Deterrence by Diplomacy

Anne E. Sartori

Princeton University[1]

October 31, 2002

[1]Preliminary and incomplete. Comments are welcome and can be sent to asartori@princeton.edu.

Chapter 4

Reputations for Honesty and the Success of Diplomacy

The earlier chapters present a theory of why diplomacy is effective. I show that diplomacy is a boon to states, and that defenders often use it to achieve their goals. This chapter tests these ideas empirically. I examine a data set containing information about a large number of international interactions to see whether the big implications of the model are borne out. To do so, I use statistical models to assess how much the probability of an outcome – such as the success of deterrence – goes up or down, depending upon the defender's reputation.

The chapter begins by using the formal theory of Chapter Three to derive empirical implications about the escalation or de-escalation of international disputes. I test two central implications: First, a defender is more likely to succeed in deterring an attack when it has a reputation for honesty. Second, a defender is more likely follow through on its threats when it has a reputation for honesty. After discussing these implications, I describe the data set, which contains information

about international interactions between 1816 and 1993, including approximately 1300 international disputes. I discuss the variables that I use to test the theory and briefly explain my statistical technique.

The chapter then turns to the empirical results. The patterns in the data are consistent with the theory:

- A typical defender with a reputation for honesty is more likely to deter an attack than a defender with a reputation for bluffing by roughly 21.5 percentage points.[1]

- A typical defender with a reputation for honesty is more likely to follow through on its threats than a defender with a reputation for bluffing by roughly 5.5 percentage points.

These results increase confidence in the theory as an explanation of states' behavior in international disputes. However, the skeptical reader might be concerned about the robustness or interpretation of these results. A final section of the chapter demonstrates that the results are robust along four main dimensions. First, the conclusions remain the same with the inclusion of several control variables that are suggested by other works on international crises. Second, they remain the same if I use either of two alternative definitions of a reputation for honesty. Third, the results cannot be due simply to differences among states in a propensity towards disputatiousness. Finally, my measure of a reputation for honesty does not simply capture the idea of a reputation for resolve.

[1] The reputations variable that I use has many values, as I explain later. By a defender with a reputation for bluffing, I mean here a defender that has more of a reputation for bluffing than 90 percent of other defenders.

Having shown the usefulness of my theory for understanding international disputes and crises, I turn in the next chapter to an investigation of empirical results that are not related to testing the theory. In that chapter, I first examine the effect of the defender's reputation for honesty or for bluffing on the challenger's decision about whether to start a dispute and the defender's decision about whether to try to deter an attack. I then investigate the effect of the balance of forces on dispute escalation.

4.1 Central Empirical Implications of the Formal Model

In Chapter Three, I conceptualize an international dispute in four steps: the challenger's decision about whether or not to threaten the use of force, the defender's decision about whether or not to counter-threaten (try deterrence), the challenger's decision about whether or not to attack after hearing the defender's threat, and the defender's decision about whether or not to defend if attacked. The equilibrium analysis focused on the defender's ability to communicate. The testable implications are about the states that are challenged and about their deterrent threats, rather than about the threat by the challenger that initiates the crisis.

In the equilibrium, a defender "babbles" when it has a reputation for bluffing; it uses diplomacy without care because it does not have a good reputation to lose. Having threatened, such a defender is more likely to back down. A defender with a reputation for honesty, on the other hand, tries to maintain that reputation; it is more likely to follow through on its threats. For example, when President Johnson

told the Soviets in 1968 that the United States was committed to the defense of Berlin, though not of Czechoslovakia, Johnson's statement was meaningful; the United States was likely to fight if the Soviets attacked Berlin (National Security Council 1968). This logic leads to the following implication:

Implication 1. A defender with a reputation for honesty is more likely to defend following a deterrence failure than is a defender with a reputation for bluffing.

I define reputations as in the game-theoretic model of Chapter Three:

Definition 1. A defender acquires or increases a reputation for bluffing if it threatens and backs down when a challenger attacks. It maintains a reputation for honesty if it acquiesces to the challenger's demands; if it threatens, the challenger attacks, and it follows through on its threats to fight; or if it engages in a successful bluff (the challenger threatens to fight and it backs down).

A challenger considers the consequences of an attack, and a defender with a reputation for bluffing is relatively good gamble. Because a defender with a reputation for bluffing is more likely to back down if deterrence fails, a challenger is more likely to attack such a defender. For example, China in the Korean War was a relatively good bet – having bluffed once, it was more likely to be bluffing again. (Tragically for the United States, this was not a bet that paid off.) Since, on average, defenders with reputations for honesty are more likely to follow through on their threats to fight, challengers are more likely to back down after hearing their threats. This logic leads to the following implication:

Implication 2. A challenger is more likely to attack following the defender's attempt at deterrence when the defender has a reputation for bluffing. (The defender's attempts at deterrence are more likely to fail when it has a reputation for bluffing.)

My theory also suggests that the balance of forces affects the progression of international disputes, but it does not provide a precise prediction about the form of the effect, as I discuss further later. Thus, I simply test the hypotheses that the balance of forces affects each stage of dispute escalation.

4.2 Data and Methodology

This section describes how I use information about states' interactions to evaluate my theory. After noting my sources of information, I discuss the measurement of the steps of dispute escalation: threat, counter-threat, attack, and defense. I also discuss the measurement of concepts that I use to explain dispute escalation, focusing on the prime variable of interest, the defender's reputation for honesty. I then explain the statistical technique that I use to analyze this information. In the next section, I turn to the main results.

4.2.1 Sources of Information

To examine the effects of reputations on the escalation of disputes, I use information from the Correlates of War (COW) project. These data, which are widely used to study international disputes and crises, include information about more

than 200 states over the course of almost 200 years.[2]

To measure the dependent variables, I use COW's Militarized International Dispute (MID) data set (Gochman and Maoz 1984; Jones, Bremer, and Singer 1996). The creators of the data set define a MID as an event in which at least one state took overt militarized action against another, where a militarized action may be as minor as a threat to use force or as major as a full-scale war (Jones, Bremer, and Singer 1996, 168-177). The action in question must be taken by the official military forces or governmental representatives of a state. MIDs are quite varied in terms of the issues involved and the degree of force used; examples include the Falklands War and British efforts to stop the Brazilian slave trade (Jones, Bremer, and Singer 1996, 178).

The data set contains information about each Militarized International Dispute that occurs between the years of 1816 and 1993. This complete coverage is crucial for evaluating the efficacy of diplomacy, since I must have information about each state's recent history of disputes in order to ascertain its reputation. I augment these data with information about which pairs of states existed in the international system in a given year. These pairs, or "dyads," also had outcomes in these years, though these outcomes were non-events. (For example, neither state began a militarized dispute with the other.) I use all of these MIDs for the coding of reputations (see below). However, because of the format of the MID data set, the same MID may appear several times if it is ongoing over several years. In order to avoid artificially inflating the number of independent observations in the data set, I use only the first year of an ongoing MID when I estimate the causes of dispute

[2]In merging and manipulating the data, I use the helpful computer program *EUGene: Expected Utility Generation and Data Management Program* (version 2.10), created by Bennett and Stam (see Bennett and Stam (2000)), and available at <http://www.eugenesoftware.org>.

escalation.[3] In other words, I use only one observation per MID for estimation.

To measure the independent variables, I again use MID data, but supplemented by information from the COW National Capabilities data set. The National Capabilities data set contains information about the military capabilities of states that are not involved in disputes, as well as of those that are, and thus provides information relevant for investigating the initiation, as well as the progress, of disputes.

To test the implications of the theory, I identify one state in a dispute as "challenger" and the other as "defender." Several states may be involved in any international incident; however, each observation in the data set corresponds to one pair of states (dyad). Within a disputing pair, one state is on the side that initiated the dispute, and the other is a target. If there is a MID, I define the challenger as the state on the initiating side of the dispute and the defender as the state on the target side. In a year with no MID, any state could have initiated a dispute. Thus, in any year with no MID, each pair of states is represented by two observations, one with the first state listed as the (potential) "challenger" and the second as the (potential) "defender," and the other with the states reversed.

[3] I include information about ongoing MIDs for purposes of coding reputations so that I do not mistakenly code a state as having a reputation for honesty because I have dropped information on its (ongoing) MID-year. After coding reputations, I keep additional observations based on an ongoing MID only if a state initiates a new dispute within a MID. When missing data and a few other issues also are taken into account, the analyses are based upon about 1300 MIDs.

4.2.2 The Dependent Variables: What is Deterrence Success? What Does it Mean for the Defender to Follow Through on its Threats?

