to evaluate the effects of state reputation building.

### **TESTING THE EFFECTS OF STATE REPUTATIONS WITH ALLIANCES**

Alliances are, by nature, public signals of state intentions in the event of conflict.<sup>4</sup> Alliance treaties provide a record of what each state has contracted to perform and can therefore be evaluated against the actions revealed by conflict. While some treaties are purposefully vague, pledges of defense, neutrality, and non-aggression are much less ambiguous than the diplomatic language that couches threats during crises.

Alliances are also not that far removed from conflict. Unlike the “low politics” of debt repayment, trade, and other negotiations, alliances are associated with conflict, and most are followed by war (Gibler and Vasquez, 1997; Levy, 1981). Thus, alliances provide an opportunity to test the effects of reputation, with a large-N approach of truly similar cases, in an environment closest to the deterrence situations of most concern to practitioners and traditional theorists. Of course, while associated with conflict, alliance behavior is more removed from the selection effects that plague the immediate deterrence literature. The Cold War was riddled with standing alliances on both sides as the US and Soviets divided the world into spheres of influence, and many of the crises that erupted did so in an environment of already existing alliances. Thus, alliances may provide some leverage as reputation tests shift from immediate to general deterrence.

---

<sup>4</sup> I exclude secret alliances and the secret provisions of public alliances from the analyses that follow.



Alliances provide another benefit for testing theories of reputation that assume interdependence, mostly because the meaning of alliance provisions varies little across region. Almost all the formal work on reputations presumes interdependence across cases, but the empirical work on crises suggests this assumption is untenable for deterrence (Press, 2005; Huth, 1997; Mercer, 1996). Crisis deterrence is specific and often relative to the threatened state because the threats are intended to have an immediate effect and are responsive to the crisis at hand. In these cases, situational attributes dominate reputation building as geography and opponent constrain the ability to build a reputation of credibility that travels to other circumstances. Put differently, threats used to defend Taiwan might not carry similar meanings in a defense of Turkey or Argentina because the interests and capabilities of the states involved differ markedly in each situation. However, since coming to one's defense means the same thing whether the prior pledge is to Taiwan, Turkey, or Argentina, a violation of an alliance promise to aid Taiwan is likely to make Turkey and Argentina worry (Jervis, 1991).

Mercer (1996: 224-226) argues that alliances are also tacitly different from deterrence during crises because the costs of alliances are borne only after war breaks out. Promises such as alliances constitute nothing more than "cheap talk" compared with the costly pre-war signals of the threats implicit in troop movements and buildups (see also Fearon, 1996). This difference in cost distribution suggests that leaders may be especially sensitive to their alliance reputations. The difficulty of making allies believe the worth of a promise compels leaders to be attentive to all signals that reinforce the initial promise, especially the signals sent from honoring similar commitments to other alliance partners.

One twist of Mercer's theory of reputation suggests that positive reputations cannot be formed because the costs of honoring a commitment will be dismissed as consistent with the situational goals of the partner that intervenes. But negative reputations can form because the perceiver attributes alliance violations and other negative behavior to dispositional explanations like reputations (1996: 225). Obviously, this is a testable proposition, one which constitutes the first set of hypotheses:

H1. All else equal, leaders will try to ally with states that have honored their alliance commitments in the past;

H2. Leaders will avoid alliances with states that have violated their previous alliance commitments.

Most theories of reputation suggest that both hypotheses should find empirical support, but again, for Mercer, only H2 should be confirmed. If reputations do not matter, then neither hypothesis will accurately describe the data.

Related to this first set of hypotheses is a second set of expectations, one which considers more than the basic yes/no dichotomy of violated commitments. For this second set of hypotheses, I consider the relative worth of each state's past actions. Honored commitments may be more meaningful when those actions make some palpable difference in the conflict, while violated commitments in unwinnable conflicts are likely to be dismissed as prudent behavior, and hence, less important for reputation building.<sup>5</sup>

H3. The likelihood of honorable reputations forming increases with the relative worth of the honored commitment.

