FEIGNING WEAKNESS

Branislav L. Slantchev
Department of Political Science, University of California – San Diego

Abstract

In typical crisis bargaining models, strong actors must convince the opponent that they are not bluffing and the only way to do so is through costly signaling. However, in a war strong actors can benefit from tactical surprise when their opponent mistakenly believes that they are weak. This creates contradictory incentives during the pre-war crisis: actors want to persuade the opponent of their strength to gain a better deal but, should war break out, they would rather have the opponent believe they are weak. I present an ultimatum crisis bargaining model that incorporates this dilemma and show that a strong actor may feign weakness during the bargaining phase. This implies that (a) absence of a costly signal is not an unambiguous revelation of weakness, (b) the problem of uncertainty is worse because the only actor with incentives to overcome it may be unwilling to do so, and (c) because of the difficulty with concealing resolve, democracies might be seriously disadvantaged in a crisis.

ACKNOWLEDGEMENTS

E-mail: slantchev@ucsd.edu. This research was supported by the National Science Foundation (grant SES-0518222). First draft: March 24, 2007. I thank Bob Powell for his extensive comments as well as Robert Walker, Kris Ramsay, Jeff Ritter, Christina Schneider, Ron Hassner, Andy Kydd, Art Stein, Shuhei Kurizaki, Barry O'Neill, Dan Posner, Ethan Bueno de Mesquita, and Rob Trager for illuminating discussions. I am grateful to the two anonymous referees whose suggestions have immensely improved the article. Presented at Washington University–St. Louis, University of California–Berkeley, University of Wisconsin–Madison, University of California–Los Angeles, University of Oxford, Yale University, University of Rochester, Princeton University, the SGIR Pan-European IR Conference (Turin, Italy), the 2007 meeting of the European Consortium for Political Research (Pisa, Italy), the Project on Polarization and Conflict (Palma de Mallorca, Spain), the 2008 meeting of the Midwest Political Science Association (Chicago), and the 2009 Workshop on Rationality and Conflict (Cowles Foundation, Yale University).

During the last days of September 1950, the U.S. administration faced a momentous decision about what to do in Korea: should American forces stop at the 38th parallel, as originally planned, or should they continue into North Korea, and turn the conflict from a war of liberation into a war of unification? The North Koreans could effect no organized resistance to the onslaught of the U.N. forces, and the only uncertainty clouding the issue had to do with the behavior of the Chinese Communists: would the People's Republic of China (PRC) intervene to forestall unification of Korea on American terms or not?

After some hesitation and an effort to ascertain Chinese intent, the U.S. administration concluded that the risk of Chinese intervention was negligible and therefore the gamble was worth taking. One crucial factor in that estimate was the lack of obvious military preparations that China would have to undertake had it seriously intended to wage war on the United States. In particular, the PRC had not sent troops in significant numbers south of the Yalu River, it had not prepared Beijing for possible aerial raids, it had not mobilized economic or manpower resources, and it had failed to move when it made best sense to do so from a military standpoint—right after General MacArthur's landing at Inchon. All the Chinese appeared to have done was issue propaganda statements in government-controlled media, send somewhat contradictory messages through a diplomatic channel known to be distrusted by the Americans, fail to make a direct statement to the United Nations, and move some token forces of "volunteers" into North Korea. Even in late November, the Far East Command estimated that there were no more than about 70,000 of these "volunteers" to face over 440,000 U.N. troops of "vastly superior firepower."[1] Confident of success, General MacArthur launched the "home by Christmas" offensive on November 24.

This U.N. offensive was shattered in a mass Chinese counter-attack. Unbeknownst to U.N. Command, the Chinese had managed to move over 300,000 crack troops into North Korea. As Appleman documents, their armies had marched in complete secrecy "over circuitous mountain roads" with

defense measures that required that during the day "every man, animal, and piece of equipment were to be concealed and camouflaged. [...] When CCF units were compelled for any reason to march by day, they were under standing orders for every man to stop in his tracks and remain motionless if aircraft appeared overhead. Officers were empowered to shoot down immediately any man who violated this order."[2] This discipline had enabled the PRC to deploy vast numbers of troops in Korea without being discovered by aerial reconnaissance prior to actual contact.