The crucial implications of the theory concern two decisions. First, if the defender tries deterrence, does the challenger nevertheless decide to attack (resulting in a deterrence failure), or does it back down? Second, if the challenger decides to attack, does the defender follow through on its deterrent threat and fight, or does it back down? As I discuss later, one cannot investigate these decisions without examining the decisions that are temporally prior to them in a dispute: the challenger's decision about whether or not to initiate a dispute, and the defender's decision about whether or not to try deterrence. Conceptually, these decisions result in four binary outcomes, corresponding to the four steps of a dispute that I discuss in Chapter 3 (e.g. no threat by challenger/threat by challenger).

To measure these concepts, I use variables in the MID data set that describe the highest "hostility level" reached by each state in a dispute. The variables contain five possible hostility levels, ranging from the absence of militarized action to war.[4] If the challenger took an action that is at least as "hostile" as a threat to use force, I describe it as having threatened. Only if the challenger threatened do I code the variables that are temporally subsequent in a dispute. I describe the defender as having tried deterrence if that state, in turn, took an action that was at least as hostile as a threat to use force. Again, only if the defender threatened do I code the subsequent variables. I describe the challenger as having attacked if it used force and/or went to war. If the challenger attacked, I code the defender

[4]The hostility levels are (in increasing order): no militarized action, threat to use force, display of force, use of force, and war.

	Hostility level	Coded outcome
	chal. at least threatened	chal. threatened
If chal. threatened:	def. at least threatened	def. threatened
If def threatened:	chal. at least used force	chal. attacked
If chal. attacked:	def. at least used force	def. fought

Table 4.1: Dispute actions

as having fought if it also used force and/or went to war. Table 4.1 shows this method of coding.

Though the MID data provide useful information for testing the theory, they are not perfect for my purpose. First, because the data set does not provide information about a state's actions prior to the one with the highest hostility level, one cannot know with certainty that a state that went to war threatened before doing so. Second, the data set does not contain information on the dates of dispute actions. For example, in a certain dispute, I might know that the defender's most-hostile action was to threaten the use of force, while the challenger went so far as to use force. The data set does not reveal with certainty that the defender tried deterrence and deterrence failed; it is possible, for example, that the defender threatened the use of force after the challenger attacked. I infer the actions that a state took prior to the one with the highest hostility level and the temporal progression of the dispute.

Note that if the progression of a dispute matches the broad outlines of my model, then the inferences that I make about the progression of the dispute (e.g. that the defender tried deterrence before fighting, if it fought) are correct. While it would be nice to test my theory using only cases that match it closely (cases to

which it very clearly applies), the MID data are the best available for my purpose because of their in-depth coverage of all disputes in a time frame and the fact that they provide a lot of information about crisis behavior.[5] If the results of the data analyses were inconsistent with my theory, one might attribute this to the data being imperfect for the test. However, I find the patterns implied by my model in the MID data. It seems likely that if the inclusion of these other cases has any effect, it is to understate the importance of diplomacy in international disputes.

4.2.3 The Independent Variables: Reputations and the Military Balance

Having described how I measure outcomes, I now turn to discussing the factors that influence the escalation of disputes. I discuss the factors that are important in the model – reputations and the military balance – and then discuss issue of including control variables.

Measuring Reputations

I use the four dependent variables I describe earlier to create a variable that describes a defender's reputation at the start of the present interaction. The four dependent variables together characterize five possible dispute outcomes: 1) there is no dispute (neither state becomes a challenger by threatening the use of

[5] I do discard some cases based on the information in the MID data set. For example, I discard cases in which the state listed as being on the initiating side did not at least threaten the use of force but the state listed as being on the target side did so. I discard these cases because these facts strongly suggest, though they do not indicate definitively, that the initiating state was not really the challenger.

force); or a dispute exists (one state threatens) and: 2) the defender does not try deterrence, 3) the defender threatens and the challenger backs down, 4) the defender threatens, the challenger attacks, and the defender backs down, and 5) both states threaten and follow through on their threats to use force.

In the equilibrium that I examine in Chapter 3, a defender acquires a reputation for bluffing if it bluffs and is caught – that is, if it threatens, deterrence fails, and it backs down (outcome 4 in the previous paragraph). I define a state as increasing its reputation for bluffing in a given year if it is the defender in a dispute and this is the outcome.[6] As in the game theoretic model, a state maintains a reputation for honesty in one of three ways. The first corresponds to the fading of a reputation: if a state is not involved in a dispute in a given year, it cannot acquire a reputation for bluffing; if it has such a reputation, it fades. The second is acquiescence: if a defender does not try deterrence (acquiesces), it maintains its reputation for honesty. The third is following through on its threats.

The theory implies that a state's reputation is temporary, but provides no other guidance as to the duration of a reputation for bluffing. In the formal analyses, a reputation for bluffing is binary: a state either has one or it does not. In the real world, memories of past events fade slowly over time; having been caught bluffing a few years in the past probably still affects a state's reputation today, but less so than having been caught bluffing a year ago. To incorporate the idea that memories of called bluffs fade over time, I define reputations in such a way that any called bluffs in the previous ten years add to a state's current reputation for

[6]Technically, the defender acquires a reputation for bluffing in the equilibrium only if the challenger is listening – that is, only if it began the interaction with a reputation for honesty. I operationalize reputations more simply, designating a state as acquiring a reputation for bluffing whenever it is caught bluffing.

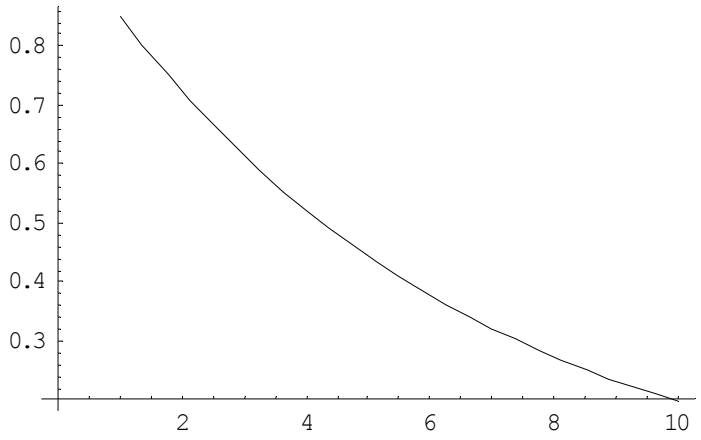

Figure 4.1: Coding of the Effect of Reputation with Years Passed

bluffing, but a bluff farther in the past has less of an effect on today's reputation than one that is more recent.

Specifically, a particular called bluff in the past ten years adds $.85^{\alpha}$ to a state's reputation for bluffing, where α is the time passed since the bluff was called. Figure 4.1 shows the amount that a particular called bluff adds to a state's reputation for bluffing under this coding, where the number on the horizontal axis represents the number of years that have passed since the state was caught in that particular bluff. For example, if a defender bluffed and was caught in each of the previous two years, but not in the eight years prior to them, its reputation would be coded as$.85 + .85^2 = 1.5725$. The variable ranges from approximately zero to 3.83. In the discussion that follows, I sometimes refer to the effect of a defender having "a reputation for bluffing" or "a reputation for honesty." Since the reputation variable has many values, this terminology is technically incorrect. What I mean is that the defender has *more of* a reputation for bluffing or *more of* a reputation for honesty.

Moving from the formal model to empirical definitions of reputations requires some judgement calls. Later in the chapter, I explore the consequences of defining reputations differently. The first alternative definition is simpler: a defender's reputation is based on whether or not it was a defender in the previous year and was caught bluffing in that year. The second is more complicated: a defender's reputation is based upon past behavior when it was a challenger, in addition to its behavior when it was a defender.

Military Power

Like much of the literature on international crises, my model of international interactions implies that the balance of military capabilities affects the course of international disputes. However, the model does not have robust implications about the directions of the effects; slight perturbations of the model imply different relationships. Since the variable appears in the formal model, I include a measure of the balance of military capabilities between the challenger and the defender (or between the two states if there is no dispute) in the statistical model. Though the estimates are not tests of the theory, they do give potentially important information about the escalation of disputes. I base the measure of the military balance variable upon the National Capabilities Index, a composite indicator of a state's military strength (Singer, Bremer, and Stuckey 1972). This index, which ranges between zero and one, is based upon a state's fraction of the total capabilities in the international system in six areas: iron and steel production, military personnel, urban population, total population, military expenditures, and energy production.