H4. Dishonorable reputations are most likely to be formed when the cost of a violation is greatest.

---

<sup>5</sup> Tomz (2001) finds that the level of difficulty in honoring commitments does matter for sovereign debt repayments – as loans repaid during tough financial times increase the likelihood that states will be perceived as good lending risks, attracting reduced future interest rates. But these results are somewhat anomalous for the debt literature (see for example, Eichengreen, 1991).

Alliance reputations can also be tested during crises. Smith (1995), for example, argues that the credibility of an outside alliance partner is a key determinant in whether a leader is willing to initiate a dispute. If past actions affect credibility, then reputable alliance partners should be able to dissuade potential attacks against their protégés. Disreputable alliance partners would be least able to provide credible signals capable of deterring these attacks.

H5. States are less likely to be targeted by a dispute if they have outside defense pacts with states that have honored their commitments in the past.

H6. States are more likely to be targeted by disputes when their alliance partners have violated commitments in the past.

Obviously, these hypotheses are not the equivalent of tests of dyadic reputation during crises but, if confirmed, would provide clear support for the expectation that leaders assess actor reputations prior to conflict.

### ***Alliances and Reputations: Measurement***

To test these hypotheses, I code six different alliance reputation variables according to whether there was a previous violation or honored commitment, and I differentiate among the dyad, region and rest of the international system when considering where the reputation-building action took place. To do this, I first code each state in an alliance at the start of an interstate war according to whether that state honored or violated the terms specified in the alliance – i.e., remained neutral, consulted, or defended an alliance partner.<sup>6</sup>

Second, I use the individual decisions to honor or violate each alliance as a proxy for each state's alliance reputation for the next 10 years. A state that violated an alliance

---

<sup>6</sup> These coding decisions replicate those described in previous works that have coded alliance reliability (Leeds, et al., 2000; Gartzke and Gleditsch, 2004).

in 1950, for example, is considered a “dishonorable” alliance partner from 1951 to 1960.<sup>7</sup> I then merged these 10-year reputation proxies with each state-year of a dataset of non-directed dyads, 1816-2000, discerning whether the alliance violation (or honored commitment) occurred in the other state’s region, within the international system outside the region, or within the dyad itself. So, building on the example above, if the United States was the state violating an alliance in 1950, then the US-Mexico dyads from 1951 to 1961 would be coded as having a previous alliance violation in the dyad. US-Canadian alliances would be coded as having an alliance violation in the region, and US-British alliances would be coded as having a previous alliance violation in the international system. The reputation scores are dummy variables only, not additive, so additional violations in the dyad, region or system carry no additional weight.

Because some violations are more heinous than others, just as the difficulty of some honored commitments varies substantially, I also code a separate variable for the relative worth of each alliance partner’s war fighting opportunity. To operationalize this variable, I divide the capabilities of each state-year by the difference in capabilities between each war-fighting state or coalition. Thus, a state-year with a CINC score of 0.20 that violated an alliance between two coalitions, one with a total CINC score of 0.40 and another with a CINC score totaling 0.60, would have a relative worth of 1.00 (or 0.20 for the state-year, divided by a difference in CINC scores of 0.20). Similarly, the relative worth of an honored commitment would also be 1.00, regardless of which side of the conflict yielded the honored commitment.

---

<sup>7</sup>I also conducted sensitivity tests with 5 and 20 year reputation thresholds, but the results do not differ from those reported here.

