The Might of the Pen: A Reputational Theory of Communication in International Disputes

Anne E. Sartori

In fall 1950 the United States and the People's Republic of China became embroiled in fighting that neither state may have wanted. President Truman sent UN forces, including U.S. troops, across the thirty-eighth parallel into North Korea. Prior to this action, Chinese leaders tried almost every diplomatic method available to communicate that China would enter the war if U.S. or UN forces crossed the parallel. On 30 September 1950, for example, Chinese foreign minister Chou En-lai publicly warned, "The Chinese people . . . will not supinely tolerate seeing their neighbors savagely invaded by the imperialists."[1] U.S. leaders said they did not want to fight China, but they misread China's myriad threats as bluffs and sent troops across the parallel.[2] China's failure to communicate its resolve contributed to the ensuing tragedy.[3] Why did China's communication fail?

The recent literature on crisis bargaining leads to a pessimistic conclusion about diplomacy: To convey information, threats must be costly , and verbal threats work only when leaders publicize them before domestic audiences that may remove them from office.[4] According to James Fearon, diplomacy is primarily a tool of democracies, not because authoritarian states do not want to use it, but because democracies are more likely to have the "audiences" that make diplomacy credible.[5]

I thank Chris Achen, David Austen-Smith, Bruce Bueno de Mesquita, Bear Braumoeller, Joanne Gowa, Fred Greenstein, Paul Huth, Ken Kollman, David Meyer, Bob Pahre, Ken Schultz, Alastair Smith, Ennio Stacchetti, William Zimmerman, anonymous reviewers, and the editors of *IO* for helpful comments. I benefited from presenting an early version of this article at the Merriam Laboratory Junior Master's Class in Formal Modeling. Beth Bloodgood provided excellent research assistance.

1. Whiting 1960, 107.
2. See Rovere and Schlesinger 1951; and Whiting 1960.
3. A few recent studies (for example, Chen 1994) argue that China would have entered the war whether or not the United States crossed the parallel because it wanted to fight the United States. Even so, the puzzle remains as to why U.S. leaders did not believe China's threats.
4. See Fearon 1992 and 1994; and Schultz 1998.
5. Fearon 1994.

International Organization 56, 1, Winter 2002, pp. 121–149

I demonstrate formally that diplomacy works in the absence of domestic audiences. It works precisely because it is so valuable. When states are irresolute, they are tempted to bluff, but the possibility of acquiring a reputation for bluffing often deters a state from bluffing. A state that has a reputation for bluffing is less able to communicate and less likely to attain its goals. State leaders often speak honestly in order to maintain their ability to use diplomacy in future disputes.[6] They are more likely to concede less important issues and to have the issues they consider most important decided in their favor. The model thus suggests that in the (more complicated) real world, states use diplomacy to attain a mutually beneficial "trade" of issues over time.

States sometimes do bluff, of course. It is impossible to measure how often they do so because opponents and researchers may not discover that a successful deterrent threat was actually a successful bluff. Nevertheless, the model I present here has a theoretical implication about when bluffs will succeed: Diplomacy, whether it be honest or a bluff, is most likely to succeed when a state is most likely to be honest. A state is most likely to be honest when it has an honest reputation to lose, a reputation gained either by its having used diplomacy consistently in recent disputes or having successfully bluffed without others realizing its dishonesty.

Since a state that uses diplomacy honestly cannot be caught in a bluff, concessions to an adversary can be a wise policy. When a state considers an issue relatively unimportant and the truth is it is not prepared to fight, bluffing carries with it the possibility of success as well as the risk of decreased credibility in future disputes. The term *appeasement* has acquired a bad name, but not all states in all situations are deterrable. Many scholars believe that Hitler would have continued his onslaught regardless of Britain's actions in response to Hitler's activities in Czechoslovakia.[7] If Britain had tried to bluff over Czechoslovakia, its attempts to deter Germany's attack on Poland would have been even less credible. Similarly, the United States' acquiescence to the 1968 Soviet invasion of Czechoslovakia was not a high point of moral policymaking; however, given that any threats regarding Czechoslovakia would have been bluffs, honest acquiescence was the best way to preserve credibility. In the latter case, U.S. leaders seemed to realize the benefits of honesty; when Russian ambassador Anatoly Dobrynin told U.S. president Johnson that U.S. interests were not affected by the Soviet action in Czechoslovakia, "in response he was told that U.S. interests are involved in Berlin where we are committed to prevent the city being overrun by the Russians."[8] Johnson's words reveal that he saw a difference between Czechoslovakia, where he was honestly admitting that there was no strong U.S. interest, and Berlin, where he was threatening and prepared to go to war.

When states do use diplomatic threats to deter actions, they often succeed in persuading their challengers to back down. For example, the Anglo–Russian Treaty

6. Jervis 1970, 80.
7. Rich 1973.
8. National Security Council 1968, 274.

of 1942 specified that both Britain and Russia were to withdraw their troops from Iran by March 1946. When the Soviets threatened to remain in parts of Iran, the U.S. chargé d'affaires in Moscow delivered a diplomatic note to the Soviet government calling for their withdrawal of troops from Iran, and the Soviet Union withdrew its troops. When Turkey threatened to invade Cyprus in June 1964, it was deterred by what its messenger called a "brutal note" from U.S. president Johnson. Johnson threatened to suspend military aid and to refuse to come to Turkey's aid in the case of Soviet intervention if Turkey proceeded with its plans.[9]

The model I present here investigates the questions of why and when states can change each other's minds about valuable information in international disputes using verbal threats to use force. The model I propose, unlike most other formal models of crisis behavior, examines states that are engaged in a series of disputes over time, potentially with different adversaries. The repeated game has an important advantage: States' past behavior and concerns about future repercussions can influence both their present behavior and that of other states. Thus, for example, reputations are endogenous to my model, whereas in models of isolated disputes they are merely assumed.[10] In contrast to these others, my model implies that states' behavior is entirely different when the defender has been caught bluffing in a recent dispute. For example, when a defender has recently been caught bluffing, its deterrence is less likely to succeed in the near future.

The reputations for honesty and bluffing that form the core of my explanation differ substantively from the reputations for resolve that form a part of rational-deterrence theory. Rational-deterrence scholars argue that resolve, or willingness to fight, is an enduring quality. They maintain that states can fight to acquire reputations for having resolve, and that possessing such a reputation enhances a state's credibility in future disputes.[11] The reputations in my model are expectations based on past behavior about whether or not a state's diplomacy will be honest in the immediate future.

The article is organized as follows. In the first section, I introduce two ideas I use in my argument: "cheap talk" and common interests. In the second, I present the model. In the third, I show how reputations arise, demonstrate that even "cheap"

9. Under-Secretary of State Ball, in Lebow and Stein 1990, 362. While Lebow and Stein classify the Iran case as "not a deterrence encounter," it was a situation in which the Soviets threatened and the United States issued a diplomatic protest in an effort to deter the threatened action. See Keesing's Contemporary Archives 1946, 7757, 7865; and Herzig 1995.

10. For an example of a work that assumes the existence of a reputation by considering a single dispute in isolation, see Fearon 1992. For political science works that model reputations for resolve using multistage games, see Nalebuff 1991; Powell 1990; Morrow 1989; and Alt, Calvert, and Humes 1988. Huth and Leng also devote attention to the form in which past behavior affects the present. See Huth 1988; and Leng 1993. A large literature in economics also examines the formation of reputations, where reputation is usually tied to, but not perfectly correlated with, some enduring quality of the reputation holder. For an excellent review, see Wilson 1985. Following Axelrod 1984, regime theory has relied extensively on insights from the infinitely repeated Prisoner's Dilemma. However, few works have created infinitely repeated games that are less stylized and more realistic representations of reality.

11. See, for example, Schelling 1966.

diplomacy often changes an adversary's mind about crucial information in a dispute, and discuss my concept of reputations and why war may be more likely in a world with diplomacy than in a hypothetical world without it. In the fourth, I revisit the Korean War case and argue that the model aids our understanding of this case. The model implies that deterrence is more likely to succeed if a state has not recently been caught bluffing; I argue that China's threats to enter the Korean War were dismissed because they came in the context of called bluffs to fight over Taiwan. I conclude by summarizing the implications of this work and providing suggestions for further research.

Cheap Talk and Common Interests

A growing literature on international crises argues that the existence of domestic audiences is the major source of diplomatic success. Drawing on a literature in economics that shows the effectiveness of costly signals, much of the recent international crisis literature argues that verbal diplomacy is costly. A costly signal is one that directly (and negatively) affects the sender's payoff. For example, if one must pay $1,000 to make a speech, that speech is a costly signal.[12] The audience-cost literature argues that domestic audiences "punish" leaders for backing down from a threat, and this punishment acts as a cost that makes diplomacy informative.[13]

This line of reasoning is problematic. If, on balance, backing down is best for a state, then most leaders and domestic audiences should support it. However, if war is best for the state, then most members of both categories should favor it. In neither case do domestic audiences as a group have any reason to punish leaders.[14] Although following through on a threat is sometimes a state's best course of action, I show that backing down can be as well. Thus, contrary to the arguments of the audience-cost literature, rational audiences should not always penalize leaders for backing down.