Rather than total capabilities, most theories suggest that the relative strength

of the challenger and the defender affects the progression of a dispute. For example, Huth (1988)[41] argues that the stronger a challenger relative to the defender, the more likely deterrence is to fail. My theory also suggests that the extent to which the challenger is favored affects the course of disputes. To ascertain the impact of the challenger's relative military strength, I create a log difference of the two states' capabilities. The log accounts for the fact that a challenger 110 times as strong as the defender is not much more likely to attack than one 100 times as strong, but a challenger 10 times as strong is much more likely to attack than one of similar strength.

Realists have long debated whether states are more likely to become involved in disputes and to fight when they are evenly matched militarily (a situation of "power parity" or when one state is much stronger than the other (see, e.g., Bremer (1992), Weede (976)). While power parity is not a part of my theory, I verify that my results about reputations do not change when the equation includes a variable measuring power parity instead of one measuring the extent of the challenger's military advantage. Following Schultz (2001), I include a measure of parity that is the military strength of the weaker state divided by that of the stronger.

Other, Irrelevant Variables?

It has become customary in statistical analyses of international relations to include a number of variables that are irrelevant to the theory being tested; these are "controls" suggested by competing or complementary theories. In this work, I take two approaches: I base most of my analyses on a simple model that includes only the concepts of my theory. However, to convince readers who disagree with this method, I also report results from analyses that include control variables

suggested by the literature on international crises.

The simpler model is the more informative. Including a long list of variables often is a substitute for careful thought about the factors implied by the theory. Moreover, these extra variables do not belong in the model; including them may mislead or confuse the researcher about their own "effects" and those of the explanatory variables of interest.

For example, according to current practice, the debate among realists over the manner in which relative power influences dispute escalation would speak to including two variables relating to the military balance in the statistical analyses: one that measures the similarity of the two sides' military strength, and another that measures the extent to which one side is stronger. If the effect of having one side much stronger disappears when the parity variable is included, a researcher might conclude that it is parity, rather than relative strength, that affects the course of disputes. However, such a conclusion would be unwarranted since this exercise is nonsensical: it amounts to considering the effect of making one side stronger, holding fixed the extent to which the sides are evenly matched.

For these reasons and others, methodologists are beginning to question the use of the regression equation as a dumping ground for unwanted variables. Achen (2003) suggests "A Rule of Three" (ART): no more than three independent variables per equation (in the absence of formal theory that points to more). Thus, I initially include only those variables indicated by the theory in my statistical analyses. While I do consider a variation of the statistical model that includes power parity, I do so as a robustness check on the results about reputations, and do not include this variable and the balance of capabilities in the same equation.

After considering the simple version of the statistical model, the chapter ex-

plores two alternative specifications for the reader who is not convinced by ART. In the first alternative, I include several "controls": whether or not each state is a major power, according to Singer and Small's definition (Singer and Small 1982, 44-45), whether or not the states are contiguous (either actually contiguous or across up to 12 miles of water), and whether or not both states in the pair are democracies. I include the power status and contiguity variables because previous studies suggest that major powers and/or neighbors are better able to go to war with each other or are more likely to interact and thus to become involved in disputes (Bremer 1992; Siverson and Starr 1990). I include the joint democracy variable because a large literature suggests that democracies are less likely to go to war with each other than non-democracies are to go to war with each other and than democracies are to go to war with non-democracies (e.g. Doyle (1983), Russett (1993), Rousseau, Gelpi, Reiter, and Huth (1996)), though both the theoretical foundations and the empirical results on this subject are contested (e.g. Farber and Gowa (1995), Gowa (1999), Layne (1994)).

Democracy is a difficult concept to define, and the literature on the democratic peace defines it in many different ways. I investigate the robustness of my results about reputations using a simple measures of joint democracy based on data from *Polity III: Regime Change and Political Authority* (Jaggers and Gurr 1995; Jaggers and Gurr 1996). Polity III contains a democracy index, created from four components: openness of executive recruitment, competitiveness of participation, competitiveness of executive recruitment and legislative constraints on the executive. It also contains an autocracy index which contains these elements and the regulation of participation. As is fairly common in the literature (e.g. Rousseau, Gelpi, Reiter, and Huth (1996)) I create a single democracy scale for each state.

I subtract the Polity III autocracy index from the democracy index, obtaining a variable that ranges between -10 and 10. I define a pair of states as jointly democratic if both have a score of seven or more.

4.2.4 Estimation Technique

I analyze two sets of statistical models, each examining the effects of military power and of reputations for honesty and for bluffing on states' decisions about crisis escalation. In the underlying theoretical model, the independent variables affect how much the states value each option (their utilities from making various choices). In practice, we do not observe the states' utilities, but only their decisions. The challenger either attacks or does not; if attacked, the defender either follows through on its deterrent threat or backs down. I model each of these decisions with a probit-type of model designed for such binary outcomes. I then translate the results into probabilities to ascertain the magnitudes of the effects.

The first statistical model, described in the following equation, represents the probability that the challenger decides to attack as an increasing function of a linear equation that includes the military balance and the defender's reputation:

$$\begin{aligned}\text{Prob}(\text{challenger attacks} \mid \text{chall. threatens, def. tries deterrence}) = \\ F(\text{constant} + \beta_1 \text{ ln balance of forces} \\ + \beta_2 \text{ def.'s reputation for bluffing}),\end{aligned}$$

where the βs, here and later, are parameters to be estimated that relate the explanatory factors to the challenger's decision to attack. I also investigate two variations on this model: 1) an analogous model that also includes the independent

variables designated as controls: each state's status as a major or minor power, whether or not the states are contiguous, and whether or not the dyad is one with two democracies, and 2) an analogous model that includes the same control variables, but with power parity instead of the natural log of the military balance.

The second set of statistical models represents the decision that the defender decides to defend as an increasing function of a linear equation that includes the same explanatory factors. The simplest version of this model is:

$$\begin{aligned}\Pr\text{ob}(\text{defender defends}|\ \text{chall. threatens, def. tries deterrence, chall. attacks}) = \\ F(\text{constant} + \beta_1\ \ln\text{ balance of forces} \\ + \beta_2\ \text{def.'s reputation for bluffing}).\end{aligned}$$

I also investigate variations on this model that include the control variables.

In order to obtain accurate estimates of the effects of the observed explanatory factors – the balance of forces and the states' reputations for honesty or for bluffing – I must account for selection in the statistical model (Achen 1986). Selection bias is a common problem in situations like this one. When challengers decide whether or not to attack following defenders' attempts at deterrence, these are all challengers that have initiated crises by making a demand coupled with the threat of force. Such challengers have "selected in" to a dispute, and they are likely to have values of unobserved variables that differ from the values of challengers that did not did not "select in" (threaten to use force).[7] Similarly, defenders that decide whether or not to defend following deterrence failure are all defenders that

[7] More technically, there are two underlying equations for each state, a selection equation and an outcome equation, and the error terms in these two equations are correlated.

have tried deterrence. Such defenders differ in unobserved ways from defenders that did not try deterrence. For this reason, ordinary probit estimates of the effects of independent variables on either state's decision to escalate an existing dispute are inaccurate (inconsistent).

As I discussed earlier, many deterrence theorists argue that states' resolve, or willingness to fight, affects the course of international disputes. My theory, too, posits the influence of a type of resolve – the states' values for the disputed issues. Resolve is notoriously difficult to measure. Like other scholars, I do not include it in the equation; if I did, I could capture the concept only poorly.[8] Thus, as in other statistical analyses of disputes, resolve forms part of the error term. This makes it necessary to account for selection in the statistical model. For example, a challenger that is extremely disadvantaged militarily may threaten the use of force (select in) because it considers the issues at stake extremely important. Because it considers the dispute important, this challenger will then be likely to escalate. If selection is not taken into account by the statistical technique, the estimates will suggest that militarily disadvantaged states are more likely to escalate than they really are. Of course, even if one could measure resolve, the problem of selection bias would remain: some of the same small factors would influence both selection

[8]I argued in Chapter 3 that a main purpose of the defender's diplomacy is to communicate information about how important it considers the disputed issue. While the defender communicates some of this information in the course of the conflict, the adversary rarely learns the defender's value for the issues precisely; it merely learns whether or not the defender valued the issues enough to fight (and sometimes it does not even learn this fact). While historians and other analysts sometimes obtain more information *ex post*, even they rarely learn a state's value for the issue precisely. For this reason, it is impossible to measure the defender's value for the issues precisely.

into disputes and the equations of interest, making ordinary probit estimates inaccurate (inconsistent).