### ***Control Variables for the Alliance Formation Tests***

Gibler and Wolford (2006) argue that the inconsistencies of the alliance literature thus far have been products of differences in research design. Confusion over the dependent variable, the selection of cases, and differences in temporal domain has each led to confusing and often contradictory results. I therefore replicate Gibler and Wolford's basic model of alliance formation which also adopts the models and data used by Lai and Reiter (2000), Leeds, et al. (2001), and Gibler and Sarkees (2004). Following Gibler and Wolford, I also slightly alter the Lai and Reiter dependent variable in two simple ways. First, I updated Lai and Reiter's original data by using the updated Correlates of War formal alliance dataset (Gibler and Sarkees, 2004). Second, I slightly change the dependent variable to a measure that better captures alliance formation. To code this variable, I include all dyads forming an alliance during the year the original alliance treaty was signed, and I exclude joiner dyads. The North Atlantic Treaty Organization (NATO) provides a good example of these coding rules – the alliance was formed in 1949, and all dyads part of the alliance in 1949 are coded as *alliance formation* dyads for that year. All alliance formation dyads and all subsequent alliance dyad-years from 1950 on, including dyads that join NATO, are *allied* dyad-years and are not considered in the hypotheses testing alliance formation.

This consistency of approach has multiple advantages. Foremost among these is the ability to conduct comparative theory tests, assessing not only the statistical significance of the independent variables but also their relative effects. Consistency across models also facilitates easier and better interpretation of quasi-experiments since

changes in the relative effects of the baseline model is directly attributable to alterations in model specification.

The baseline alliance formation model includes a one-year lag that measures whether the dyad was allied in the previous year, and this variable should have a positive effect on alliance formation in all models. As a large number of dyads engaged in multiple, overlapping regional and security alliances, the formation of an alliance to cooperate on one issue will likely influence the willingness to cooperate on other issues. For example, the United States and Great Britain were allied via three different alliances during the late 1950s, 1960s, and early 1970s; the first alliance was dedicated to cooperation in Europe (NATO), the second in Southeast Asia (SEATO), and the third in the Middle East (Baghdad Pact) (see also Gibler and Wolford, 2006: 138, footnote 9).

I again keep all of Lai and Reiter's (2000) independent and control variables for the alliance formation analyses. First, the effects of regime type are assessed using measures of joint democracy and regime similarity in the dyad. Regime type is based on the 21-point combined autocracy-democracy scores from Polity IV. Joint democracy is dichotomous and is coded as one for all dyads possessing two states that score 5 or higher on this scale; regime similarity is the absolute difference of the two regime scores (2000: 213). The second group of independent variables captures cultural similarities in the dyad. Using the Cultural Composition of Interstate System Members dataset from the Correlates of War Project, dummy variables identify dyads with similar ethnicity, language, and religion (2000: 214). The third group of independent variables captures the level of joint threat experienced by the two states of the dyad. These variables include whether the states were on opposite sides of a Militarized Interstate Dispute

(MID), whether the states of the dyad had the same enemy in a MID, and the total number of MIDs experienced by each state in the dyad; each of these measures were coded using data from the ten years prior to each dyad-year (2000: 214). Finally, Lai and Reiter (2000: 215) specify variables for overall trade (the lower score of each state's exports and imports within the dyad divided by its gross domestic product), the presence of at least one major power, distance between capitals controlling for contiguity, and a learning measure based on Reiter's (1996: 84) earlier work.<sup>8</sup> Once again, I keep each of their variables, permitting an easier comparison to their results and the results reported in Gibler and Wolford (2006).<sup>9</sup>

#### *Control Variables for the Militarized Interstate Dispute (MID) Tests*

To test how alliance credibility affects the likelihood of being targeted by a dispute, I construct a dataset of directed, politically relevant dyad-years between 1816 and 2000. The dependent variable in these tests is whether the second state of the dyad is targeted by any type of MID initiation. I use Zeev Maoz's (2001) dyadic dispute dataset, and I code each case as targeted by a MID if the first state was revisionist and initiated a dispute. I conducted separate tests on "fatal" MIDs that experienced at least one fatality, but those do not differ from the results presented here.