But if the Chinese wanted to deter the Americans, why did they not make their mobilization public? When they knew the Americans doubted their resolve, why did they not choose an action that would reveal it? Whereas it is doubtless true that the Chinese benefitted from the tactical surprise once fighting began, they practically ensured that the Americans would not believe their threats. As Schelling puts it,

> It is not easy to explain why the Chinese entered North Korea so secretly and so suddenly. Had they wanted to stop the United Nations forces at the level, say, of Pyongyang, to protect their own border and territory, a conspicuous early entry in force might have found the U.N. Command content with its accomplishment and in no mood to fight a second war, against Chinese armies, for the remainder of North Korea. They chose instead to launch a surprise attack, with stunning tactical advantages but no prospect of deterrence.[3]

This behavior is indeed puzzling, especially when we consider the logic of costly signaling in crisis bargaining. When two opponents face each other with conflicting demands, the only way to extract concessions is by persuading the other that rejecting the demand would lead to highly unpleasant consequences such as war. The focus is on credible communication of one's intent to wage war should one's demands are not met. As is well known, to achieve credibility, an actor must engage

in an action which he would not have taken if he were unresolved even if the act of taking it would cause the opponent to become convinced that he is resolved. In other words, the action must be sufficiently costly or risky (or both) to make bluffing unattractive. Because a weak actor would not attempt to bluff his way into concessions with such an action, the act of taking it signals strength. Conversely, the absence of such an act can be taken as *prima facie* evidence of weakness.

In this light, the American administration was justified in drawing what turned out to be a wildly incorrect assessment about Chinese intent. The Chinese had not backed up their threats with any costly or risky actions, and even their demands had been somewhat watered down. For instance, at one point they said that it would be acceptable for South Korean troops to cross the parallel as long as the American forces remained south of it. This unwillingness by the Chinese to take actions that were available to them, and that they could have expected to produce concessions from the U.S. at an acceptable cost provided they were resolved to forestall unification, eventually persuaded the Americans that the threats were not serious, causing them to embark on unification.[4]

Since the Chinese goal was to deter unification, the logic of crisis bargaining suggests that the Chinese should not have concealed their preparations, and should have made the (admittedly much riskier) public demand for U.N. forces to remain south of the parallel. The fact that concealment had significant tactical advantages cannot, by itself, explain the decision to mobilize in secret because such an argument presupposes that the Chinese preferred to fight over Korea rather than prevent unification through deterrence, which is a highly dubious assumption.

In this article, I propose a development of our crisis bargaining models that could help shed some light on the puzzling failure to signal strength. First, I show that in a war, a strong player can obtain serious tactical advantage from an opponent who mistakenly believes him to be weak. This is intuitive and unsurprising although it is not without merit to have this emerge as result of optimal behavior by both actors instead of assuming it. Second, I consider a crisis model of the type in which

strong actors can obtain better negotiated outcomes when their opponent correctly infers that they are strong. I show that when bargaining in a crisis can end in war, a strong actor has contradictory incentives. On one hand, he wants to obtain a better negotiated deal, which requires him to convince his opponent that he is strong. On the other hand, should persuasion fail and war break out, he wants his opponent to believe that he is weak. Somehow, this actor must simultaneously signal strength and weakness.

I show that this contradiction is resolved in equilibrium by the strong actor sometimes feigning weakness during the crisis bargaining phase itself. He pretends to be weak by mimicking the smaller demand of a weak type. Even though this puts him at a disadvantage at the negotiation table, the loss is offset by the gain of tactical surprise on the battlefield that he can achieve if war follows anyway. This explanation also provides a rationale for the Chinese decision to forego the potential benefits of deterrence in order to gain tactical advantages in case deterrence failed.[5]

1 Signaling Strength in Crises

When two actors with conflicting interests lock horns in a crisis, the only way to secure concessions is to convince the opponent that such concessions, however painful, are preferable to the consequences of failure to comply with one's demands. In an interstate crisis, the threatened consequences are in the form of a costly and risky war. The stronger an actor is, the worse the expected war outcome for the adversary, and the more that adversary should be prepared to concede in order to avoid it. If there is one conclusion that emerges from our studies of crisis bargaining, it is that actors must signal credibly their strength if they are to obtain better deals from their opponents. Pretending to be weak does not pay.