To obtain accurate estimates of the effects of interest, I model the challenger's two decisions together and the defender's two decisions together. That is, I use one selection-probit model that represents both the challenger's decision about whether to threaten and its decision about whether to attack. I use another model that represents both the defender's decision about whether to threaten and its decision about whether to defend if attacked. Of course, disputes involve a second type of nonrandom selection: challengers select defenders, and vice versa. Accounting for even one stage of selection is complicated. Since I believe that a state's own decisions are more highly correlated than are its decisions and its opponent's, I account for self selection in the statistical model.[9]

The commonly used Heckman selection model (Heckman 1974; Heckman 1976) requires that the explanatory variables differ in the two equations being considered.[10] Like many game-theoretic explanations, my theory of diplomacy posits

[9]The two equations in the text show a state's decision as also conditional on prior choices made by the other state. This is true empirically and in the construction of the dependent variables; for example, the defender only decides whether to defend if the challenger attacks. Each statistical model represents only the decisions made by one state, however, since the statistical model is a model of a state's utility (Sartori 2003). The first implication that I test concerns the challenger's response to a deterrent threat; thus, if there is no threat, I drop the case for purposes of analyzing the challenger's decision. Similarly, the second implication concerns the defender's response to an attack after a deterrent threat; thus, if there is no attack, I drop the case for purposes of analyzing the defender's decision.

[10]Technically, one can estimate the Heckman model with identical explanatory variables in both equations, but then the model is identified only by the nonlinearity inherent in the discrete-choice first stage. Thus, in practice, an exclusion restriction is required. The Heckman model assumes a continuous dependent variable, which also makes it a poor choice for these data, but

that the same factors – in this case, reputations and the balance of forces – influence all decisions in the model.

I use a new estimator that is intended for a situation in which identical explanatory variables influence the two decisions, represented by dichotomous dependent variables (Sartori 2003).[11] The new estimator is similar to the version of the Heckman model that represents dichotomous dependent variables (Dubin and Rivers 1990). However, it adds an assumption: that the unobserved factors that drive each of a state's two decisions are the same.[12] As I argue in Sartori (2003), this assumption tends to be reasonable (and more reasonable than identifying from functional form alone) when one believes that the two process have the same causes, as they do in my theoretical model. Sartori (2003) shows through simulations that the estimator is better than the Heckman-type estimator when identical explanatory factors influence the two decisions, even when the assumption that the unobserved factors are identical is fairly inaccurate.

4.3 The Escalation of International Disputes: Tests of the Theory

In this section, I examine the escalation and de-escalation of international disputes to determine whether the implications of my theory are borne out in the data. I begin with cross-tabulations to provide some intuition about the data. I then

Dubin and Rivers (1990) and Van de Ven and Van Praag (1981) create extensions for binary dependent variables.

[11] The paper that presents the selection estimator is available at www.princeton.edu/~asartori.

[12] Technically, the estimator adds an identifying restriction: that the error term for an observation is the same in the selection equation and the outcome equation.

present results from the statistical analyses.

4.3.1 Raw Data

As a first look at the data, Tables 4.2 and 4.3 present percentages based on cross-tabulations. For the tables, I separate defenders into two categories: those with more-honest values of the reputation variable (values less than or equal to .32, that is, reputations that in less than the 90th percentile of the reputation variable), and those with less-honest values of the reputation variable (those in at least the 90th percentile of the reputation variable). The first table examines the percent of the time that a challenger attacks following a defender's attempt at deterrence, comparing the frequency for defenders with more-honest reputations to that with less-honest reputations. The second table examines the percent of the time that a defender fights if the challenger attacks, again comparing defenders with more-honest reputations to those with less-honest ones.

The tables show the pattern implied by the theory for the defender's behavior: the defender is more likely to defend when it has more of a reputation for honesty. They do not show the implied pattern for the challenger's behavior: contrary to what the model implies, the challenger is more likely to attack a defender with more of a reputation for honesty.

The cross-tabulations, however, do not take into account issues of nonrandom selection. A rough consideration of selection issues suggests that the possession by the defender of a reputation for bluffing is likely to be more positively related to the probability of deterrence failure than simple cross tabulations or ordinary probit analysis of the challenger's decision to attack would indicate. Probit analysis shows that the challenger is more likely to threaten the use of force, initiating

		Challenger attacks (deterrence fails)?	
		No	Yes
Def rep.	More honest (rep. for honesty)	15.0%	85.0%
	Less honest (rep. for bluffing)	21.2%	78.8%

Table 4.2: Percent of time challenger attacks, by defender's reputation

		Defender fights if deterrence fails?	
		No	Yes
Def rep.	More honest (rep. for honesty)	15.1%	84.9%
	Less honest (rep. for bluffing)	32.0%	68.0%

Table 4.3: Percent of time defender fights if attacked, by defender's reputation

a militarized dispute, when either state has a reputation for bluffing. A challenger may nevertheless threaten the use of force when the defender has a reputation for honesty, but it is only likely to do so when it has a high value of the error term – in particular, when it has a high value for the disputed issue. Challengers are more likely to attack when they consider the issue important, all else being equal. Thus, challengers are more likely to attack when defenders have reputations for honesty. This is not *because* the defenders have reputations for honesty, but because the challengers of these defenders also consider the issue important. Put differently, challengers are *less* likely to attack defenders with reputations for bluffing because challengers are willing to threaten the use of force against defenders with reputations for bluffing even if the challengers do not consider the issues particularly vital.

The selection-probit estimator ascertains the effect of the independent variables, correcting for nonrandom selection. The results of this estimation therefore are likely to show that the association between the defender's reputation for bluffing and the challenger's probability of attacking following a deterrence attempt is more positive than the cross-tabulation would indicate. Similar logic suggests that the association between the defender's reputation for bluffing and the defender's probability of defending following a deterrence failure is more negative than the cross-tabulations would indicate.[13]

[13]Ordinary probit indicates that the defender is less likely to try deterrence when either state has a reputation for bluffing, though the effect is small.

4.3.2 The Challenger's Decision; Does Deterrence Succeed?

Here and in the next section of the paper I discuss the main results of interest to the theory: the defender's reputation affects the probability of deterrence success and the probability that it backs down if deterrence fails. Following the tests of the theory, I discuss robustness of the results. In the next chapter, I turn to empirical results that are not directly related to testing the theory: the effect of the defender's reputation on each state's decision to become involved in a dispute, and the relationship between the military balance and dispute escalation.

Tables 4.4 and 4.6 display the major results, based on the estimation that takes nonrandom selection into account. In each of these tables, the second column of results shows the estimates when I include control variables measuring whether or not the states are major powers, whether or not they are contiguous, and whether or not both states in the dyad are democracies. The third column shows the estimates when I include a measure of power parity instead of a measure of the balance-of-forces. As with probit or Heckman-probit estimates, the results must be translated into probabilities in order to be interpreted substantively. However, the signs of the coefficients represent the signs of the estimated effects and the standard errors show the precision of the estimates.

Table 4.4 reports the estimated impacts of reputations and the military balance on the challenger's decisions to initiate a dispute and to attack following a deterrence attempt on the part of the defender. The table shows that the possession by the defender of more of a reputation for bluffing has the effect that the theory implies: it increases the challenger's propensity to attack, or the chances of deterrence failure. The standard error is small relative to the coefficient, indi-

cating that one can be fairly confident that the true relationship is positive. The second column shows that the estimate is similar when I include control variables in the equation, though slightly smaller in magnitude. The third column shows that the estimate is similar if I include a measure of power parity instead of the balance-of-forces variable.

Because the table shows probit-type coefficients, the coefficients do not immediately reveal the substantive importance of the variables. Table 4.5 below translates the coefficients into predicted probabilities to investigate the magnitude of the impact of the defender's reputation for bluffing on the challenger's decision to attack. The predicted probabilities in Table 4.5 use the first set of estimates in Table 4.4.

In Table 4.5, I perform a series of thought experiments: "How does variation in the defender's reputation affect the probability of an attack by a challenger that is involved in a dispute, *holding fixed the probability that a challenger becomes involved in a dispute?*" This thought experiment speaks to the effect of the defender's reputation on its decision to attack, assuming that a defender with the specified features was randomly selected (correcting for selection bias). Each row of the table shows the effect for a different value of the balance of forces (challenger's/defender's). A military balance of 1.46-to-one in favor of the challenger is the median balance for states that become involved in disputes.