---

<sup>8</sup> A state is coded as 1 if it had a lesson favoring alliance; -1 if it has a lesson favoring neutrality; and 0 if it had no lesson. A state is coded as having a lesson favoring neutrality if it was neutral during World War I or II and was not invaded or if it was allied during the war and was invaded. A state is coded as favoring alliance if it was allied during a world war and was not invaded or was neutral during the war and was invaded (Reiter 1996: 84). The dyadic learning variable varies from -2 to 2, according to the cumulative learning experiences of each state in the dyad.

<sup>9</sup> The trade variable narrows the temporal domain to 1950, and so I only report the results for the full model (1816-2000) which excludes trade. Results with the trade variable are no different from those reported in Gibler and Wolford (2006) and do not affect the results reported for the reputation variables.

The principle variables of interest in these tests are the reputation of states allied with the potential target state. I use five different independent variables to measure these reputations. First, a dummy variable captures whether the potential target has an alliance with another state that has, in the previous 10 years, honored another alliance commitment; a second dummy variable is used for alliances with states that have violated an alliance commitment, also within the previous 10 years. I use the same two continuous variables to measure the worth of each violation or honored commitment, and finally, I code a dummy variable if a target state has an alliance member which has built no reputation within the previous 10 years.

*Control Variables.* MID initiations obviously do not occur in a vacuum, and I therefore employ several standard control variables in these analyses. First, contiguity is a dummy variable, and I expect contiguous states to be more likely to have disputes. Jointly democratic dyads, also a dummy, should be more pacific, while those states at relative parity – a continuous variable of the weakest state’s capabilities divided by the dyad’s strongest state’s capabilities – are likely to conflict. Finally, a defense pact in the dyad – a dummy variable for Correlates of War Type I alliances (Gibler and Sarkees, 2004) – often denotes the tabling of threats within the dyad and is therefore likely to be peaceful.

## **REPUTATIONS, ALLIANCE FORMATION, AND MID AVOIDANCE**

The first set of analyses has alliance formation as the dependent variable and begins with a replication of the baseline model. Replicating the findings in Gibler and Wolford (2006), column 1 of Table 1 demonstrates once again that six variables are associated with an increased likelihood of alliance formation – another alliance in the



previous year, shared religions and languages in the dyad, a joint enemy, major power status, and lessons learned from formative events. The five variables associated with a statistically significant drop in the likelihood of alliance formation – joint democracy, polity differences, joint ethnicities, large amounts of threat to the dyad, and increased distances – also retain coefficients and standard errors identical with the previously reported results.

\*\*\*\*\*TABLE 1 ABOUT HERE\*\*\*\*\*

Column 2 adds the set of dyadic reputation variables to the baseline model. Quite surprisingly, dyadic violations in the previous 10 years have no statistically significant effect on alliance formation, and even if the coefficient were significant, the sign for the relationship is positive. This finding makes more sense when considered in conjunction with the relative worth of violation variable, which is significant ( $p < 0.10$ ) and negative. Thus, the actual violation matters less than whether the potential intervention could have made a difference. Both dyadic honorable reputation variables are statistically significant. States are more likely to ally with those states that have honored their dyadic commitments in the past, all the more so if those commitments meant a major difference in the conflict.

All but two of the original predictors of alliance formation remain virtually unchanged from the baseline model. The two variables affected are the alliance lag and the learning variable. First, the addition of the dyadic reputation variables decreases the marginal effect of a previous alliance by more than 50%.<sup>10</sup> For the second variable, the

---

<sup>10</sup> I calculated this effect by toggling the alliance lag between 0 and 1 in each model, holding all other variables at their mean. The probit coefficients can sometimes be deceiving. For example, the coefficient for the joint ethnicity variable decreases almost the same amount as the alliance lag; however, the marginal

coefficient for learning is cut in half while the standard error almost doubles. I believe that both these results lend added support to the validity of the reputation variables. Of all the independent variables in the baseline model, only these two attempt to measure the historical dependence within the dyad, and since the effects of the reputation variables are concentrated principally on these two variables, these results suggest that reputation is also capturing at least a portion of that historical dependence. Reputation matters more than previous formative events, at least for the dyadic learning variable, and even the brute force measure of a lag dependent variable is affected by the addition of reputations.