Signals need not be costly to convey information to a listener. A second literature in economics has examined the effectiveness of "cheap talk" (costless) signals.[15] As Robert Gibbons observes, "The key feature of such a cheap-talk game is that the message has no *direct* effect on either the Sender's or the Receiver's payoff. The

12. See Robert Gibbons' definition of *costless signals* below.

13. See Martin 1993; Fearon 1994; and Schultz 1998.

14. Smith (1998) also makes this point. Smith argues that cheap-talk diplomacy can be effective, but his argument rests on a type of audience cost. He proposes that domestic audiences punish leaders for backing down because backing down is a signal of incompetence; however, if backing down is sometimes the best course of action, then one might think that backing down could sometimes signal competence.

15. See Crawford and Sobel 1982; and Farrell and Gibbons 1989. For examples of cheap-talk models in political science, see also Austen-Smith 1990 and 1992. The model in this article differs from standard cheap-talk games but was motivated by that literature's idea of common interests.

only way the message can matter is through its informative content: by changing the Receiver's belief about the Sender's type, a message can change a Receiver's action, and thus indirectly change both players' payoffs."[16]

Diplomacy is the epitome of "cheap talk." It includes speeches, conversations between state leaders, and diplomatic notes. Actions are part of the lexicon,[17] but often they are also "cheap" relative to the issue at stake in a dispute, such as protecting the independence of an ally, avoiding a war, or maintaining the illusion of military superiority. Any state would be willing to pay the small monetary costs of moving an aircraft carrier if doing so would decide a crisis in its favor. Thus, the monetary costs of diplomacy meet the technical requirement for successful costly signaling, but not the real-world test.

How does one state convince another? A defender's ability to convince a challenger stems from both its and its challenger's behavior and expectations. A challenger will believe a defender's threats only if the challenger expects the defender to be using diplomacy honestly; in other words, some honesty is a prerequisite for effective diplomacy. For diplomatic threats to make a difference, states that threaten must be more likely to fight than those that do not.

The cheap-talk literature shows that communication through words and other costless signals is more likely when the speaker and the listener have interests in common. If two parties have the same goals, the speaker has an incentive to speak honestly and the listener has no reason to doubt the speaker's information. For example, during the Persian Gulf War, the United States agreed to provide Israel with early warning of Iraqi missile launches against it.[18] The United States had an incentive to give truthful information because it wanted the Israelis to be able to defend themselves. Israel, in turn, had every reason to trust the information.

When the listener may use the information to the speaker's disadvantage, the speaker has an incentive to lie. Incentives to lie are stronger when states have divergent preferences. Again in the Gulf War, the Iraqis dramatically announced that coalition forces had hit a "baby milk" factory. The United States countered that only the part of the factory producing biological weapons had been hit. In this case, it is possible that each party was not being fully honest in an effort to manipulate the United States' partners in the coalition.[19]

Interests do not have to be perfectly aligned for communication to occur. When preferences are neither identical nor opposite, as is often the case in international disputes, some communication may still be possible.

The Model

The model rests on five main assumptions:

16. Gibbons 1992, 212 (italics in original).
17. Jervis 1970.
18. Freedman and Karsh 1994.
19. Ibid., 319.

1. *War is costly.* The costs include lives lost and weaponry and other resources used.
2. *Leaders care more about some issues than about others.* States engage in disputes about different issues across time, and these are of varying importance. For example, leaders usually consider home territory more important than other territory; Vietnam was more important to the North Vietnamese than to the United States. As Thomas C. Schelling has argued, for this reason it usually is easier to establish credible deterrence than credible extended deterrence.[20] A state also may consider one ally more important than another. Sometimes, security issues may be most important; at other times, domestic politics may make symbolic or ideological issues primary.
3. *Leaders are unsure of the value other states place on the issues.* In particular, at the start of a dispute states do not know each other's value for the disputed issue well enough to know whether the adversary is prepared to fight. A state cannot precisely infer an adversary's value for the issue from information gathered in previous disputes; even if the issues under contention in two disputes are similar, they are never the same. For example, the disputes in Europe over the Balkans prior to and at the start of World War I differed in a variety of ways.[21] Thus, in each dispute, an adversary must attempt to ascertain anew whether the state values the issue highly enough that it is prepared to fight.
4. *States believe that some positive probability exists that they will still be engaging in international relations tomorrow.* Tomorrow's interaction need not be with the same allies or adversaries as today's.
5. *Each state interacts with many others over time, and tomorrow's interaction is unlikely to be with the same state as today's.*[22] The assumption does not necessarily imply that *every* state interacts with every other, only that each state interacts with many others over time. The states in the model could represent any state, or, for example, the states in a particular region.

This last assumption simplifies the analysis, but it is not necessary for the results. Without the assumption, the challenger and the defender in the model would represent the same two states engaging in disputes about different issues across time. In that case, the argument and the equilibrium proofs still hold. The model would then explain how reputations arise between two states and lead to effective diplomacy.[23]

20. Schelling 1966, 33.
21. See Albertini 1952, 577; and Turner 1970, 95.
22. In technical language, the assumption is that the opportunity for disputes arises "purely randomly" between any pair of states in two interacting groups.
23. Assumption 5 simplifies the model because it rules out the possibility of reputations for resolve. This is not problematic for my analysis because it is not my intention to prove the impossibility of such

The literature on reputations for resolve questions whether reputations generalize.[24] Do states pay attention only to interactions that involve particular adversaries or regions, or do they take note of all interactions? For the reasons discussed earlier, the model here shows that any of these kinds of reputation is logically possible.

The model describes a series of international interactions between states. In each interaction, both the use of diplomacy and the use of force are possible. The stage game, depicted in Figure 1, represents a one-period interaction between a randomly chosen "challenger" and a randomly chosen "defender." Many deterrence models, purely verbal and formal, share a similar form.[25] The repeated game consists of an infinite number of repetitions of the stage game—that is, in each period, each state believes there is a positive probability that it will interact with some other state tomorrow.

At the beginning of each period, a challenger meets a defender. Each is drawn from a pool of potential challengers or defenders. At the start of the stage game, the defender is in possession of the territory or other disputed issue. Each interaction may or may not become a dispute, then a crisis, then a war. At the beginning of the interaction, each state knows the other's relevant history; it is aware of any speeches made and of all actions in its opponent's previous two disputes.

At the beginning of a new dispute, each player "receives" a new value for the issue at stake. This value is the realization of a random variable uniformly distributed between zero and 1.[26] Each state knows how important it considers this issue, but it does not know its adversary's value for the issue; it knows only that its opponent has a value equally likely to be anywhere between zero and 1. The challenger's issue value in this time period is i_t^c, and the defender's is i_t^d.[27] A state's valuation of an issue is its "type."

Each interaction begins with a "talk" stage. In the stage game in Figure 1, the challenger moves first, deciding whether or not to turn the interaction into a crisis. The interaction becomes a dispute if the challenger makes a speech threatening to attack the defender if the defender does not resolve some issues in the challenger's favor. If the challenger does not threaten, the game ends with the status quo maintained, and each state receives a one-period payoff of zero. If the challenger

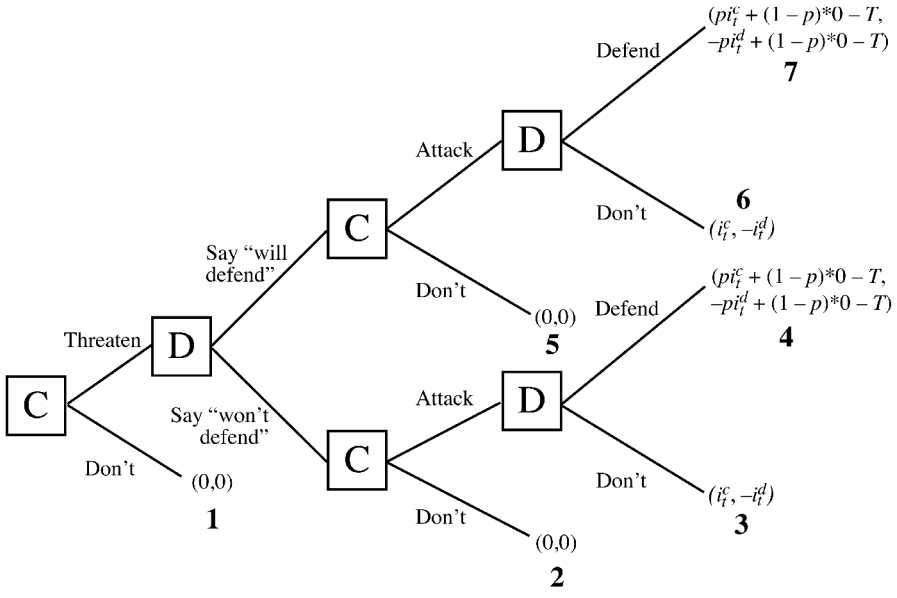
reputations. When states interact repeatedly with the same adversary or over similar issues, it is possible that they gain both reputations for honesty and reputations for resolve. In this case, I would speculate that states learn something about the adversary's value for the issues but never learn that value precisely. Since the adversary's type is never completely known, there is always information to communicate. Thus, reputations for honesty play the role discussed here: They allow a state to communicate about what is unknown. Exploring this scenario fully is beyond the scope of this article.

24. Huth 1997.

25. For example, see Russett 1963; Schelling 1966; Huth 1988; Bueno de Mesquita and Lalman 1992; and Morrow 1989.

26. The upper and lower bounds are chosen for the sake of mathematical convenience, since the payoffs are Von Neumann–Morgenstern utilities and hence are unique only up to a positive affine transformation. Two points on the scale can be chosen arbitrarily.