	Implication of the Theory	Equation Implied by Theory	With Control Vars I	With Control Vars II
Issue a challenge?				
Constant (standard error)		-3.07 (.00911)	-3.34 (.0138)	-3.42 (.0173)
Ln capabilities ratio [-12,12] (challenger's/defender's)		.0175 (.00299)	.00865 (.00435)	–
Power parity			–	.279 (.0344)
Defender's rep for bluffing [0,3.8]		.319 (.0156)	.227 (.0192)	.236 (.0191)
Challenger is major power		–	.504 (.0275)	.558 (.0237)
Defender is major power		–	.385 (.0306)	.371 (.0273)
States are contiguous		–	1.09 (.0207)	1.06 (.0209)
Both are democracies		–	-.517 (.0576)	-.515 (.0577)
Attack, if threatened and defender counter-threatened?				
Constant (standard error)		-3.12 (.00980)	-3.36 (.0146)	-3.43 (.0180)
Ln capabilities ratio [-12,12]		.0136 (.00348)	.0129 (.00487)	–
Power parity (0,1)			–	.240 (.0373)
Defender's rep for bluffing	+	.288 (.0177)	.202 (.0211)	.205 (.0208)
Challenger is major power		–	.436 (.0294)	.501 (.0225)
Defender is major power		–	.397 (.0319)	.371 (.0282)
States are contiguous		–	1.03 (.0220)	1.02 (.0224)
Both are democracies		–	-.553 (.0655)	-.549 (.0656)
Sample Size		1050479/ 1337	1008216/ 1261	1014852/ 1281

Table 4.4: Influences on the Challenger's Decisions, to Threaten and to Attack.

	Prob(attack| threatened) defender rep. honesty second stage	Prob(attack| threatened) defender rep. bluffing second stage	Change in prob of deterrence failure
capabilities 1-to-1	60.3%	82.2%	21.9%
capabilities 1.46-to-1	60.1%	81.7%	21.6%
capabilities 7-to-1	59.1%	80.3%	21.2%

Table 4.5: Relationship between defender's reputation and deterrence success or failure.

To perform these thought experiments, I hold the defender's reputation constant in the equation that represents the challenger's decision to threaten the use of force and vary its reputation in the equation that represents whether or not the challenger attacks (whether or not deterrence fails).[14] In the table, a reputation for bluffing means that a defender's reputation has a value of .32, so its reputation is less-honest than those of 90% of defenders. While this may seem extreme, remember that a defender has a reputation for bluffing with value .85 if it was caught in a bluff in the previous year. (Note that one must read across the rows of the table. One cannot read down the columns of this table to see the effect of varying the balance of forces on deterrence success/failure.)[15]

[14]For this thought experiment, I assume that the defender has a reputation for bluffing when the challenger decides whether or not to threaten the use of force.

[15]To see this effect, one must hold the balance of forces fixed in the selection stage and vary it only in the selection equation. I investigate the effect of the balance of forces in the next

The advantages of a reputation for honesty are substantial. A defender with a reputation for bluffing is more likely to experience deterrence failure by about 21 or 22 percentage points. The effect is similar in magnitude regardless of the challenger's military advantage. As the theory implies, a reputation for honesty helps the defender to deter an attack.

4.3.3 The Defender's Decision: Does its Threat Turn out to be a Bluff?

Table 4.6 explains why challengers are more likely to attack when defenders have reputations for bluffing: defenders in that situation are less likely to fight if deterrence fails to deter an attack. Again, the standard error is small relative to the coefficient on the defender's reputation, indicating that one can be fairly confident that the true relationship is positive. The second column of results shows that this result is robust to controlling for contiguity, major power status, and bilateral democracy. The third column shows that it is robust to including a measure of power parity, instead of a measure of the balance of forces.

In Table 4.7, I again perform a series of thought experiments. This time, I ask, "How does variation in the defender's reputation affect the probability that the defender fights if attacked, *holding fixed the probability that a defender tries deterrence in the first place?*" These thought experiments speak to the effect of the defender's reputation on its decision to follow through on its threats, should deterrence fail, assuming that a defender with the specified features was randomly selected (correcting for selection bias). Each row of the table shows the effect for a different value of the balance of forces.

chapter.

	Implication of the Theory	Equation Implied by Theory	With Control Variables I	With Control Variables II
Try deterrence if challenger threatened?				
Constant (standard error)		.661 (.0405)	.792 (.0678)	.700 (.0724)
Ln capabilities ratio [-12,12] (challenger's/defender's)		-.123 (.0142)	-.0937 (.0203)	–
Power parity (0,1)		–	–	.0875 (.131)
Defender's rep for bluffing [0,3.8]		-.0815 (.0673)	-.0415 (.0758)	-.00849 (.0741)
Challenger is major power		–	-.363 (.0898)	-.602 (.0748)
Defender is major power		–	.0272 (.119)	.343 (.0986)
States are contiguous		–	.0179 (.0738)	.0452 (.0742)
Both are democracies		–	-.467 (.205)	-.506 (.206)
Fight if challenger threatened, deterrence tried and failed?				
Constant (standard error)		.303 (. 0372)	.363 (.0634)	.277 (.0693)
Ln capabilities ratio [-12,12]		-.0725 (.0137)	-.0599 (.0194)	–
Power parity (0,1)		–	–	.203 (.123)
Defender's rep for bluffing	-	-.316 (.0624)	-.233 (.0687)	-.205 (.0676)
Challenger is major power		–	-.231 (.0870)	-.362 (.0720)
Defender is major power		–	-.120 (.109)	.0516 (.0894)
States are contiguous		–	.0829 (.0602)	.0742 (.0716)
Both are democracies		–	-.348 (.206)	-.314 (.208)
Sample Size		1562/1115	1476/1065	1476/1065

Table 4.6: Influences on the Defender's Decisions, to Try Deterrence and to Defend if Attacked

The defender's reputation again has a substantively large impact: a defender with a reputation for bluffing is more likely to back down from its deterrent threats by between 5.3 and 6.1 percentage points. The results again are substantively similar, regardless of the balance of forces. Thus, as the theory implies, defenders with reputations for honesty use diplomacy more carefully: they are less likely to back down once they have threatened the use of force.

	Prob(defend| tried deterrence defender rep honesty second stage	Prob(defend| tried deterrence defender rep. bluffing second stage	Change in prob of defense
capabilities 1-to-1	84.0%	78.7%	-5.3%
capabilities 1.46-to-1	87.2%	81.8%	-5.4%
capabilities 7-to-1	86.3%	80.2%	-6.1%

Table 4.7: Relationship between defender's reputation and probability that the defender follows through on its threats.

4.4 Robustness of the Empirical Results

The tables that I present earlier show that my conclusions are robust to one kind of change: the inclusion of several control variables suggested by the literature on international crises. The second column of each table (4.4 and 4.6) shows that the estimates are of the same sign, though slightly smaller in magnitude, when I control for the balance of forces, the status of each state as major or minor power, whether or not the states are contiguous, and whether or not they are

both democracies. The third columns of these tables show that the results about reputations also are robust to including a measure of power parity rather than a measure of the balance of forces.[16]

In this section, I respond to three other possible criticisms of the results. First, one might question my method of measuring reputations for honesty and for bluffing and wonder if my results were an artifact of that method. I show that the conclusions remain the same if I use either of two alternative definitions. Second, one might posit that some states are more disputatious than others, and that my results merely capture this fact. I show that this alternative hypothesis is inconsistent with the data. Finally, one might wonder if my measure of reputations for honesty were capturing the effect of reputations for resolve. I show that my reputations variable affects the course of international interactions, even parsing out the overlap between the two concepts of reputations.

I first consider two alternative definitions of the defender's reputation for bluffing or for honesty. Earlier, I defined reputations in terms of whether or not the defender was caught bluffing in each of the previous ten years, with called bluffs discounted more the farther they are in the defender's past. Here, I consider robustness to two alternative definitions, one more simple and one more complicated. The first, and simpler, is: the defender has a reputation for bluffing if it was caught bluffing in the previous year (it tried deterrence, the challenger nevertheless attacked, and the defender backed down); otherwise, it has a reputation for honesty.

The second definition includes behavior as a challenger as part of the basis of

[16] See my discussion of variables above for an explanation of why I do not include power parity and the balance of forces in the same specification.

the defender's reputation. The formal analyses of Chapter 3 argues that *defenders* obtain reputations for bluffing or for honesty, and defenders with reputations for honesty are more able to use threats effectively. The definition of the defender's reputation that I used earlier follows closely from these formal analyses. In practice, however, challengers clearly also do communicate. It is possible that states also acquire reputations for honesty or for bluffing from interactions in which they are challengers. In fact, the Korean War case study suggested that this is the case. That is, like the Chinese communists prior to the Korean War, a state may threaten to use force to change the *status quo* in its favor, and then fail to follow through when a defender states that it is willing to fight. When it does not follow through, it is shown to be bluffing: it has said that it is willing to use force, but now indicates that it is unwilling to do so. Though my formal work does not speak to this subject, it is also possible that challengers have reputations, and that these reputations affect the course of international disputes.