Column 3 reports that only one of the regional reputation variables is statistically significant. Honoring a previous commitment in one region will apparently make a state more sought after in that region. Neither the cost of that honored commitment nor any type of alliance violation alters the likelihood of future alliance opportunities.

As I report in the fourth column of Table 1, the system-level reputation variables have a great effect on alliance formation. Honored commitments and previous violations are both statistically significant, and their signs are in expected directions. Further, costly involvements have a positive effect on future alliance formation.

The variegated results for the dyadic, regional, and system-level variables demonstrate some interesting differences in how reputations may be perceived. In the dyad, where information about the alliance partner's commitment is greatest, the cost of the violation matters more than the actual violation. But in the system, when information

---

effect of ethnicity remains virtually unchanged across the models. (All these results are available from the author.)

about the specific commitment is most likely lowest, only the dichotomous violation variable has an effect.<sup>11</sup>

The differences in information flows is also apparent when considering the changes in the temporal control variables (the alliance lag and the learning variable) as the reputation variables move from dyad, to region, to system. The alliance lag has the greatest effect in analyses using regional or system-level reputations, and the learning variable is statistically significant only in the system-level reputation model, though even in this case, the marginal effect is half as strong as in the baseline model.

\*\*\*\*\*TABLE 2 ABOUT HERE\*\*\*\*\*

I use Table 2 to describe the marginal effects of each reputation variable. As I describe above, one of the key benefits of using similar research designs across studies is the ability to conduct comparative theory tests, and I do so here by assessing the relative effects of reputation against established predictors of alliance formation. The likelihood of alliance formation in the dyad is quite small; though more numerous than dyadic conflicts, alliance formation is still exceedingly rare. Thus, Table 2 presents the results as percent changes from the base likelihood of alliance formation.

The probability of alliance formation for the baseline model is 0.01% for any given dyad-year. The probability of alliance formation increases by 0.26% if the dyad has the same language and decreases by 0.05% if the dyad is jointly democratic. Increasing the amount of threat against the dyad has a seemingly perverse effect – the likelihood of alliance formation decreases by 0.04%. Note that these numbers are indeed fractions of a percentage point and not some typographical error with a misplaced

---

<sup>11</sup> This is true for all time periods. In results that I do not present here, separate analyses using only Cold War dyads or pre-1945 dyads both produce almost identical findings to results using the entire dataset.

decimal point. Not only is alliance formation rare, we also have very few good predictors of what correlates with alliance formation in the dyad.<sup>12</sup>

The effects of reputation are large when compared with the baseline predictors of alliance formation. In the dyadic reputation model, honored commitments increase alliance formation likelihoods by 0.94%; alliance violations decrease these likelihoods by 4.53%. In the model with system-level reputations, the results are in the same direction for these dichotomous variables, but the effects are weaker by an order of ten. Real marginal differences occur only when considering the worth of honored commitments. Moving from the mean honored commitment to situations in which the ally made the maximum difference in a war, the likelihood of alliance formation for that state increases by over 80% if that war opportunity occurred in the dyad and over 66% if it occurred elsewhere in the system. Obviously, these marginal effects dwarf the effects of both the control variables and the dichotomous variables.<sup>13</sup>

The clear conclusion from these results is that reputations matter, at least for alliance formation. Leaders pay attention to the past actions of other states when considering their alliance partners; leaders seek states that honor their commitments and avoid those states that do not. These results confirm both Hypotheses 1 and 2 and provide moderate support for the propositions that form the basis for Hypotheses 3 and 4. The next step is determining whether these alliance reputations have an effect on crisis

---

<sup>12</sup> I initially thought that the alliance lag variable was overwhelming the predictive abilities of the other variables in the baseline model. However, the marginal effects of the variables in Table 2 are only slightly greater when based on a model that excludes the alliance lag.