27. More formally, at the beginning of time t, the challenger learns its value for the issue i_t^c, and the defender learns its value i_t^d, where i_t^c, $i_t^d \sim unif[0, 1]$. Each state's value is private information.



Payoffs (challenger's, defender's):
C: Challenger's move
D: Defender's move

Stage-game parameters:
p: Probability of challenger's winning if war occurs.
T: Cost of fighting, which both states pay in case of war.
i_t^c, i_t^d: Challenger's and defender's values for the territory or other issue in this iteration, respectively. The defender loses its value for the issue if it gives up through acquiescence, war, or backing down. In case of war, its expected value is $p*(-i_t^d) - T$.

FIGURE 1. *The stage game*

does threaten, the defender responds with its own speech; it says either that it "will defend" or that it "won't defend" if an attack occurs. If the defender says "won't defend," it acquiesces to the challenger's demands and gives up the issue.

The challenger then may attack. If it does not do so, the status quo is maintained, and each player has a one-period payoff of zero. If the challenger attacks and the defender resists, war occurs. The challenger has a predetermined probability of winning (p); the probability of winning is a function of the balance of forces

between the two states.[28] If the challenger wins, it acquires the issue (adding i_t^c to its payoff) and the defender loses it (subtracting i_t^d). If the defender wins, which occurs with probability $(1 - p)$, then the status quo is maintained and both players receive a payoff of zero. No matter which player wins, both pay the costs of war (subtracting T from their payoffs). Thus, the challenger's one-period expected payoff from war is $(p * (i_t^c) + (1 - p) * 0 - T)$, and the defender's one-period expected payoff is $(p * (-i_t^d) + (1 - p) * 0 - T)$. If the defender backs down, the issue is resolved in the challenger's favor without war. The challenger gets a one-period expected payoff of i_t^c, and the defender gets a one-period payoff of $-i_t^d$. (For simplicity, the model assumes that an attack with no defense amounts to the defender giving up the issue.) As in Gibbons' definition of a cheap-talk game, a message in this game has no direct effect on payoffs.

When this interaction is over, each state begins another. Note that the state's type does not persist from one dispute to the next. At the beginning of time $(t + 1)$, the state will interact with a new potential adversary and the issue at stake will be different.

Effective Diplomacy

The assumptions of the model suggest that even states involved in disputes have common interests. While each state prefers to have all issues go its way, in practice such one-sidedness is usually impossible. Each state has a particular interest in obtaining a favorable resolution of disputes about issues it considers the most important. Since the outcome of war is uncertain, a state is better off if it achieves deterrence success more frequently when it considers the issues more important. A state that considers an issue crucial sometimes finds itself interacting with another that cares much less about it. But the second state will someday find itself in a similar situation: considering an issue extremely important and facing an opponent that does not. All states can be better off over time if they are more likely to concede when issues are relatively unimportant to them and to resist on issues they consider relatively important. This process is a type of "trade" of issues over time.

As I show formally later, states' communications often are effective because of a communications "norm," or entrenched pattern of behavior. When states act according to the norm, they acquire or assign a reputation for bluffing or for honesty. States disregard the pronouncements made by other states that have bluffed recently and been caught. Since they have no "honest" reputation to lose, states that have been caught bluffing recently try deterrence, whether or not they intend to back down. Since states benefit from diplomacy as long as others are listening, the

28. The assumption that p is fixed is made for analytic tractability. A modified version of the game that relaxes the assumption of fixed p is available from the author. The results discussed in this article carry over to the modified game.

possibility of acquiring a reputation for bluffing provides an incentive for states to use diplomacy honestly.

The Equilibrium

Finding a logical outcome of this situation amounts to finding an equilibrium of the model. Like all infinitely repeated games, this one has many equilibria.[29] I characterize a perfect Bayesian equilibrium (PBE) that corresponds to my substantive argument. PBE is a solution concept for games, like this one, in which some players have information others lack. In a PBE, players update their beliefs rationally, according to Bayes's rule, whenever Bayes's rule applies. I use the technique of factorization, which allows one to characterize equilibria of infinitely repeated games.[30] Equilibria characterized using factorization involve credible threats and promises. The equations that characterize the equilibrium are presented in the appendix.[31]

Disputes are rarely isolated events, and threats often do convey information. To show formally that cheap-talk diplomacy can be effective, I characterize an equilibrium in which the defender's threats often convey information to the challenger about the defender's value for the issue at stake. In the equilibrium I study, there are two sets of strategies: one in which the defender has a reputation for honesty and can partially communicate, and one in which it has a reputation for bluffing and cannot.[32]

I define two concepts: A defender may bluff successfully or it may be caught. In the game, a defender "bluffs" if it says that it "will defend" without intending to follow through if the challenger attacks. A defender is *caught* bluffing in period t if (1) the defender has begun the dispute with a reputation for honesty (the challenger is listening), (2) the defender says that it "will defend," (3) the challenger attacks, and (4) the defender does not follow through. In Figure 1, then, the defender is caught bluffing in period t if the outcome in that period is outcome 6 and if the defender began that period with a reputation for honesty. In equilibrium, defenders are more likely to bluff, and challengers punish the defender by not listening for two periods after a defender has been caught bluffing.

29. Any cheap-talk game has an equilibrium in which talk is meaningless. I discuss one of these in the appendix. The model certainly also has equilibria in which states acquire reputations for certain behaviors (fighting, for example) but talk is meaningless. I do not analyze these explicitly here because my goal is to demonstrate that diplomacy can be effective.

30. Abreu, Pearce, and Stacchetti 1986 and 1990.

31. The equations that characterize the equilibrium in this case are highly nonlinear. I solve numerically; the solution results in a set of equilibria, each of which corresponds to a pair of values of the exogenous variables $\{T, p\}$. The equilibria are all of the same form, described in the text and shown in Figure 2. Thus, to avoid confusion, I refer to them as "the equilibrium" in the discussion.

32. For reasons of tractability, I ignore the possibility that the challenger, too, can communicate. To represent a situation in which the challenger's threats are uninformative, I examine an equilibrium in which the challenger always threatens. The stage game effectively begins at the second node of the game tree shown in Figure 1.

While the duration of the reputation for bluffing—two periods—is chosen for mathematical convenience, a reputation is relatively short-lived for substantive reasons.[33] Not all possible durations of reputations, or "punishments," make sense. On the one hand, for a reputation to exist, it must last for at least one period. On the other hand, it is implausible that the reputation would last forever; a reputation that lasted forever would correspond to a story in which a state caught bluffing could *never* use diplomacy again. In practice, a state does not lose its credibility for all time if it bluffs unsuccessfully once. For example, even after the Soviet Union backed down in the Cuban Missile Crisis, it did not abandon diplomacy thereafter. Thus, a plausible punishment lasts for at least one period but not forever.

In the part of the equilibrium in which the defender has a reputation for honesty, its diplomacy is partially effective. The defender often tells the truth, and so its diplomacy convinces the challenger that it is more likely to fight than the challenger had previously thought to be the case (though it does not convey to the challenger its precise value for the issue). Thus, the defender sometimes can succeed in deterring attacks, either honestly or by bluffing. However, a defender that bluffs may be caught—that is, the challenger may attack. By reducing the effectiveness of a state's diplomacy in the immediate future, a reputation for bluffing harms the state's ability to deter attacks, including its ability to bluff successfully. Thus, when the defender has a reputation for honesty, it bluffs only when the temptation to do so is very strong.[34] The challenger listens to the defender's diplomacy because it wants to know whether or not its adversary will fight. Since the defender often is honest, the challenger can learn something from the defender's threats.

In the part of the equilibrium in which the defender has a reputation for bluffing (the punishment phase), its diplomacy is ineffective. In this situation, the defender cannot lose a reputation for honesty, because it has none. Thus, it has every incentive to bluff. Because of its desire to avoid being duped, the challenger does not pay attention to the defender's diplomacy. Under these circumstances, the defender bluffs more often, but its threats (bluffs or truthful statements) are less credible.

Note that the "punishments" in this equilibrium are substantively sensible, unlike the common punishment schemes in infinitely repeated games. In my equilibrium, states "punish" those that are caught bluffing by refusing to listen; this makes sense because states that have been caught bluffing are less likely to be conveying information. (In technical terms, they are "babbling.") In many equilibria of infinitely repeated games, the states doing the punishing punish only in order to

33. Proofs of the central results about reputations and effective diplomacy do not depend on the length of the reputation; the results hold for any equilibrium of the form depicted in Figure 2. The equations that characterize an equilibrium with a one-period reputation are available from the author; these have numerical solutions for many values of the exogenous variables.

34. As I discuss later, the temptation to bluff is strongest when the defender cares a middling amount about the issues.

avoid being punished themselves. The states that must punish them, in turn, will do so only because they will be punished if they fail to punish, and so on.

The transition from the noncommunicative part of the equilibrium back to the communicative part represents the fading of a reputation. As time passes, the defender regains its reputation and its incentive to use diplomacy honestly. Why should reputations fade? Empirically, as discussed earlier, one observes that they do. Theoretically, the fading of a reputation represents behavior that is optimal in the circular way that all equilibrium behavior is optimal. As long as the challenger believes that the defender will drop its bluffing behavior, it is in the challenger's interest to assign the defender a reputation for honesty. As long as the defender is likely to be honest, the challenger listens because it wants to know whether or not the defender will fight. However, as long as the challenger is willing to listen, it is in the defender's interest to use diplomacy honestly because the defender is more likely to attain its goals with communication. Thus, as long as the challenger is listening, the defender makes listening worth the challenger's time.