In the second alternative definition, I still consider a state to increase its reputation for bluffing if it is caught bluffing when it is a defender. I now also consider a state to increase its reputation for bluffing in a given year if it is a challenger and it is revealed to be bluffing (it threatens the use of force; the defender indicates its willingness to fight, and the challenger backs down). I again code the reputations from a state's behavior over the previous ten years, with events farther in the past carrying less weight in the same way. However, I now code each state's reputation as based on all such behavior – regardless of whether the state is a challenger or a defender now and regardless of whether it was a challenger or a defender in a previous dispute. This leads to a reputations variable for the challenger and one

for the defender, each ranging between 0 and 4.55. [17] In the statistical analyses that use this definition, I also include a variable that indicates the interaction between the two states' reputations. If the challenger's reputation does affect its ability to use diplomacy, one might expect each state's decisions to be different when both have severe reputations for bluffing than when only one has such a reputation.

My overall conclusions are the same with either alternative specification. In each case, the estimated effects of the defender's reputation are of the expected sign and quite precise. Tables 4.8 and 4.9 investigate the magnitudes of the effects using predicted probabilities.

In table 4.8, I again perform thought experiments to answer the question, "How does variation in the defender's reputation affect the probability of an attack by a challenger that is involved in a dispute, *holding fixed the probability that a challenger becomes involved in a dispute?*" Remember that this thought experiment speaks to the effect of the defender's reputation on the challenger's decision to attack, assuming that a defender with the specified features was randomly selected (correcting for selection bias).[18] In Table 4.9, I perform analogous thought experiments to investigate the effect of the defender's reputation on its own probability of following through on its threats, should deterrence fail. In each thought experiment, I assume that the balance of forces is one-to-one. In the thought experiments that involve the more-complicated alternative definition,

[17] Since there are more ways to increase a reputation for blufifng by this method of coding, the maximum reputation value is higher than in the previous chapter.

[18] For the thought experiment, I again must hold the defender's reputation constant in the selection equation. I assume that when deciding whether or not to try deterrence, the defender has a reputation for bluffing.

I must also hold the challenger's reputation constant, since I include it and an interaction term in the equation. I assume that the challenger has a reputation for honesty, but the change in predicted probability is similar if the challenger has a reputation for bluffing. When the reputation is based upon the simpler definition, the variable has a value of one if the defender has a reputation for bluffing. When I use the more-complicated reputation, a "reputation for bluffing" in the table signifies that the defender's reputation is worse than those of 90% of other defenders.

	Prob(attack threatened) defender rep. honesty second stage	Prob(attack| threatened) defender rep. bluffing second stage	Change in prob of deterrence failure
result reported earlier	60.3%	82.2%	21.9%
def.'s rep. based on behavior in past year	22.6%	76.5%	53.9%
def.'s rep. based on behavior as chal. or def.	54.7%	86.6%	31.9%

Table 4.8: Relationship between defender's reputation and deterrence success or failure, military balance one-to-one.

	Prob(defend| deterrence failed) defender rep. honesty second stage	Prob(defend| deterrence failed) defender rep. bluffing second stage	Change in prob defender follows through
result reported used earlier	84.0%	78.7%	-5.3%
def.'s rep. based on behavior in past year	88.0%	68.9%	-19.1%
def.'s rep. based on behavior as chal. or def.	87.6%	77.6%	-10.0%

Table 4.9: Relationship between defender's reputation and defender's decision to follow through on its threats and defend, military balance one-to-one

The first row of each table shows the predicted probabilities using my original definition of the defender's reputation. The next two rows show that, using either alternative definition, the defender remains better able to deter an attack when it has a reputation for honesty. Similarly, using either alternative definition, the defender is more likely to follow through on its threats, should deterrence fail, when it has a reputation for honesty. Both effects appear greater in magnitude when I use either alternative definition of the defender's reputation.

Having shown that my conclusions are the same with any of three definitions of the defender's reputation. I now turn to two other possible criticisms. First, some readers may posit an alternative explanation for my results: that some states are simply more disputatious, and that disputes in which these are involved are more likely to escalate towards war. This alternative hypothesis is consistent with the

challenger's behavior. That is, a defender only obtains a reputation for bluffing if a previous dispute escalated to the point at which it could back down from threats. Perhaps there is something "disputatious" about such defenders, and this disputatiousness, not the defenders' reputations, leads challengers to attack them more often.

However, this alternative hypothesis is inconsistent with the defender's observed behavior. If it is defenders' disputatiousness, not their inability to use diplomacy, that leads challengers to attack more often, then these defenders should be more likely to follow through on their threats should deterrence fail. The data show just the opposite: as my theory implies, these defenders are more likely to back down if deterrence fails.

Second, other readers may be suspicious that my measures of reputations for honesty are really measures of reputations for resolve. Deterrence theorists argue that some states are simply more willing to fight than others, and that their willingness to fight is an enduring quality. One of the situations in which I argue states acquire reputations for honesty is also a situation in which deterrence theorists argue they acquire reputations for resolve: A state acquires or increases a reputation for honesty if it follows through on its deterrent threats following deterrence failure. The two arguments are thus difficult to distinguish empirically.

However, I also argue that a state can acquire or increase its reputations for honesty in two ways that do not coincide with a reputation for resolve: by having no dispute in a given year, in which case its reputation for bluffing fades, and by acquiescing to a challenger's demands. As a check that the effect that I capture is not merely that of reputations for resolve, I create two dichotomous "reputation for honesty" variables. The first has a value of one if the present defender either

acquiesced to a challenger's demands or was not involved in a dispute in the previous year; it has a value of zero otherwise. The second has a value of one if the present defender tried deterrence and fought after deterrence failed in the previous year; it has a value of zero otherwise.

When I estimate (as described earlier, including the natural log of the balance-of-forces in the equation), both of these variables have the effects posited by my theory. The challenger is less likely to attack a defender with either kind of reputation for honesty than a defender with a reputation for bluffing. A defender with either kind of reputation for honesty is more likely to follow through on its threats. *Both findings show that reputations for honesty – as distinct from reputations for resolve – affect the course of international disputes.*

Three caveats are in order: First, the variable that indicates that a defender followed through on its threats also indicates that it either acquiesced or had no dispute; in the sample, in any year in which a defender tried deterrence and followed through, it also acquiesced or had no dispute. (The reverse is not true.) Second, if my theory is correct, the two reputation for honesty variables do not belong in the same equation because they are proxies for the same underlying concept; including them both may produce incorrect estimates.[19] Finally, when I do include both in the same equation, the estimate of the effect of a reputation for honesty that comes from acquiescence or not having a dispute on the defender's decision to follow through on its threats is small and imprecise. The estimate suggests that the effect is positive, but does not show with much certainty that there is no effect. The estimate of the effect on the challenger's decision, however, is

[19] Achen (1985) shows that including two proxies for the same concept in the same regression equation can result in an estimate with an incorrect sign if the variables are measured with error.

large and precise. In sum, reputations for honesty and reputations for resolve are overlapping concepts and are therefore difficult to distinguish empirically. Nevertheless, the data suggest that my measure of reputations for honesty is capturing something different from deterrence theory's concept of reputations for resolve.

The implications of the model that I discuss at the beginning of the chapter are borne out by the data, when I analyze the data in a number of different ways. The defender is more likely to succeed in deterring an attack, and more likely to follow through if deterrence fails, when it has a reputation for honesty. This result is quite robust to alternative specifications and is unlikely to be produced by two leading alternative explanations.

4.5 Conclusion

The empirical analyses in this chapter reveal two facts: when a state has a reputation for honesty, it is substantially more likely to attain deterrence success; when it has a reputation for bluffing, it is substantially more likely to back down if deterrence fails. The second fact explains the first. Defenders' deterrent threats are more likely to succeed (challengers are less likely to attack after hearing them) when those defenders have reputations for honesty precisely because defenders with reputations for honesty are more likely to mean what they say. Thus, as I suggested earlier, a reputation for honesty helps the defender to communicate that it is willing to fight, but this ability comes at a cost: the defender must actually be willing to fight more often if deterrence fails in order to obtain this greater credibility.

The empirical analyses in this chapter yield "on average" results: on average,

states behave in ways that the theory implies. They corroborate what individual cases suggest: like China in the Korean War, states that recently have been caught bluffing have a hard time deterring attacks. The reason for this is that a state that has bluffed once is more likely to bluff again in the near future.

These empirical results are consistent with the implications of the theory I present in Chapter 3, and so provide evidence that the theory helps to explain the progression of international disputes. The next chapter turns to several results of the empirical analyses that are not directly related to testing the theory. These concern two subjects: the effect of the defender's reputation on the states' decision to threaten the use of force, and the impact of the balance-of-forces on dispute escalation.