<sup>13</sup> The differences in marginal effects are not due to differences in the scale of these variables. The large marginal effects do not change for models that replace these reputation variables with standardized variables that I constrained to vary between 0 and 1.

behavior – does a reputation of upholding alliance commitments affect the credibility of alliance threats?

\*\*\*\*\*TABLE 3 ABOUT HERE\*\*\*\*\*

Table 3 presents results from several models that predict whether a state is likely to be targeted by a dispute. Using directed dyads with varying temporal domains, the dependent variable is whether the second state of the dyad was targeted by any type of MID initiation.<sup>14</sup> I use common controls for predicting MIDs – defense pacts within the dyad and joint democracy decrease the likelihood of onset while parity and contiguity increase the chance of conflict. All these variables are statistically significant and in the expected direction.

The reputation variables in these models assess what type of defense pact partner each state has.<sup>15</sup> If reputations affect the credibility of a commitment, then outside alliances with states that have previously violated alliances should be less credible and, hence, less likely to deter MID targeting. This is exactly what I find: disreputable alliance partners are correlated with an increased number of disputes targeting the state. But that result seems limited to alliance violations only. The dichotomous honored commitment variable has no effect, and while the interaction of honored commitment and the worth of commitment does affect conflict initiation, the result for this interaction is an unexpected direction. Allies that provide the bulk of capabilities when honoring their commitments are correlated with an increase in the likelihood of MID initiation.

The honored commitment results may be more complex than they first appear. First, if leaders choose allies based on their reputations for upholding alliances in the

---

<sup>14</sup> Results using “fatal” MIDs only do not differ from any of the results presented in Table 3.

<sup>15</sup> The results again do not vary substantially if I use all alliances rather than just defense pacts. Reputations are based on all previous alliance commitments.

past, as Tables 1 and 2 demonstrate, then it also seems likely that, in times of intense danger, these allies will be even more sought after. Thus, a selection effect may be present. States that honor difficult commitments are brought into crises that are likely to find deterrence failure. It could also be the case that these allies, though sought after, do not provide many advantages for extended deterrence. The very fact that these states could not deter previous conflicts, even though their capabilities greatly outweighed the capabilities of the combatants, suggests opponents might conclude the ally is exceptionally weak or perhaps even exhausted from the previous conflict.

Alliance reputations apparently matter, especially when compared with the baseline of outside defense pact partners with no reputation. To assess the marginal effects of these reputations, I compare each variable to the baseline controls and report these results in Table 4.

\*\*\*\*\*TABLE 4 ABOUT HERE\*\*\*\*\*

Once again, the likelihood of conflict targeting is exceedingly rare; in the full model (1816-2000), the likelihood of “state b” being targeted by any type of dispute is 1.13% in any given directed dyad-year. The control variables alter this probability only slightly. Contiguity increases the chance of a dispute by 2.74%, while complete shifts from preponderance to parity lead to a 2.09% increased dispute risk. Defense pacts within the dyad decrease dispute likelihood by only 0.30%

The alliance reputation variables have effects similar to the variable measuring defense pacts within the dyad. Outside defense pact allies with no reputation decrease the chance of targeting by 0.60%, but disreputable allies are correlated with 0.90% increase in dispute initiation. This last probability is 3 times the effect of the intra-dyad

defense pact variable, and alliances are manipulated much more easily than either contiguity or capabilities.

## **CONCLUSION**

The results of this paper confirm that state reputations have an effect on state decisions to ally and initiate conflict. States that have honored their commitments in the past are more likely to find alliance partners in the future, just as alliance violations decrease the likelihood of future alliance formation. Alliance reputations also matter in disputes. States with disreputable outside alliance partners are more likely to be targeted by rival states. Each of these results is consistent across time period and across model specification.