Figure 2 shows the equilibrium strategies more formally. If the defender was not caught bluffing in time $t - 1$ or $t - 2$, both states play the strategies in the top half of Figure 2; if the defender was caught bluffing, both states play the strategies shown in the bottom half of Figure 2. Recall that the values of the possible issues are scaled to be between zero and 1; the horizontal lines in Figure 2 represent the possible values of issues that could arise. A state whose issue value in the given dispute is in the specified range (for example, between zero and l) plays the strategy shown above that range. For example, if the defender can communicate, the challenger is deterred (it challenges, but it attacks only if the defender does not try deterrence) if it considers the issue at stake to be worth less than j.[35] Figure 2 presents both the generic equilibrium, in which the thresholds for switching strategies are represented by letters (j, l, m, o, and q), and a numerical example of an equilibrium of the game.[36] The precise thresholds between strategies depend on the balance of forces *(p)* and the costs of war *(T)*. For example, the defender acquiesces when it considers the issue least important, but the cutoff l between issues that the defender concedes and issues over which it bluffs depends on the balance of forces and the likely costs of war.

Efficacy of a Threat

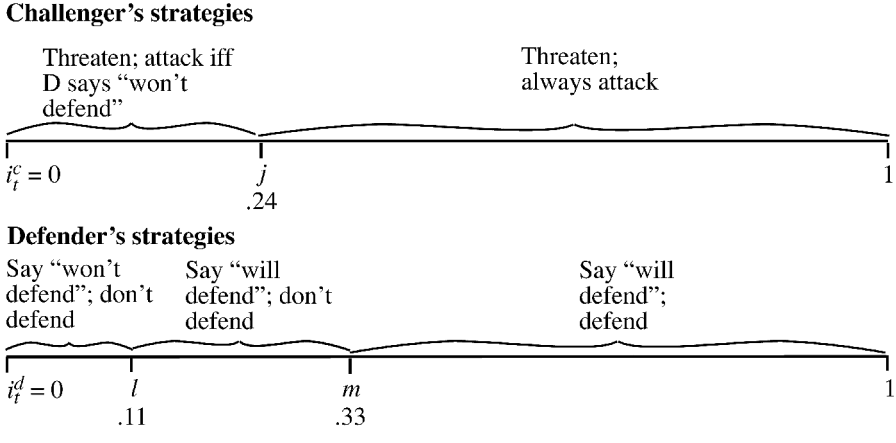
The goal of a deterrent threat is to make a challenger less likely to attack than it was before hearing the threat. Loosely following Schelling,[37] I define the efficacy of a

35. Since the values for the issues are uniformly distributed, there is a j chance that a challenger is involved in a dispute over issues that are so unimportant to it.

36. In the numerical example, the probability that the challenger wins if these two states go to war is 50 percent. The costs of war are 20 percent of the maximum possible ($T = 0.2$). Additional sample equilibria are given in Table 1 in the appendix.

37. Schelling 1960, 6.

a. If the defender was not caught bluffing in either of the previous two periods:



b. If the defender was caught bluffing in either of the previous two periods:

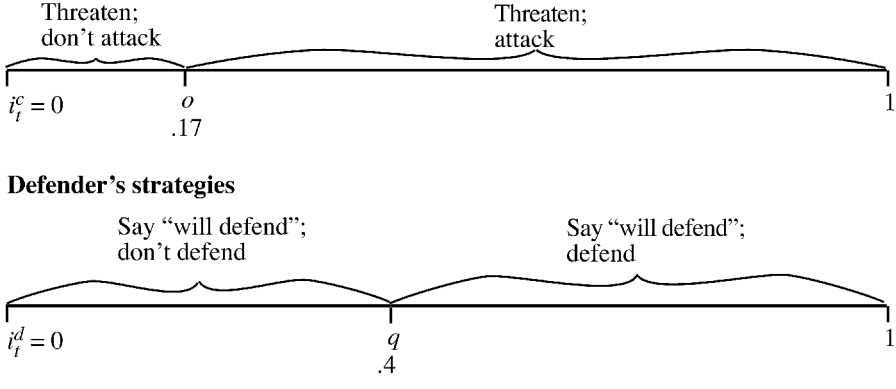


FIGURE 2. *Equilibrium strategies*

threat as the "value-added" of the defender's diplomacy. The efficacy of a deterrent threat is the *change* the defender's threat causes in the challenger's beliefs about the defender's likelihood of fighting. When the defender begins a dispute with a reputation for honesty (top of Figure 2), the challenger begins a dispute believing that the probability that the defender will fight is $1 - m$. Afterward, it believes that this probability is $(1 - m)/(1 - l)$.

For the equilibrium shown in Figure 2, the defender's threats have efficacy when the defender has a reputation for honesty. This fact is easily proven:[38]

1. Existence of an interior equilibrium implies that $0 < l, m < 1$.
2. $(1 - m) < (1 - m)/(1 - l)$.

$\rightarrow$ The defender's efficacy is positive.

The defender's threat serves to convince the challenger that the defender is more resolved than the challenger had previously believed.

Similarly, the defender's threats do not have efficacy when the defender has a reputation for bluffing. The challenger begins a dispute believing that the probability that the defender will fight is

$$\frac{(1 - q)}{(1 - q) + (q - 0)},$$

or $(1 - q)$. All defenders try deterrence, and

$$\frac{(1 - q)}{(1 - q) + (q - 0)} \text{ of these do not defend.}$$

The defender's efficacy is zero. Thus, the game has an equilibrium in which reputations for honesty and for bluffing allow for effective, "cheap" diplomacy.

Effective diplomacy is not an inevitable outcome of international interactions. Though the communications-norm equilibrium is a logical outcome of international interactions, the model has an alternative equilibrium in which states do not acquire reputations and diplomacy is completely ineffective. Thus, the model confirms the logic of an argument made by Ted Hopf and others: Since each dispute involves new issues, there is no logical reason why states' behavior today *must* be tied to how they behaved in the past.[39] Nevertheless, these two equilibria are not equally plausible from an empirical standpoint, since states do obtain reputations for bluffing.

The model predicts that we will observe bluffs and reputations for bluffing, but few of each. Moreover, states will be more likely to bluff for a short time after they are caught bluffing (though not always); this is why they are less likely to obtain diplomatic success.[40] Robert Axelrod and William Zimmerman find little deception in the Soviet press about Soviet foreign policy over a thirty-five–year period.[41] Nevertheless, examples of bluffs do exist. In 1962 India pursued a "forward policy," placing troops forward of Chinese outposts. China repeatedly threatened to respond

38. See the appendix for proof of existence of interior equilibria.
39. Hopf 1994.
40. The model also explains the prevalence of acquiescence. Bueno de Mesquita and Lalman find 109 cases of acquiescence among 707 international disputes. Bueno de Mesquita and Lalman 1992, 68.
41. Axelrod and Zimmerman 1981.

with force and backed down. Accordingly, the Indian government dismissed these threats as bluffs, even though China was the stronger power.[42] Indian general J. N. Chaudhuri said, "It was a game of Russian roulette, but the highest authorities of India seemed to feel that the one shot in the cylinder was a blank. Unfortunately for them and for the country it was not so. The cylinder was fully loaded."[43] The Chinese attacked in what became the Sino–Indian War.

In the Balkan conflict of 1908–1909, Russia backed down from its initial support for Serbia, leading Serbia to recognize Austria-Hungary's annexation of Bosnia-Herzegovina.[44] When the Austro–Serbian conflict flared up again in 1912, Russia bluffed again, accepting Albanian independence when it became clear that Germany would support Austria-Hungary.[45] Russia's threats in 1914 then were ineffective. Given the issues, one might have expected Russia's threats to be *more* credible in 1914. The issue in 1914 was the invasion of a Slavic state by Austria; based on estimates of Russia's interests, other states should have believed Russia likely to fight. Not so: In a 1914 letter to the German ambassador to Britain, the German foreign secretary writes, "The more boldness Austria displays, the more strongly we support her, the more likely is Russia to keep quiet. There is certain to be some blustering in St. Petersburg, but at bottom Russia is not now ready to strike."[46] Of course, we cannot know if Germany would have been deterrable in 1914 if German leaders had believed the Russian threats. What we do know is that German leaders did not believe the threats and that Russia was not bluffing. Once the Russians became convinced that diplomacy was failing, they mobilized for war. A few days later, Germany mobilized in turn and declared war on Russia.[47]

In discussions of reputations for resolve, Jervis questions the idea that backing down is bad for credibility, suggesting instead that states that have backed down will attempt to rebuild their reputations by fighting.[48] Since states that are rebuilding their reputations are more likely to fight, the diplomacy of states that retreated recently should be more rather than less credible. As applied to this work, Jervis's argument (which is directed at reputations for resolve rather than reputations for honesty) suffers from two main weaknesses. First, this behavior probably is not a logical outcome of international interactions. In my model, if states were to possess increased credibility after bluffing, there would be no disincentive to bluff. Without a disincentive to bluff, states would bluff often and diplomacy never would be effective. Second, the empirical record does not support the conjecture that states obtain a surge in credibility after bluffing. To the contrary, like China and Russia in

42. See Bueno de Mesquita and Lalman 1992, 200–202; and Maxwell 1970, 226, 237.
43. Maxwell 1970, 171.
44. Albertini 1952, 190–300. This case is classified as a "called-bluff" adversary crisis in Snyder and Diesing 1977, 137.
45. Helmreich 1938.
46. Turner 1970, 85.
47. Turner 1970.
48. For example, see Jervis 1997, 266–71.

the cases just discussed, states that have been caught bluffing obtain reputations for bluffing and suffer losses of credibility.

