In the latter half of the 1980s, Ronald Reagan’s approach to the Soviet Union was often characterized by the maxim “trust but verify,” a phrase which simultaneously acknowledged the pressures of the security dilemma and the realities of coexistence in a time of great change. We are now at another time of great change in international politics, but as North Korea and Iran advance their nuclear programs, the Bush administration’s approach has been one of “never trust, only verify.” A hard-line realist might contend that this approach is absolutely correct, that anarchy forces states to distrust one another’s motives, and that the first half of the phrase “trust but verify” was an effort to placate the electorate rather than a true reflection of Reagan’s policy beliefs.

In two recent and important books, Andrew Kydd of the University of Pennsylvania and Anne E. Sartori of Princeton University offer alternative explanations for a policymaker’s choice to be distrustful of another state’s motives. Both authors approach the issue of crisis bargaining by employing the expected utility model of conflict. The problem of war, as both Kydd and Sartori view it, is one of uncertainty. States are strategic actors with an incentive to pursue their goals at the lowest expected cost, which means they have strong motivation to make threats that are bluffs and to misrepresent information about the costs they are willing to bear to win. Kydd and Sartori reach different conclusions about how states can more effectively communicate in bargaining situations, but in the end, each presents a model of behavior in which reputation plays a key role.

In Trust and Mistrust in International Relations, Kydd crafts a theory of how trust between actors is damaged or reinforced through aggressive or cooperative behavior, and more importantly, how a reputation for trustworthiness...
affects the probability of future cooperative behavior. His definition of trust is grounded in expectations about preferences, or in his words, “that trust is a belief that the other side prefers mutual cooperation to exploiting one’s own cooperation, while mistrust is a belief that the other side prefers exploiting one’s own cooperation to returning it.” Kydd introduces the term “Bayesian Realism” which is a learning-focused synthesis of defensive and offensive realism, to describe his view of crisis bargaining. States, he contends, cannot trust one another’s motives initially, but through interaction and efforts at cooperation, can update their beliefs about one another’s trustworthiness.

Kydd employs detailed game theoretic analysis and an extended case study of U.S.-Soviet negotiations during the Cold War to examine three general situations in which trust plays a role in enhancing or impeding cooperative behavior. First, he examines classic security dilemmas, or spiral models, to use a term coined by Robert Jervis. Kydd delineates states into one of two groups—security-seekers, which are interested solely in their own defense, and aggressors, which are expansionist and offensive. The tragic equilibrium in a spiral model is that even security-seeking states, which Kydd contends are inherently trustworthy, are forced to engage in arms races with other security-seeking states.

Second, Kydd expands his analysis to include multilateral bargaining situations, particularly those that involve a hegemon as a potential third-party. He concludes that only trustworthy hegemons—those that prefer cooperative behavior, such as the United States in the second half of the twentieth century—promote cooperative behavior among other actors. Hegemons that are not trustworthy have a particularly negative effect, since they not only stymie cooperation between themselves and other states, but also prevent it from occurring between dyads of otherwise-trustworthy states.

Finally, Kydd explores reassurance conditions, in which pairs of states send signals in order to convey their trustworthiness. When these signals are costly, they allow the recipient to distinguish more accurately between trustworthy states making truthful statements and untrustworthy states making deceitful statements. At the end of the Cold War—the height of the “trust but verify” era—the Soviet Union began to send costly signals of their trustworthiness, including signing the Intermediate Nuclear Forces treaty in 1987, announcing social policy shifts at the nineteenth Party Conference in 1988, and free elections for the Congress of People’s Deputies in 1989.

Sartori, in Deterrence by Diplomacy, also examines the question of how naturally suspicious states can resolve conflicts, avoid war, and achieve
cooperative outcomes. Sartori asks why, if cheap talk is as meaningless and information-free as most of the deterrence literature contends it is, diplomacy is ever effective. Why, when states that are truly resolute and those that are actually irresolute make the same speeches and employ the same diplomatic language, are threats to retaliate or promises to cooperate sometimes believed? Like Kydd, Sartori argues that the answer involves the efficient transmission of information between parties in a bargaining situation. Specifically, she contends that diplomacy is effective because it is perceived by states to be a critical tool of statecraft and one they fully expect to utilize in future bargaining situations. She employs a reputation-based theory of diplomacy’s effectiveness, and concludes that fear of acquiring a dishonest identity will deter states from bluffing.