The differences between the alliance formation models and the dispute initiation models is interesting and may underscore the difficulty in establishing the relative worth of reputations during times of intense hostility. Reputations in the alliance formation models elicit much stronger marginal effects than the dispute targeting models. I think this highlights the overwhelming effects that situational variables like power and interest can have on rival decision-making during times of crisis. When hostilities are lower, reputations provide information about the intentions of actors. But when hostilities are highest, reputations matter less than the ability and willingness to fulfill threats against the state.

**Table 1: Probit analyses of the effects of reputation on alliance formation for all dyad-years, 1816-200**

	Model 1	Model 2	Model 3	Model 4
Dependent Variable:	Alliance Formation	Alliance Formation	Alliance Formation	Alliance Formation
Time Period:	1816-2000	1816-2000	1816-2000	1816-2000
<u>Reputation Measures†</u>				
Alliance Violated within last 10years...				
Violation of alliance within the dyad		.061(0.162)		
Violation in dyad X worth of involvement		-3.771(2.241) *		
Violation of alliance in region by one state of dyad			-.072(0.140)	
Violation in region X worth of involvement			-3.073(2.417)	
Violation of alliance in system by one state of dyad				-.162(0.089) *
Violation in system X worth of involvement				-1.431(2.650)
Alliance Honored within last 10years				
Honored alliance within the dyad		.778(0.107) ***		
Alliance in dyad honored X worth of involvement		4.247(1.886) **		
Honored alliance in region by one state of dyad			.221(0.083) **	
Alliance in region honored X worth of involvement			4.387(2.897)	
Honored alliance in system by one state of dyad				.302(0.058) ***
Alliance in system honored X worth of involvement				4.082(2.095) **
<u>Allied in previous year</u>	.178(.033) ***	.149(.044) ***	.209(.047) ***	.208(.045) ***
<u>Regime Type</u>				
Joint Democracy	-.271(.043) **	-.273(.056) ***	-.360(.061) ***	-.379(.063) ***
Polity Difference	-.013(.002) ***	-.011(.003) ***	-.013(.003) ***	-.011(.003) ***
<u>Culture</u>				
Joint Religion	.337(.027) ***	.322(.033) ***	.301(.035) ***	.308(.035) ***
Joint Language	.444(.034) ***	.440(.043) ***	.440(.045) ***	.443(.043) ***
Joint Ethnicity	-.165(.047) ***	-.095(.042) ***	-.116(.070)	-.083(.068) ***
<u>Threat</u>				
Conflict Relations	.040(.048)	.068(.082)	.172(.085) **	.158(.084) *
Joint Enemy	.636(.028) ***	.631(.045) ***	.647(.047) ***	.645(.048) ***
Amount of Threat	-.012(.003) ***	-.009(.004) ***	-.001(.005)	-.003(.005)
<u>Other Controls</u>				
Distance	-.012(.000) ***	-.011(.001) ***	-.011(.001) ***	-.012(.001) ***
Major Power	.287(.035) ***	.330(.049) ***	.394(.055) ***	.331(.052) ***
Learning	.093(.019) ***	.047(.029)	.044(.031)	.056(.031) *
Constant	-2.345(.040) ***	-2.413(.056) ***	-2.433(.062) ***	-2.411(.062) ***
N	411,033	411,033	411,033	411,033
Wald c <sup>2</sup> (12)	3566.120 ***	2168.260 ***	2261.350 ***	1871.850 ***
Pseudo R <sup>2</sup>	0.215	0.285	0.256	0.263

†An alliance violation/honored alliance is a dummy variable capturing the performance of either state in the last 10 years of alliances in the dyad, region or system. The relative worth of the alliance performance is a continuous variable, with the capabilities of the honoring/violating state divided by the difference in capabilities between the stronger and weaker side in the conflict. See text for more detailed explanations of both of these measures.