Situational Reputations

Scholars usually argue that reputations are dispositional: They stem from underlying differences among people, firms, or states.[49] For example, David M. Kreps and Robert Wilson argue that monopolists may be strong or weak, and that they wish to develop reputations for strength in order to convince upstart firms not to enter the market.[50] Schelling suggests that states may be resolute or irresolute (and that resolve is at least partly an enduring characteristic), and that they wish to acquire reputations for resolve in order to increase their credibility in international conflict.[51]

In contrast, states in my model have no permanent differences in disposition. This difference reflects my belief that the crucial information—how much the state values the disputed issue—varies from one dispute to the next. Whether or not a state has the "will to fight" depends on what the issues are and so must be communicated anew through the use of diplomacy.

Since, by assumption, there are no enduring dispositions, the reputations that emerge in this model are situational. Like a dispositional reputation, a situational reputation is an assessment—based on past, observed behavior—that a state is likely to behave in a certain way in the future. Defenders lie or tell the truth depending on the situation they are in, and they acquire reputations based on this behavior. Since a defender that is caught bluffing is actually more likely to lie in the near future, it is rational to assign that defender a reputation for dishonesty. While these situational reputations could be considered a modeling simplification, they also may be an interesting real-world phenomenon.

In their minds, leaders may attribute reputations for bluffing to disposition, thinking *this state tends to lie*. Such an error would be in keeping with my model; as long as leaders are less likely to believe the bluffer in its next interaction, the communications norm will work. Such an attribution error also would be consistent with portions of psychological theory, which argues that persons overattribute undesirable behavior to disposition.[52] A detailed comparison of the beliefs in my model with those in psychological theory is not explored here, and a precise match is unlikely to occur. However, this consistency hints at similarities between rational choice and psychological approaches to decision making.

49. For a recent argument to this effect, see Mercer 1996.
50. Kreps and Wilson 1982.
51. Schelling 1966.
52. On attribution error as a psychological bias, see Ross and Anderson 1982; and Fiske and Taylor 1984. For applications in political science, see Jervis 1976; and Mercer 1996.

Diplomacy, Commitment, and War

I suggested at the beginning of this article that diplomacy allows states to realize common interests. To pursue this line of reasoning, I now engage in a thought experiment: I compare the real world, in which states do communicate, to a hypothetical world in which communication is impossible. I do so by comparing two equilibria: the partially communicative equilibrium discussed earlier (which represents the real world) and an equilibrium in which states never can communicate (which represents the hypothetical world without communication). In the noncommunicative equilibrium, states play the strategies in the bottom half of Figure 2 forever rather than alternating between the two sets of strategies.

Two results related to states' well-being stem from the fact that credibility allows states to use diplomacy as a form of commitment. Commitment is a double-edged sword. The defender is less likely to suffer an attack when it can commit itself than when it cannot. However, for its diplomacy to be effective, the defender must be willing to follow through.[53] Thus, the defender gets more of what it wants when it can communicate. However, war can be more common with communication than without.[54]

For reasons of tractability, I examine a situation in which the challenger cannot communicate. In an analogous equilibrium with bilateral communication, I conjecture that the logic of the equilibrium would carry over. The challenger, too, would obtain issues that it valued more over time, on average, with the help of communication. Thus, states would attain a "trade" of issues over time as suggested at the beginning of the article, conceding less important issues and keeping or acquiring more important ones.

The Korean War Revisited

This article began with a reference to China's failed attempt to deter the United States from crossing the thirty-eighth parallel during the Korean War. Many U.S. leaders believed that a fight over Korea was not in China's best interest.[55] Yet, as I noted earlier, Chinese leaders stated flatly that China would enter the war if U.S. troops crossed the parallel, and U.S. leaders dismissed the Chinese threats as bluffs.

Scholars and policymakers have proposed many reasons for the communication failure. These include the timing of the threats prior to a crucial UN vote, the fact that some of the threats were relayed through the mistrusted Indian ambassador K. M. Panikkar,[56] and "groupthink" on the part of U.S. leaders.[57]

53. This argument is similar to an argument Fearon makes about costly signaling: "Audience costs and risks of preemptive war work to separate states according to resolve *precisely* by posing the risk of unwanted escalation." Fearon 1992, 172.

54. See the appendix for further discussion of these results.

55. See, for example, Rees 1964, 113.

56. See Acheson 1969; Truman 1956; and Rees 1964.

57. Janis 1983.

These explanations are unsatisfying for a variety of reasons. The choice of Panikkar as a messenger is a poor explanation because many of the threats were not made by Panikkar; they were issued directly by Chinese government officials or published in the official press. For example, on 22 September 1950 China admitted sending aid to the North Koreans and threatened to send more. A Ministry of Foreign Affairs spokesman said that China "will always stand on the side of the Korean people."[58] The timing of the threats is an equally poor explanation for U.S. officials' disbelief; the warnings continued after the vote in question, and U.S. leaders still considered them to be bluffs.

Groupthink is an insufficient explanation because President Truman, Secretary of State Acheson, and others had major disagreements with General MacArthur. Janis argues, "The mutual support for risk-taking, it seems to me, was part of a more general pattern of concurrence-seeking behavior, which also fosters uncritical acceptance of stereotypes of out-groups and a sense of unanimity about the wisdom and morality of past decisions."[59] Yet this case involves much behavior that cannot be explained as concurrence-seeking: Truman and others were notably and consciously unhappy with many of MacArthur's views, statements, and actions over the course of the war, particularly with his desire to involve the Taiwanese in the Korean conflict.[60] Rather than groupthink in this crisis, there was substantial disagreement—but almost no disagreement over whether or not the Chinese were bluffing.[61]

Some scholars argue that MacArthur was eager to proceed into North Korea regardless of whether he believed China would carry out its threats and that Truman ultimately agreed with his view.[62] That debate is interesting, but the relevant question here is whether MacArthur and other top leaders believed China's threats, not what MacArthur or the United States would have done had they believed them. At his Wake Island conference with Truman, MacArthur stated that he did not believe China would enter the war.[63] Part-way through the war, MacArthur was relieved of his command and Congress held an unusual set of hearings into the military situation in the Far East and into the circumstances surrounding his relief.[64] At these hearings, MacArthur mentioned a November CIA report that "said that they felt there was little chance of any major intervention on the part of the Chinese forces."[65] In fact, it is remarkable that in an administration replete with power struggles over the conduct of the war, the president, the secretary of state, members

58. Whiting 1960, 106.
59. Janis 1983, 58.
60. See Truman 1956; Acheson 1969; and Rovere and Schlesinger 1951.
61. While political scientists continue to teach groupthink, and probably should, the evidence in the psychological literature is inconclusive. See 't Hart 1991; and Esser 1998.
62. Neustadt 1990, 114.
63. See Harriman and MacArthur 1951; and Truman 1956.
64. The hearings by the Joint Foreign Relations and Armed Services Committees, commonly known as the "MacArthur Hearings," were held in May–June 1951. For a discussion, see Rees 1964, chap. 15.
65. U.S. Senate 1951, 18.

of the Joint Chiefs of Staff, and MacArthur all were aware of the Chinese threats and to various degrees dismissed them as bluffs.[66] Though they were eager to blame each other for the conduct of the war, none of these leaders argued that the others did believe the Chinese threats. One reason for MacArthur's disbelief was that he considered China's military capabilities inferior. This fact does not, however, explain other leaders' dismissal of the threats as bluffs. Even the Joint Chiefs of Staff believed that the Chinese had sufficient capabilities to enter the war if they wanted to do so.[67] The question thus remains: Why did so many U.S. leaders disbelieve the threats?

A final explanation is that China had acquired a reputation for bluffing through a series of bluffs over Taiwan.[68] The model of diplomacy I develop here helps to shed light on China's failed communication by explaining when and why states acquire reputations for bluffing and the negative consequences of such a reputation for a state's diplomacy. In the period before the war and in its early months, China was almost, though not quite, ready to fight over Taiwan. China's leaders went so far as to make invasion plans but did not follow through.[69] China threatened to enter the Korean War in the context of its having made a series of unfulfilled threats to fight over Taiwan. For example, early in the summer of 1950, Mao stated that the Chinese would invade Taiwan by September, but they did not do so.[70] Many of China's threats over Korea came in September and October 1950, just after this particular threat to invade Taiwan had been revealed as bluff.[71] China's history of bluffing contributed to the widespread belief in the United States that China's threats to intervene in the Korean War were bluffs.