Bibliography

Abreu, D., D. Pearce, and E. Stacchetti (1986). Optimal cartel equilibria with imperfect monitoring. Journal of Economic Theory 39(1), 251–269.

Abreu, D., D. Pearce, and E. Stacchetti (1990). Toward a theory of discounted repeated games with imperfect monitoring. Econometrica 58(5), 1041–1063.

Achen, C. H. (1985). Proxy variables and incorrect signs on regression coe¢cients. Political Methodology 11, 299–316.

Achen, C. H. (1986). The Statistical Analysis of Quasi-Experiments. Berkeley: University of California Press.

Achen, C. H. (2003). An agenda for the new political methodology: Microfoundations and art. forthcoming in the Annual Review of Political Science.

Acheson, D. (1969). Present at the Creation. New York: W.W. North and Company Inc.

Albertini, L. (1952). Origins of the War of 1914, Volume I. London: Oxford University Press.

Alt, J. E., R. L. Calvert, and B. D. Humes (1988, June). Reputation and hegemonic stability: A game-theoretic analysis. American Political Science Review 82(2), 445–466.

Axelrod, R. (1984). The Evolution of Cooperation. New York: Basic Books.

Axelrod, R. and W. Zimmerman (1981, April). The Soviet press on Soviet foreign policy: A usually reliable source. British Journal of Political Science 11, 183–200.

Bennett, D. S. and A. Stam (2000). Eugene: A conceptual manual. International Interactions 26, 179–204.

Bremer, S. (1992). Dangerous dyads: Conditions a¤ecting the likelihood of interstate war, 1816-1965. Journal of Con‡ict Resolution 36, 309–41.

Bueno de Mesquita, B. (1981). The War Trap. New Haven: Yale University Press.

Bueno de Mesquita, B. and D. Lalman (1986). Reason and war. American Political Science Review 80, 1113–31.

Bueno de Mesquita, B. and D. Lalman (1992). War and Reason. New Haven: Yale University Press.

Bueno de Mesquita, B., J. D. Morrow, and E. R. Zorick (1997, March). Capabilities, perception, and escalation. American Political Science Review 91, 15–27.

Chen, J. (1994). China's Road to the Korean War. New York: Columbia University Press.

Christensen, T. J. (1992, Summer). Threats, assurances, and the last chance for peace: The lessons of mao's korean war telegrams. International Security 17, 122–154.

Crawford, V. P. and J. Sobel (1982). Strategic information transmission.

Econometrica 50(6), 1431–1451.

Donovan, R. J. (1982). Tumultuous years : the Presidency of Harry S. Truman, 1949- 1953. New York: Norton.

Doyle, M. W. (1983, summer). Kant, liberal legacies, and foreign a¤airs. Philosophy and Public A¤airs 12(3), 205–235.

Dubin, J. A. and D. Rivers (1990). Selection bias in linear regression, logit and probit models. In J. Fox and J. S. Long (Eds.), Modern Methods of Data Analysis, pp. 410–443. Newbury Park: Sage Publications.

Esser, J. (1998). Alive and well after 25 years: A review of groupthink research. Organizational and Behavioral and Human Decision Processes 73(2/3), 116–141.

Farber, H. and J. Gowa (1995). Polities and peace. International Security 20(2), 123–46.

Farrell, J. and R. Gibbons (1989). Cheap talk can matter in bargaining. Journal of Economic Theory 48(1), 221–237.

Fearon, J. D. (1992). Threats to Use Force: Costly Signals and Bargaining in International Crises. Ph. D. thesis, University of California, Berkeley.

Fearon, J. D. (1994, September). Domestic political audiences and the escalation of international disputes. American Political Science Review 88(3), 577–591.

Fearon, J. D. (1995, Summer). Rationalist explanations for war. International Organization 49(3), 379–414.

Fearon, J. D. (1997, February). Signaling foreign policy interests: Tying hands versus sinking costs. Journal of Con‡ict Resolution 41, 68–90.

Fiske, S. T. and S. E. Taylor (1984). Social Cognition, Chapter 4, pp. 72–91. New York: Random House.

Foot, R. (1985). The Wrong War: American Policy and the Dimensions of the Korean Con‡ict, 1950-1953. Ithaca: Cornell University Press.

Foot, R. (1991). Making known the unknown war: Policy analysis of the Korean con‡ict in the last decade. Diplomatic History 15, 411–431.

Foot, R. (1996). Leadership, perception, and interest: Chinese-American relations in the early Cold War. Diplomatic History 20(3), 473–482.

Freedman, L. and E. Karsh (1994). The Gulf Con‡ict; 1990-1991. Princeton: Princeton University Press.

Gaddis, J. L. (1995). We Now Know; Rethinking Cold War History. Oxford: Clarendon Press.

Garson, R. (1994). The United States and China Since 1949. London: Pinter Publishers Ltd.

George, A. and R. Smoke (1989, January). Deterrence and foreign policy. World Politics 41(2), 170–182.

George, A. L. and R. Smoke (1974). Deterrence in American Foreign Policy. New York: Columbia University Press.

Gibbons, R. (1992). Game Theory for Applied Economists. Princeton: Princeton University Press.

Gochman, C. S. and Z. Maoz (1984). Militarized international disputes, 1816-1976: Procedures, patterns,and insights. Journal of Con‡ict Resolution 28(4), 585–615.

Gowa, J. (1999). Ballots and Bullets. Princeton: Princeton University Press.

Guisinger, A. and A. Smith (2002). Honest threats: The interaction of reputation and political institutions in international crises. Forthcoming, Journal of Con‡ict Resolution.

Halperin, M. H. (1963). Limited War in the Nuclear Age. New York: Wiley.

Harry S. Truman Presidential Oral History Collection (1990). University Publications of America. Microform.

Heckman, J. J. (1974). Shadow prices, market wages, and labor supply. Econometrica 42, 679–694.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection, and limited dependent variables and a simple estimator for such models. The Annals of Economic and Social Measurement 5(4), 475–492.

Helmreich, E. C. (1938). The Diplomacy of the Balkan Wars, 1912-1913. Cambridge: Harvard University Press.

Herzig, E. (1995). Iran and the Former Soviet South. London: Royal Institute of International A¤airs, Russian and CIS Programme.

Hopf, T. (1994). Peripheral Visions: Deterrence Theory and American Foreign Policy in the Third World, 1965-1990. Ann Arbor: University of Michigan Press.

Huth, P. K. (1988). Extended Deterrence and the Prevention of War. New Haven: Yale University Press.

Huth, P. K. (1997, Autumn). Reputations and deterrence: A theoretical and

empirical assessment. Security Studies 7(1), 72–99.

Huth, P. K. (1998). Standing Your Ground; Territorial Disputes and International Con‡ict. Ann Arbor: University of Michigan Press.

Huth, P. K., C. Gelpi, and D. S. Bennett (1993, September). The escalation of great power militarized disputes: Testing rational deterrence theory and structural realism. American Political Science Review 87(3), 609–623.

Huth, P. K. and B. Russett (1984, July). What makes deterrence work? cases from 1900 to 1980. World Politics 36(4), 496–526.

Huth, P. K. and B. Russett (1993, March). General deterrence between enduring rivals: Testing three competing models. American Political Science Review 87(1), 61–73.

Jaggers, K. and T. R. Gurr (1995, November). Tracking democracy's third wave with the Polity III data. Journal of Peace Research 32(4), 469–82.

Jaggers, K. and T. R. Gurr (1996). Polity iii: Regime change and political authority, 1800-1994. Ann Arbor, MI: Inter-university Consortium for Political and Social Research (distributor). 2nd ICPSR ed. Boulder, CO: Keith Jaggers/College Park, MD: Ted Robert Gurr (producers), 1995.

Janis, I. L. (1983). Groupthink : psychological studies of policy decisions and …ascoes. Boston: Houghton Mi - in.

Jervis, R. (1970). The Logic of Images in International Relations. Princeton: Princeton University Press.

Jervis, R. (1976). Perception and Misperception in International Politics. Princeton: Princeton University Press.

Jervis, R. (1997). System E¤ects: Complexity in Political and Social Life. Princeton: Princeton University Press.

Jones, D. M., S. A. Bremer, and J. D. Singer (1996, Fall). Militarized international disputes, 1816-1992: Rationale, coding rules, and empirical patterns. Con‡ict Management and Peace Science 15, 163–213.

Kahn, H. (1965). On Escalation. New York: Praeger.

Keesing's Contemporary Archives (1946). Keesing's Publishing Limited. March 2-9 and April 27-May 7.

Keesing's Contemporary Archives (1992). Keesing's Publishing Limited.

Keesing's Contemporary Archives (1993). Keesing's Publishing Limited.