\*p<0.10; \*\*p<0.05; \*\*\*p<0.01



**Table 2: Marginal Effects of Reputation on Alliance Formation**

*Change in probability in a dyad for each of the following events (measured from the base probability of alliance formation):*

*Control variables--*

Dyad becomes jointly democratic:	-0.05%
Dyad has same language:	0.26%
Number of MIDs involving both states changes from Mean to Maximum:	-0.04%

*Reputation and alliances--*

One state in the dyad...	
<b>Honored</b> an alliance in the dyad in the past 10 years:	0.94%
<b>Honored</b> a dyadic alliance (last 10 years), and provided a maximum difference in the war:	80.77%
<b>Violated</b> a dyadic alliance (last 10 years) but could have provided maximum difference in war:	-4.53%
<b>Honored</b> an alliance in the system in the past 10 years:	0.14%
<b>Honored</b> an alliance in the system (last 10 years), and provided a maximum difference in the war:	66.22%
<b>Violated</b> an alliance in the system (last 10 years)	-0.04%

*Note: Except for the system-level reputation variables, all probabilities are calculated using the results presented in Table 1, column 2 (all non-directed dyad-years, 1816-2000). System-level probabilities are based on Table 1, column 4.*

**Table 3: Probit analyses of the Likelihood of being Targeted by a Militarized Interstate Dispute (MID)**

**Dataset: Politically Relevant, Directed Dyad-Years**

**Dependent Variable: State B is target of a MID**

**Temporal Domain:**

	Model 1 1816-2000	Model 2 1816-1945	Model 3 1946-2000
<u>Reputation and Alliance Measures†</u>			
Defense pact in the dyad	-0.117(.061) *	-0.103(.087)	-0.162(.081) **
<i>Potential Target State has a...</i>			
defense pact partner with <b>NO</b> reputation	-0.193(.047) ***	-0.336(.061) ***	-0.095(.067)
defense pact partner with <b>HONORABLE</b> reputation	0.050(.043)	0.041(.058)	0.090(.061)
defense pact partner with <b>DISHONORABLE</b> reputation	0.263(.045) ***	0.549(.072) ***	0.116(.058) **
defense pact partner with <b>HONORABLE</b> reputation x Worth of conflict	0.005(.001) ***	0.009(.002) ***	0.003(.001) **
defense pact partner with <b>DISHONORABLE</b> reputation x Worth of conflict	-0.001(.002)	-0.087(.099)	0.001(.002)
<u>Other controls</u>			
Contiguous dyad	0.609(.052) ***	0.391(.071) ***	0.791(.070) ***
Capability Ratio (Weaker/Stronger)	0.473(.073) ***	0.757(.105) ***	0.305(.095) ***
Joint democracy	-0.358(.060) ***	-0.284(.102) **	-0.367(.070) ***
Constant	-2.354(.037) ***	-2.469(.200) ***	-2.429(.060) ***
<i>N</i>	176,042	70,934	105,108
<i>Wald chi-square</i>	359.680 ***	224.810 ***	256.130 ***
<i>Pseudo R2</i>	0.092	0.078	0.114

\*\*\* $p < 0.001$ ; \*\* $p < 0.01$ ; \* $p < 0.05$

Robust standard errors are reported in parentheses next to the coefficients.

**Table 4: Marginal Effects of Reputation and Alliances on MIDs**

*Change in probability in a dyad for each of the following events (measured from the base probability of MID targeting):*

*Control variables--*

Dyad changes from non-contiguous (major-minor) to contiguous: 2.74%

Dyad changes from preponderance (0.01) to parity (0.99), weaker/stronger: 2.09%

Dyad becomes jointly democratic: -0.79%

*Reputation and alliances--*

Potential target has defense pact with potential initiator: -0.30%

Potential target has external defense pact partner...  
with no reputational history: -0.60%  
that previously **violated** alliance: 0.90%  
that **honored** an alliance, and provided a maximum difference in the war: 0.20%

*Note: all probabilities are calculated using the results presented in Table 3 (all politically relevant dyad-years, 1816-2000).*