Why was China prepared to fight over Korea but not quite prepared to invade Taiwan in 1950? One reason is that Chinese leaders worried that U.S. forces would march straight through North Korea into China. For example, Mao cabled Chou En-lai that "not to have intervened . . . would have meant that 'the reactionaries at home and abroad would be swollen with arrogance when the enemy troops press on toward the Yalu River [the border between North Korea and China].' "[72]

Would China's threats over Korea have been more successful had it not threatened Taiwan? Any answer to this question is speculative. Nevertheless, prior to these conflicts, U.S. leaders seem to have believed that China was more likely to

66. See Truman 1956; Acheson 1969; and U.S. Congress 1951. George Kennan was one of few to express strong concerns about Chinese intervention; his views were marginalized. See Donovan 1982, 217.

67. U.S. Congress 1951, 759.

68. See Lichterman 1963; Zelman 1967; and Neustadt 1990, fn. 23, 334–35. The Indian ambassador was considered not credible in part because he, too, had warned of a Chinese invasion of Taiwan that had not materialized. See Rovere and Schlesinger 1951, 133.

69. Whiting 1960, 64–65.

70. Rovere 1950, 54.

71. See Whiting 1960; Acheson 1969; and Truman 1956.

72. Gaddis 1995, 80, quoting Mao Zedong.

fight over Taiwan than over Korea.[73] If, in the early 1950s, China had said it was willing to let Taiwan live in peace, this news would have come as a surprise to the United States. If China soon thereafter had said it was unwilling to tolerate U.S. troops in North Korea, U.S. leaders might well have concluded that preventing a U.S. occupation of North Korea was quite important to the Chinese.

Conclusion

In this article I demonstrate the roles of honesty and reputation in crisis bargaining. I argue that diplomacy works because it is beneficial, and states have an incentive to preserve their ability to use it. Using diplomacy, states communicate information about which issues they consider particularly important and thus increase their chances of prevailing when these issues are at stake. The flip side is that states must concede those issues they consider less crucial; by doing so, they maintain credibility for disputes in which they consider the stakes higher.

This analysis has several implications for understanding crisis behavior and communication. First, the work fills a gap in theories of crisis behavior by explaining why cheap talk can change an adversary's mind about a state's resolve. I do not argue that costly signals are meaningless; states' choices to use cheap talk or costly signals remains an important subject for future research. One conjecture is that a state will turn to costly signals when the currency of cheap talk has been devalued by bluffing and the state considers the issue exceptionally important—Russia at the beginning of World War I might be an example. Of course, if it were always possible to use costly signaling to avoid the repercussions of bluffing, then cheap talk would never be credible.

Second, the model has novel, testable implications about states' behavior in international disputes. Here, I develop one central implication: As was the case for China in the Korean War, a state's deterrence success is less likely to be effective and the state is less likely to attain its goals when it has a reputation for bluffing; it is more likely to succeed when it has a reputation for honesty. This hypothesis is testable because the concept of reputations incorporated in this model can be readily operationalized. If a state acquiesced or fought a war recently, it currently has a reputation for honesty; otherwise, it has a reputation for bluffing. Of course, the model is based on a simplifying assumption that a state knows nothing about its opponent's value for the issues at the beginning of a new international dispute. In the real world, I would expect that *a priori* beliefs about the importance of a dispute to an adversary would also affect the probability of deterrence success. The reputations hypothesis in this article differs from that of deterrence theory: The model I propose suggests that acquiescing in one dispute leads to a reputation for

73. For example, Acheson lists the "lack of real advantage to China itself of coming in" as one of the reasons he believed the Chinese would not enter the Korean War. U.S. Congress 1951, 2000.

honesty, so acquiescing is positively associated with success in the next dispute; deterrence theory suggests that acquiescing in one dispute leads to a reputation for lacking resolve, so it is negatively associated with success in the next dispute. The reputations hypothesis also is contrary to audience-cost models, which suggest no association between a state's recent behavior and the outcome of the current dispute.[74]

As mentioned earlier, some deterrence theorists have questioned the idea that reputations for resolve generalize. For example, Paul Huth argues that states might pay attention only to their regional neighbors or to their partners in enduring rivalries, ignoring the outcomes of unrelated disputes.[75] This same question arises with reputations for honesty. The results presented here hold even if we consider the challenger and the defender always to be the same two states interacting over time. Nevertheless, the question of when states assign reputations is an important one, and the answer affects the precise implications of the model. If states pay attention to all disputes (as in an informal version of the Law Merchant[76]), then whether or not a state is caught bluffing in the current dispute affects its next interaction, regardless of the adversary. If states ignore disputes outside of their regions, then a reputation affects a state's next dispute in its region. Since either answer is logically possible, the question can best be addressed empirically. This empirical work is another topic for future research.

Third, encouraged by Robert Putnam's well-known study, a growing number of international relations scholars have turned their attention to domestic influences on foreign policy.[77] While this research program is valuable, my analysis here sounds a warning note about the dangers of neglecting systemic factors. I demonstrate formally that domestic audiences are unnecessary for effective signaling. Moreover, I show that crisis bargaining models that assume domestic audiences punish leaders who back down are assuming irrational behavior on the part of domestic publics. While bluffing is rare, it nevertheless can be the best policy.[78] Backing down is *always* best for a state that is caught bluffing; if a state considered the issue important enough to be worth a fight, it would not be bluffing. Thus, a rational domestic audience should not punish a leader for backing down, as the audience-cost literature assumes it will. The different models have different implications: The audience-cost literature suggests a correlation between regime type and deterrence

74. However, see Guisinger and Smith forthcoming, which combines the reputational theory suggested here with a domestic politics model.

75. Huth 1997.

76. Milgrom, North, and Weingast 1990.

77. Putnam 1988.

78. In equilibrium, defenders with reputations for bluffing often bluff because challengers expect them to be likely to bluff and do not pay much attention to their diplomacy. Even defenders with reputations for honesty bluff when they consider the issues moderately important—too important for acquiescence but not enough to warrant going to war.

success, whereas the model I propose suggests that democracies and autocracies should be equally able to use diplomacy and equally likely to deter attacks.[79]

Fourth, my analysis provides an example of the benefits of the theory of infinitely repeated games as a tool for understanding international relations. Most international relationships are ongoing. Without repeated games, scholars must make many more assumptions about the ways in which the past and the future influence present choices. While such simplifications often yield valuable insights, they also distort the effects of the past and the future.

Finally, although this work directly concerns only diplomacy, it suggests ways to understand communication more generally. Like the "cheap talk" literature in economics, I argue that common interests help actors communicate. In most human interactions, the ability to communicate has significant value. Precisely because dishonesty places the ability to communicate in the future at risk and all parties understand this, persons and states can convey information through the use of language.

Appendix

Technical Characterization of the Equilibrium

The text describes an equilibrium of the game in which the defender's threats convey information to the challenger. The parameters in Figure 2 can be considered thresholds; for instance, a challenger with a value above the threshold o finds it optimal to attack in time t if the defender was caught bluffing in time $t - 1$ or $t - 2$. Choosing thresholds amounts to choosing strategies. Thus, formally characterizing the equilibrium consists of finding equations for the thresholds such that the thresholds constitute optimal solutions to the states' maximization problems.

The equilibrium described here is a perfect Bayesian equilibrium, which is an extension of the Nash equilibrium concept to games of incomplete information. At every information set, no state can wish to deviate from its prescribed equilibrium strategy, given the beliefs and strategies of the other states. In addition, beliefs must, if possible, be updated according to Bayes's rule.

The beliefs of players off the equilibrium path are as follows. Since a player gets a new type at the beginning of each interaction, a player beginning a new interaction believes that the other player is equally likely to be of any type at that point in time, even if the player's previous actions have been off the equilibrium path. Within an interaction, the only off-path beliefs that could affect play are the beliefs of a challenger when the defender is being ignored (punished). The equilibrium I examine specifies that a defender that is being ignored always says that it will defend the issues. If the challenger observes a deviation (the defender saying that it will not defend), the challenger's off-path beliefs are that the straying defender is equally likely to have any value for the issues. In this event, the challenger's strategies after

79. Of course, this hypothesis assumes all else is equal. For reasons outside of both this model and audience-cost models, democracies may be more likely to try to resolve disputes through diplomacy.

a discovered bluff are not tied to the defender's words; after observing this deviation, challengers with issue values below o refrain from attack and those with values above o attack. Thus, the defender weakly prefers its equilibrium strategy.

Equilibrium Analysis

I begin by using factorization to write the defender's expected payoffs from time t onward for arbitrary values of the thresholds j, l, m, o, and q.[80] Factorization is the Bellman principle applied to a game-theoretic model. After any t-period history, any equilibrium must specify a successor equilibrium. An equilibrium can be factorized into the strategies it specifies for the current period and the strategies it specifies from tomorrow onward. Thus, verifying that a player does not wish to deviate from the equilibrium strategies requires checking that it is not worthwhile to deviate once (today), given that the players play their equilibrium strategies from tomorrow onward. In the equilibrium described here, one must keep in mind that the successor strategies depend on today's play and that there are thus two different continuation values.

In particular, the defender's expected payoff from t onward depends on whether the defender was or was not caught bluffing in time $t-1$ or $t-2$. Call the discount factor δ, the total expected payoff from t onward for the defender who was not caught bluffing in time $t-1$ or $t-2$, "$w1$," and the total expected payoff from t onward for the defender who was caught bluffing in time $t-1$, "$w2$." There is a separate expected payoff from t onward for a defender who was caught bluffing in time $t-2$.