Sartori opens the book with a discussion of resolve, which she defines as a state’s willingness to fight over a specific issue. She levels appropriate criticism against traditional deterrence theory for conflating resolve itself with the communication of information about that resolve between adversaries. Rational deterrence theory often dismisses diplomacy as ineffective cheap talk, ignores it entirely, or assumes it away as negligible “offstage” behavior, rather than recognizing that credible transmission of information is critical to effective and efficient dispute resolution. Her basic model, like Kydd’s, is centered on a Bayesian type of information-updating. The challenger in a conflict has prior beliefs about the probability a target will acquiesce or resist. The defender’s response to a challenge contains information about its capabilities and the price it is willing to pay to suffer for victory, which is transmitted to the challenger and may or may not change its assessment of the defender’s resolve. According to Sartori, rational deterrence theory fails to describe the mechanism through which a defender can change a challenger’s beliefs about its resolve.

This mechanism, Sartori contends, is diplomacy, which she defines as, “the use of language and other signals by one state in an attempt to convey information to another.” More specifically, states see diplomacy as a valuable foreign policy tool, and thus, have an interest in maintaining a reputation for honest negotiation. Bluffing, or making promises and not carrying through, weakens a state’s reputation for honesty and makes future diplomatic efforts less effective. More important and more interesting than that result, however, is her distinction between a reputation for honesty and a reputation for resolve. Here, too, she differs sharply from traditional deterrence theory. A state gains a reputation for resolve by refusing to back down in the face of challenges and through being willing to fight over a variety of issues. A reputation for honesty, on the other hand, is gained through not backing
down when challenged on important issues, but also by granting concessions on those issues not critical to a state’s national interest. This is one of Sartori’s key contributions to bargaining theory—by sometimes acquiescing, a state gains credibility in those situations in which it refuses to do so. In other words, a state that makes concessions when it truly is not willing to fight over the issue earns a reputation for diplomatic honesty.

It follows that, in addition to a reputation for honesty, occasional acquiescence may also earn a state a reputation as irresolute. However, as Sartori ably demonstrates in a formal bargaining model and a probit estimation, a reputation for honesty is a more effective mechanism for communicating resolve than a straightforward reputation for high resolve. This interesting counterintuitive conclusion grows out of the strong case Sartori makes for the value of so-called “cheap talk.” A reputation for honesty, her empirical results demonstrate, “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.”

Overall, Kydd and Sartori share an interest in the mechanisms that generate effectiveness in bargaining, which is a critical and growing topic in security studies. Each author offers critical insights into the process through which interstate disputes can be resolved nonviolently. War, they contend, is an inefficient outcome, since there is always a negotiated settlement that both parties would prefer. The problem is reaching that negotiated settlement in an environment of mutual suspicion and strong incentives to misrepresent private information about resolve. Kydd contends that more efficient bargaining outcomes depend on reputations for trustworthiness, while Sartori argues that the mechanism is diplomatic honesty. Reagan and Gorbachev’s policy of “trust but verify” reflects the assumptions shared by Kydd and Sartori—that cooperation begets cooperation through reputation-based mechanisms.

What, then, can Kydd and Sartori tell us about the effectiveness of the current administration’s “never trust, only verify” approach to Iran and North Korea’s nuclear ambitions? The answer, according to the frameworks Kydd and Sartori have skillfully laid out, lies in the actors’ respective reputations for trustworthiness and diplomatic honesty, or more specifically, in the willingness of the Bush administration to compromise and to overlook labels like “evil” and “irrational.” As these situations evolve and progress toward unknown and unpredictable outcomes, we would expect negotiation to be more successful if both parties manage to engage in behavior that demonstrates trustworthiness and honesty. The current administration, through the policies it has pursued to this point, appears to be unwilling to engage in a good-faith bargaining effort. In Kydd’s
model, mutual distrust is the appropriate starting point; however, it appears the current administration either has chosen not to engage in, has found ineffective, or has not been convinced by the type of information-updating Kydd describes. According to Sartori’s model, on the other hand, the most effective and efficient method of dispute resolution necessitates a willingness on the part of the Bush administration to employ costly signals of its intent to avoid violence and to grant concessions on non-central issues—strategies that, so far, it appears unwilling to implement. In either case, this unwillingness bodes poorly for the likelihood of resolving these bargaining crises without violence.