Loosely speaking, the logic goes as follows. If an actor's expected payoff from war is high, his

minimally acceptable peace terms are more demanding relative to what they would have been if he were weak. Because actors are loath to concede more than is absolutely necessary, they need to ascertain what the minimally acceptable terms of the opponent might be. A simple assertion from an actor that he expects to do well in war will not do. If the opponent were to believe it and concede, there would be no risk or cost in making that statement. But then even a weak actor could assert it, which means that the opponent cannot take it at face value. The only way to persuade the opponent that one is strong is by taking an action that is so costly or risky that even if it were to succeed the weak type could not benefit from imitating it.

We have studied many mechanisms that allow a strong actor to distinguish himself from a weak one by taking some such action. For instance, an actor could make public statements that increase the domestic political costs of backing down, allow his domestic political opponents to contradict him for political gain, put his international reputation on the line, engage both domestic and international audiences, or generate an autonomous risk of inadvertent war.[6] As Banks has proven for a general class of models, strong types can expect to obtain better negotiated deals but only at the cost of taking actions that are too risky for the weak types to imitate.[7]

The crisis bargaining models that are central to these studies rely on a conceptualization of war as a costly lottery. Both actors must pay to participate in it but only one can win it. The expected payoff from war, usually referred to as the *distribution of power*, is a fundamental primitive in these models and is assumed to be exogenous. This assumption is carried over to the crisis bargaining models that treat war as a process rather than a costly lottery.[8]

Why does it matter that the distribution of power is assumed to be exogenous? For one, if we maintain this assumption, we cannot study military investment decisions because these presumably change the distribution of capabilities, and as such influence the distribution of power. Powell shows that when the expected payoff from war depends on strategic decisions about how to allocate

resources between consumption and arming, the necessity to spend on mutual deterrence creates a commitment problem which may lead to war when peace becomes too expensive to maintain.[9]

More directly related to crisis bargaining, this assumption excludes any actions that might alter the distribution of power. Slantchev (2005) argues that military moves—mobilization and deployment of troops, for instance—must necessarily affect it, and as such their use as instruments of coercion may have effects that do not obtain in models that do not take that into account. He shows that strong types do not, in fact, have to run higher risks in order to obtain better deals: the costliness of increasing military capability discourages bluffing while the concomitant improvement in the distribution of power reduces the opponent's expected war payoff and makes her more likely to concede.

These are theoretical reasons for treating the distribution of power as endogenous. The puzzle of Chinese intervention in the Korean War suggests at least one substantive reason to do so. As the admittedly cursory sketch of that episode illustrates, the PRC concealed its military preparations so thoroughly to gain tactical surprise. It was well known at the time that the superior air power of the U.N. forces put the Chinese at a serious disadvantage, which is why they tried to hard to obtain Soviet air cover for their land action.[10] If they were to expose their preparations, they risked having their forces annihilated before getting a change to engage the enemy. If the U.S. administration had made up its mind on unification, the revelation of the extent of Chinese mobilization could have also caused the United States to increase its effort in the war, which would similarly have jeopardized the chances of success of the PRC offensive.[11] The upshot is that for both actors, the expected payoff from war depended on the behavior they thought their opponent might engage in. If the Chinese revealed their mobilization, they might have succeeded in deterring the U.S. but they might have also considerably reduced their payoff from war if deterrence failed. If, on the other hand, they concealed their mobilization, they might not have been able to deter the U.S. but they would have

increased their payoff from war. In other words, the expected distribution of power depended on the actions taken during the crisis.