The threshold for a defender who was caught bluffing in time $t-2$ is just q (the same as for a defender who was caught bluffing in time $t-1$), because the defender's choice of strategy in either of these situations does not influence its continuation value. Similarly, the challenger's threshold when the defender was caught bluffing in time $t-2$ is o.

Referring to Figure 2 for the meaning of the thresholds and the associated equilibrium strategies, the expected payoffs can be written as follows:

$$w1 = \left[l\left(\frac{-l}{2} + \delta w1\right) = (m-l)\left(j\delta w1 + (1-j)\left(\frac{-m-l}{2} + \delta w2\right)\right) \right. \tag{1}$$
$$\left. + (1-m)\left(\delta w1 + (1-j)\left(p\,\frac{-m-1}{2} - T\right)\right)\right]$$

$$w2 = \left[(1+\delta)\left(q(1-o)\left(\frac{-q}{2}\right) + (1-q)(1-o)\left(p\,\frac{-q-1}{2} - T\right)\right) + \delta^2 w1\right]. \tag{2}$$

To see the meaning of these value functions, consider the first term on the right side of equation (1):

$$l\left(\frac{-l}{2} + \delta w1\right).$$

80. Abreu, Pearce, and Stacchetti 1986 and 1990.

The first part of this term, l, represents the probability that a defender that was not caught bluffing in time $t-1$ or $t-2$ finds itself in time t with an issue value between zero and l, since the distribution of values is uniform. The expression

$$\left(\frac{-l}{2}+\delta w1\right)$$

represents the expected payoff of a defender that finds itself with an issue value between zero and l in the communicative situation; the payoff consists of such a defender's expected value today plus its expected continuation value. All defenders with issue values less than l say "won't defend," are attacked by all challengers, and give up the issues. Therefore, a defender in this situation receives a one-period payoff of $-i_t^d$; expected over its possible values for the issues, the one-period payoff of a defender between zero and l is thus $\frac{-l}{2}$. Since defenders pursuing this strategy are never caught bluffing, their expected payoff from time $t+1$ onward is $w1$. In time t, this expected payoff is worth only $\delta w1$. Of course, the term $l\left(\frac{-l}{2}+\delta wl\right)$ represents only part of the defender's utility, the part that comes from those times when its issues are between zero and l. The defender also expects that the interaction may be over issues worth between l and m or between m and 1, and assigns a probability to each of these possibilities. The remaining terms on the right-hand side of equation (1) represent the expected utility that comes from those possibilities.

In equilibrium, each state chooses its thresholds at time t to maximize its expected utility from time t onward. If the defender enters time t and was not caught bluffing in time $t-1$ or $t-2$, the state chooses its strategies to maximize $w1$; the maximization problem is then as follows:

$$\begin{aligned}\mathrm{Max}_{l,m}\Bigg[\, l\left(\frac{-l}{2}+\delta wl\right)+(m-l)\left(j\delta w1+(1-j)\left(\frac{-m-l}{2}+\delta w2\right)\right)\\ +(1-m)\left(\delta w1+(1-j)\left(p\,\frac{-m-1}{2}-T\right)\right)\Bigg].\end{aligned} \tag{3}$$

The first-order conditions for this maximization problem result in a set of two equations characterizing the defender's strategies in the communicative phase in equilibrium. These are

$$-l+\delta w1=j\delta w1+(1-j)(-l+\delta w2) \tag{4}$$

$$-m+\delta w2=-pm-T+\delta w1. \tag{5}$$

The challenger's maximization problem is theoretically similar; however, in practice the problem is simpler because the challenger's actions today do not affect its continuation values. Its first-order condition is

$$j=\frac{(1-m)T}{(m-l+p-mp)}. \tag{6}$$

TABLE 1. *Sample equilibrium points*

T	p	j	l	m	o	q
.1	.3	.22	.005	.14	.21	.14
.1	.4	.21	.06	.14	.17	.17
.1	.5	.19	.11	.15	.13	.2
.1	.6	.16	.16	.17	.11	.25
.4	.3	.28	.04	.55	.24	.57
.4	.4	.29	.14	.57	.17	.67
.4	.5	.31	.25	.58	.09	.80

If the defender was caught bluffing in time $t - 1$ or $t - 2$, both states' maximization problems are simple, because their continuation values are the same regardless of their play in the current interaction. The resulting first-order conditions are

$$q = \frac{T}{(1 - p)} \tag{7}$$

$$o = \frac{(-T^2 - pT + T)}{(-p^2 + p - pT + T)} \tag{8}$$

The value-function equations, equations (1) and (2), and the five first-order-condition equations, equations (4) through (8), constitute a set of seven equations and seven unknowns and characterize the strategies played on the equilibrium path. In combination, these equations are highly nonlinear. (For example, the equations can be reduced to characterize one threshold, m, with all other endogenous variables eliminated. This equation is very lengthy and is not presented here.) I solve the system of equations numerically; each numerical solution is a set of values for the parameters $\{j, l, m, o, q, w1, w2\}$, where the values satisfy the seven equations. The numerical work shows that the equilibrium exists for many pairs of the exogenous variables T and p, since the equations can be satisfied by meaningful values of the parameters (values for which thresholds are between zero and 1 and $l < m$). Some pairs $\{T, p\}$ lead only to solutions of the system of equations such that one or more states choose threshold values outside of the interval (0, 1), so that an equilibrium of the posited form does not exist for these pairs. I discuss a sample equilibrium in the text and provide several others in Table 1.

The thresholds in the table are rounded to the nearest percentage (except 0.005, because zero would be a case of nonexistence for the interior equilibrium). Notice, for example, that if a defender threatens, that threat is more likely to be successful when the defender has a reputation for honesty than when it has a reputation for bluffing ($o < j$).

Comparing a World with Communication to One Without

To evaluate the welfare effects of communication, I compare a world with communication to a hypothetical world without—that is, I compare the communicative portion of the equilib-

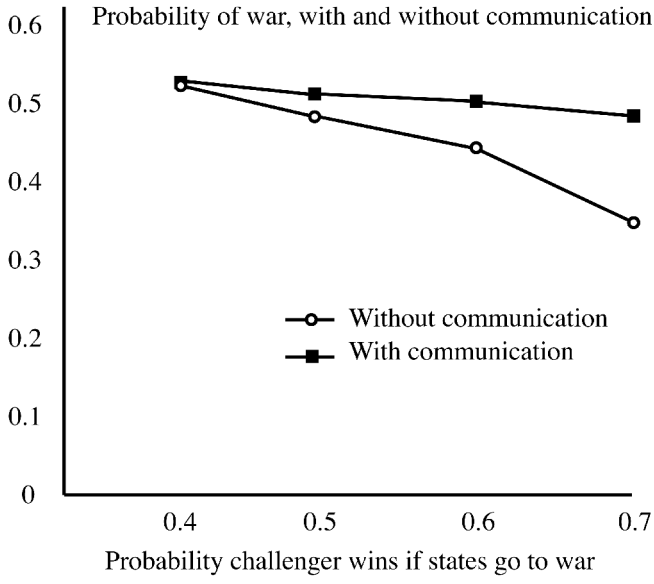


FIGURE 3. *Comparing a world with communication to a hypothetical world without*

rium discussed in the text to an equilibrium in which states never can communicate. In this noncommunicative equilibrium, states play the noncommunicative strategies at the bottom of Figure 2 in every period; strategies are not history-dependent, so neither player's continuation values affect its current play. This method of comparing the partially communicative equilibrium discussed in the text to the noncommunicative equilibrium contains a simplification: In the equilibrium in the text, there are some periods without communication. However, the complete payoffs in the full "communicative" equilibrium are a convex combination of the two sets of payoffs, communicative and noncommunicative. Thus, the patterns I discuss later generalize immediately to the full equilibrium, though the precise numbers do not.

Figure 3 compares the probability of war with and without communication for several values of the balance of forces (p). In these examples, war is relatively cheap ($T = 0.2$). The figure shows a pattern: The probability of war usually is higher with communication. When the defender can communicate, it is more likely to deter an attack, but it is also more prepared to fight. The probability of war is the probability that the challenger attacks a defender that tries deterrence (denoted *"A"*) multiplied by the probability that the defender defends (denoted *"D"*).[81] The challenger attacks under fewer circumstances when it expects the defender's diplomacy to be effective, so A is always lower with communication than it is

81. Technically, the probability of war includes the probability that the challenger attacks a defender that does not try deterrence multiplied by the probability that such a defender defends, but the latter probability is always zero in this equilibrium.

without it. However, the defender sometimes finds itself choosing to fight when, were it not committed, it would have chosen to back down. Thus, D is always higher when the defender has a reputation for honesty. The resulting probability of war ($A * D$) may be either higher or lower but is usually higher.[82]

The defender also does better, on average, when it can communicate than when it cannot. The figures do not show this fact, but Figure 2 shows why: The defender is, on average, happier with those issues that are resolved in its favor when it is able to communicate. When a defender with a reputation for honesty considers the issue to be worth less than l in Figure 2, it acquiesces to the challenger's demand, so these issues are never among those resolved in favor of the defender.

In this equilibrium, the ability to commit works to the defender's advantage. When the defender has not recently been caught bluffing, the defender is able to deter many attacks, and its expected payoff rises sharply. The challenger's payoff, however, falls. The challenger holds an informational advantage, since the defender's value for the issues is partially revealed. However, the informational advantage comes to naught, since the information reveals that the defender is likely to defend. This disparity between the challenger and the defender is interesting, but it probably is not a feature of the real world. In a more general equilibrium of the same model, the norm might apply both to defenders and to challengers. In this situation, the challenger's average payoff might be higher than its average payoff in this equilibrium.