Kreps, D. M. and R. Wilson (1982). Reputation and imperfect information. Journal of Economic Theory 27(2), 253–279.

Kydd, A. (1992). Whose side are you on? mediation as cheap talk.

Layne, C. (1994). Kant or cant: The myth of the democratic peace. International Security 19, 5–49.

LeBow, R. N. (1981). Between Peace and War. Baltimore: The Johns Hopkins University Press.

Lebow, R. N. and J. G. Stein (1990, April). Deterrence: The elusive dependent variable. World Politics 42(3), 336–369.

Leng, R. J. (1993). Interstate Crisis Behavior, 1816-1980: Realism Versus Reciprocity. Cambridge: Cambridge University Press.

Leng, R. J. and C. S. Gochman (1984, Summer). Dangerous disputes: A study of con‡ict behavior and war. American Journal of Political Science 26(4),

664–687.

Levi, W. (1953). Modern China's Foreign Policy. Minneapolis: University of Minnesota Press.

Lichterman, M. (1963). To the Yalu and back. In H. Stein (Ed.), American Civil-Military Decisions. Birmingham: University of Alabama Press.

Lindsay, M. (1955). China and the Cold War. London: Cambridge University Press.

Maoz, Z. (1983, June). Resolve, capabilities, and the outcomes of interstate disputes. Journal of Con‡ict Resolution 27(2), 195–229.

Martin, L. (1993, April). Credibility, costs, and institutions: Cooperation on economic sanctions. World Politics 45(3), 406–32.

Max, D. (2001). The making of the speech. The New York Times Magazine, 32–37. October 7.

Maxwell, N. (1970). India's China War. London: The Trinity Press.

Maxwell, S. (1968). Rationality in Deterrence. London: Institute for Strategic Studies.

Mercer, J. (1996). Reputation and International Politics. Ithaca: Cornell University Press.

Millett, A. R. (1997, July). A reader's guide to the Korean War. The Journal of Military History 61, 583–97.

Morgan, P. (1977). Deterrence; A Conceptual Analysis. Beverly Hills: Sage Publications.

Morgenthau, H. J. (1967). Politics Among Nations. New York: Alfred A. Knopf.

Morrow, J. D. (1989, November). Capabilities, uncertainty, and resolve: A limited information model of crisis bargaining. American Political Science Review 33(4), 941–972.

Morrow, J. D. (1994). Game Theory for Political Scientists. Princeton: Princeton University Press.

Munro-Leighton, J. (1992, Spring). A postrevisionist scrutiny of America's role in the Cold War in Asia, 1945-1950. Journal of American-East Asian Relations 1, 73–98.

Nalebu¤, B. (1991). Rational deterrence in an imperfect world. World Politics 43(3), 313–335.

National Security Council (1968). Summary notes of the 590th meeting of the national security council. In J. E. Miller (Ed.), Foreign Relations of the United States, 1964-1968, Volume XVII, pp. 272–278. United States Government Printing O¢ce.

Neustadt, R. (1990). Presidential Power and the Modern Presidents. New York: The Free Press.

Nicolson, H. (1963). Diplomacy. New York: Oxford University Press.

of State, D. (1976). Foreign Relations of the United States; 1950; Volume VII; Korea. United States Government Printing O¢ce.

Orme, J. D. (1987). Deterrence failures: A second look. International Security 11(4), 109–112.

Panikkar, K. (1955). In Two Chinas; Memoirs of a Diplomat. London: George Allen Unwin Ltd.

Powell, R. (1990). Nuclear Deterrence Theory. Cambridge: Cambridge University Press.

Rees, D. (1964). Korea: The Limited War. London: MacMillan Co.

Rich, N. (1973). Hitler's War Aims. New York: W.W. Norton and Company.

Ross, L. and C. A. Anderson (1982). Shortcomings in the attribution process: On the origins and maintenance of erroneous social assessments. In D. Kahneman, P. Slovic, and A. Tversky (Eds.), Judgement Under Uncertainty: Heuristics and Biases, pp. 129–152. Cambridge: Cambridge University Press.

Rousseau, D. L., C. Gelpi, D. Reiter, and P. K. Huth (1996, September). Assessing the dyadic nature of the democratic peace, 1918-88. American Political Science Review, 512–533.

Rovere, R. H. (1950). Letter from Washington. The New Yorker, 52–54. September 2 issue, August 25 submission date.

Rovere, R. H. and A. M. Schlesinger, Jr. (1951). The General and the President. New York: Farrar, Straus and Young.

Russett, B. (1963). The calculus of deterrence. Journal of Con‡ict Resolution 7(2), 97–109.

Russett, B. (1993). Grasping the Democratic Peace. Princeton: Princeton University Press.

Sartori, A. E. (2003). An estimator for some binary-outcome selection models without exclusion restrictions. Political Analysis 11(2).

Schelling, T. C. (1960). The Strategy of Con‡ict. Cambridge: Harvard Univer-

sity Press.

Schelling, T. C. (1966). Arms and In‡uence. New Haven: Yale University Press.

Schnabel, J. F. (1992). Policy and Direction; The First Year. Washington: U.S. Government Printing O¢ce.

Schnabel, J. F. and R. J. Watson (1979). The History of the Joint Chiefs of Sta¤; Volume II; The Korean War. Wilmington, Delaware: Michael Glazier, Inc.

Schultz, K. (1998). Domestic opposition and signaling in international crises. American Political Science Review 92(4), 829–844.

Schultz, K. (2001). Democracy and Coercive Diplomacy. Cambridge: Cambridge University Press.

Silber, L. and A. Little (1997). Yugoslavia; Death of a Nation. New York: Penguin Books.

Singer, J., S. Bremer, and J. Stuckey (1972). Capability distribution, uncertainty, and major power war, 1820-1965. In B. Russett (Ed.), Peace, War and Numbers. Beverly Hills: Sage.

Singer, J. D. and M. Small (1982). Resort to Arms; International and Civil Wars; 1816-1980. Beverly Hills: Sage Publications.

Siverson, R. M. and H. Starr (1990). Opportunity, willingness, and the di¤usion of war. American Political Science Review 84(1), 47–67.

Smith, A. (1998a). The e¤ect of foreign policy statements on foreign nations and domestic electorates. In R. M. Siverson (Ed.), Strategic Politicians, Institutions and Foreign Policy. Ann Arbor: University of Michigan Press.

Smith, A. (1998b). International crises and domestic politics. American Political Science Review 92(3), 623–638.

Snyder, G. H. (1961). Deterrence and Defense. Princeton: Princeton University Press.

Snyder, G. H. and P. Diesing (1977). Con‡ict Among Nations. Princeton: Princeton University Press.

Spence, A. M. (1974). Market Signalling. Cambridge, MA: Harvard University Press.

Steiner, H. A. (1958). The International Position of Communist China. New York: Institute of Paci…c Relations.

Stueck, W. (1995). The Korean War; An International History. Princeton: Princeton University Press.

't Hart, P. (1991). Irving L. Janis' victims of groupthink. Political Psychology 12(2), 247–278.

Truman, H. S. (1956). Memoirs; Volume Two; Years of Trial and Hope. Garden City, NY: Doubleday and Company, Inc.

Turner, L. (1970). Origins of the First World War. New York: W.W. Norton and Company.

United States Congress Senate (1951). Military situation in the Far East; hearings before the Committee on Armed Services and the Committee on Foreign Relations. United States Government Printing O¢ce. Eighty-Second Congress, First Session, Part I, May 3, 1951.

Van de Ven, W. P. and B. Van Praag (1981). The demand for deductibles in

private health insurance. Journal of Econometrics 17, 229–252.

Weede, E. (976). Overwhelming preponderance as a pacifying condiiton among contiguous asian dyads. Journal of Con‡ict Resolution 20, 395–412.

Whiting, A. S. (1960). China Crosses the Yalu. New York: The Macmillan Company.

Wilson, R. (1985). Reputations in games and markets. In A. E. Roth (Ed.), Game-Theoretic Models of Bargaining, pp. 27–61. Cambridge: Cambridge University Press.

Wu, S. (1990). To attack or not to attack. Journal of Con‡ict Resolution 34, 531–52.

Zelman, W. A. (1967). Chinese intervention in the Korean War: A bilateral failure of deterrence. Technical report, University of California, Los Angeles. Security Studies Paper 11.

Zhang, S. and J. Chen (1996). Chinese Communist Foreign Policy and the Cold War in Asia; Documentary Evidence, 1944-1950. Chicago: Imprint Publications.

Zhang, S. G. (1992). Deterrence and Strategic Culture: Chinese-American Confrontations, 1949-1958. Ithaca: Cornell University Press.