References

Abreu, Dilip, David Pearce, and Ennio Stacchetti. 1986. Optimal Cartel Equilibria with Imperfect Monitoring. *Journal of Economic Theory* 39 (1):251–69.

———. 1990. Toward a Theory of Discounted Repeated Games with Imperfect Monitoring. *Econometrica* 58 (5):1041–63.

Acheson, Dean. 1969. *Present at the Creation.* New York: W.W. North and Company.

Albertini, Luigi. 1952. *Origins of the War of 1914. Volume I.* London: Oxford University Press.

Alt, James E., Randall L. Calvert, and Brian D. Humes. 1988. Reputation and Hegemonic Stability: A Game-Theoretic Analysis. *American Political Science Review* 82 (2):445–66.

Austen-Smith, David. 1990. Information Transmission in Debate. *American Journal of Political Science* 34 (1):124–52.

———. 1992. Strategic Models of Talk in Political Decision Making. *International Political Science Review* 13 (1):45–58.

Axelrod, Robert. 1984. *The Evolution of Cooperation.* New York: Basic Books.

Axelrod, Robert, and William Zimmerman. 1981. The Soviet Press on Soviet Foreign Policy: A Usually Reliable Source. *British Journal of Political Science* 11 (April):183–200.

Bueno de Mesquita, Bruce, and David Lalman. 1992. *War and Reason.* New Haven, Conn.: Yale University Press.

Chen, Jian. 1994. *China's Road to the Korean War.* New York: Columbia University Press.

Crawford, Vincent P., and Joel Sobel. 1982. Strategic Information Transmission. *Econometrica* 50 (6):1431–51.

Donovan, Robert J. 1982. *Tumultuous Years: The Presidency of Harry S. Truman, 1949–1953.* New York: Norton.

82. The probability of war when the defender has a reputation for honesty is $(1 - j) * (1 - m)$ in Figure 2. The probability when the defender has a reputation for bluffing is $(1 - o) * (1 - q)$.

Esser, J. K. 1998. Alive and Well After 25 Years: A Review of Groupthink Research. *Organizational and Behavioral and Human Decision Processes* 73 (2/3):116–41.
Farrell, Joseph, and Robert Gibbons. 1989. Cheap Talk Can Matter in Bargaining. *Journal of Economic Theory* 48 (1):221–37.
Fearon, James D. 1992. Threats to Use Force: Costly Signals and Bargaining in International Crises. Ph.D. diss., University of California, Berkeley.
———. 1994. Domestic Political Audiences and the Escalation of International Disputes. *American Political Science Review* 88 (3):577–91.
Fiske, Susan T., and Shelley E. Taylor. 1984. *Social Cognition.* New York: Random House.
Freedman, Lawrence, and Efraim Karsh. 1994. *The Gulf Conflict, 1990–1991.* Princeton, N.J.: Princeton University Press.
Gaddis, John Lewis. 1995. *We Now Know: Rethinking Cold War History.* Oxford: Clarendon Press.
Gibbons, Robert. 1992. *Game Theory for Applied Economists.* Princeton, N.J.: Princeton University Press.
Guisinger, Alexandra, and Alastair Smith. Forthcoming. Honest Threats: The Interaction of Reputation and Political Institutions in International Crises. *Journal of Conflict Resolution.*
Harriman, W. A., and Douglas MacArthur. 1951. Memorandum of Conversation. President Harry S. Truman's Office Files Part 2, Reel 16/26, Frame 00544. Frederick, Md.: University Publications of America, 1989. Microform.
Helmreich, Ernst Christian. 1938. *The Diplomacy of the Balkan Wars, 1912–1913.* Cambridge, Mass.: Harvard University Press.
Herzig, Edmund. 1995. *Iran and the Former Soviet South.* London: Royal Institute of International Affairs, Russian and CIS Programme.
Hopf, Ted. 1994. *Peripheral Visions: Deterrence Theory and American Foreign Policy in the Third World, 1965–1990.* Ann Arbor: University of Michigan Press.
Huth, Paul K. 1988. *Extended Deterrence and the Prevention of War.* New Haven, Conn.: Yale University Press.
———. 1997. Reputations and Deterrence: A Theoretical and Empirical Assessment. *Security Studies* 7 (1):72–99.
Janis, Irving Lester. 1983. *Groupthink: Psychological Studies of Policy Decisions and Fiascoes.* Boston: Houghton Mifflin.
Jervis, Robert. 1970. *The Logic of Images in International Relations.* Princeton, N.J.: Princeton University Press,
———. 1976. *Perception and Misperception in International Politics.* Princeton, N.J.: Princeton University Press.
———. 1997. *System Effects: Complexity in Political and Social Life.* Princeton, N.J.: Princeton University Press.
Keesing Publishing. 1946. *Keesing's Contemporary Archives* (2–9 March, 27 April–7 May) 6:7757, 7865. Bristol: Keesing's Publishing Limited..
Kreps, David M., and Robert Wilson. 1982. Reputation and Imperfect Information. *Journal of Economic Theory* 27 (2):253–79.
Lebow, Richard Ned, and Janice Gross Stein. 1990. Deterrence: The Elusive Dependent Variable. *World Politics* 42 (3):336–69.
Leng, Russell J. 1993. *Interstate Crisis Behavior, 1816–1980: Realism Versus Reciprocity.* Cambridge: Cambridge University Press.
Lichterman, Martin. 1963. To the Yalu and Back. In *American Civil-Military Decisions*, edited by Harold Stein, 569–639. Birmingham: University of Alabama Press.
Martin, Lisa. 1993. Credibility, Costs, and Institutions: Cooperation on Economic Sanctions. *World Politics* 45 (3):406–32.
Maxwell, Neville. 1970. *India's China War.* London: The Trinity Press.
Mercer, Jonathan. 1996. *Reputation and International Politics.* Ithaca, N.Y.: Cornell University Press.

Milgrom, Paul R., Douglass C. North, and Barry R. Weingast. 1990. The Role of Institutions in the Revival of Trade: The Law Merchant, Private Judges, and the Champagne Fairs. *Economics and Politics* 2 (1):1–23.

Morrow, James D. 1989. Capabilities, Uncertainty, and Resolve: A Limited Information Model of Crisis Bargaining. *American Political Science Review* 33 (4):941–72.

Nalebuff, Barry. 1991. Rational Deterrence in an Imperfect World. *World Politics* 43 (3):313–35.

National Security Council. 1968. Summary Notes of the 590th Meeting of the National Security Council. In *Foreign Relations of the United States, 1964–1968*, vol. 17: *Eastern Europe*, edited by James E. Miller, 272–78. Washington, D.C.: U.S. Government Printing Office, 1996.

Neustadt, Richard. 1990. *Presidential Power and the Modern Presidents*. New York: The Free Press.

Powell, Robert. 1990. *Nuclear Deterrence Theory*. Cambridge: Cambridge University Press.

Putnam, Robert D. 1988. Diplomacy and Domestic Politics: The Logic of Two-Level Games. *International Organization* 42 (3):427–60.

Rees, David. 1964. *Korea: The Limited War*. London: MacMillan.

Rich, Norman. 1973. *Hitler's War Aims*. New York: W. W. Norton.

Ross, Lee, and Craig A. Anderson. 1982. Shortcomings in the Attribution Process: On the Origins and Maintenance of Erroneous Social Assessments. In *Judgment Under Uncertainty: Heuristics and Biases*, edited by Daniel Kahneman, Paul Slovic, and Amos Tversky, 129–52. Cambridge: Cambridge University Press.

Rovere, Richard H. 1950. Letter from Washington. *The New Yorker*, 2 September, 52–54.

Rovere, Richard H., and Arthur M. Schlesinger, Jr. 1951. *The General and the President*. New York: Farrar, Straus and Young.

Russett, Bruce. 1963. The Calculus of Deterrence. *Journal of Conflict Resolution* 7 (2):97–109.

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

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

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

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

Snyder, Glenn H., and Paul Diesing. 1977. *Conflict Among Nations*. Princeton, N.J.: Princeton University Press.

't Hart, Paul. 1991. Irving L. Janis' Victims of Groupthink. *Political Psychology* 12 (2):247–78.

Truman, Harry S. 1956. *Memoirs*, vol. 2, *Years of Trial and Hope*. Garden City, N.Y.: Doubleday.

Turner, L. C. F. 1970. *Origins of the First World War*. New York: W. W. Norton.

U.S. Senate. 1951. *Military Situation in the Far East. Hearings Before the Committee on Armed Services and the Committee on Foreign Relations*. Pts. 1–3. 82nd Cong., 1st sess., 3 May, 18.

Whiting, Allen S. 1960. *China Crosses the Yalu*. New York: Macmillan.

Wilson, Robert. 1985. Reputations in Games and Markets. In *Game-Theoretic Models of Bargaining*, edited by Alvin E. Roth, 27–61. Cambridge: Cambridge University Press.

Zelman, Walter A. 1967. Chinese Intervention in the Korean War: A Bilateral Failure of Deterrence. Security Studies Monograph 11. Los Angeles: University of California, Los Angeles.