This episode not only provides a rationale for treating the distribution of power as endogenous, it also suggests a particular *timing* of decisions if one is interested in investigating analogous cases. In Powell's and Slantchev's models, actors make their military allocation decisions that fix the distribution of power for the duration of the war *before* the actual choice to attack.[12] The decision to fight is then taken after they observe each other's military preparations in light of the distribution of power that results from their actions. The Chinese tactic in the Korean War intervention, on the other hand, was to conceal the actual distribution of power until *after* the battle was joined. That is, they managed to lull the Americans into a false sense of security which was designed to prevent them from formulating an even more formidable offensive plan that would have attacked whatever vulnerability the Chinese revealed. In that sense, the episode suggests that we might want to think about war fighting decisions made *after* bargaining breaks down but in the light of information revealed *during* the bargaining phase.

One simple model with a structure that could address this situation would be an ultimatum crisis bargaining game in which the distribution of power is endogenously determined by actions taken after the ultimatum is rejected. This means that the expected payoff from war will depend on what the actors do when they go to war but that these decisions will be based on the information they obtain during the crisis. This structure allows us to investigate the contradictory incentives the Chinese faced in November: on one hand they wanted to signal that they are serious and the Americans should not advance to the Yalu River, but on the other hand they wanted to keep the Americans in the dark about their actual military preparations. As we shall see, this dilemma appears in the model in the following terms: should the strong actor choose a demanding ultimatum that would reveal his strength but put him at a fighting disadvantage if the demand is rejected, or should

he choose a middling demand that is not very attractive and will cause the opponent to think he might be weak but that would give him a tactical advantage if it is rejected?

2 THE MODEL

The model is designed as a simple setting that captures the contradictory incentives of strong players, and has three characteristics: (i) bargaining—an ultimatum to distribute an infinitely divisible benefit; (ii) endogenous distribution of power—military effort determines the expected payoff from war; and (iii) signaling—military effort can be contingent on information obtained from the crisis bargaining phase.

Two risk-neutral players, $i \in \{1, 2\}$ are disputing the two-way partition of a continuously divisible benefit represented by the interval $[0, 1]$. An agreement is a pair $(x, 1-x)$, where $x \in [0, 1]$ is player 1's share and $1-x$ is player 2's share. The players have strictly opposed preferences with $u_1(x) = x$ and $u_2(x) = 1 - x$. Player 1 begins by making a take-it-or-leave-it demand $x \in [0, 1]$ that player 2 can either accept or reject.[13] If she accepts, the game ends with the agreement $(x, 1 - x)$. If she rejects, she decides whether to mobilize additional resources, at cost $k_2 > 0$, or fight with what she already has. In any case, war occurs and each player pays costs $c_i > 0$. The winner obtains the entire benefit.

The outcome of the war depends on the distribution of power summarized by the probability that player 1 will win. This probability itself depends on player 2's arming choice: if she mobilizes additional resources, player 1's chances of victory decrease. We shall leave the precise functional form of the relationship between arming and victory unspecified. Instead, assume that player 1 can be either weak or strong. If player 2 does not arm, the weak type prevails in the war with probability w_n and the strong type prevails with probability $s_n > w_n$. If player 2 arms, the weak type prevails

with probability $w_a < w_n$, and the strong type prevails with probability $s_a < s_n$ such that $s_a > w_a$ (that is, player 2's additional mobilization cannot make the strong type's chance of winning lower than the weak type's). If player 1 is weak, his expected war payoffs are $W_w^n = w_n - c_1$ if player 2 does not arm, and $W_w^a = w_a - c_1$ if she does. If player 1 is strong, his expected war payoffs are $W_s^n = s_n - c_1$ if player 2 does not arm, and $W_s^a = s_a - c_1$ if she does.

Player 2's war payoff against a weak opponent is $1-w_n-c_2$ without arming, and $1-w_a-c_2-k_2$ with arming. Hence, she will not arm against a weak type when $k_2 > w_n - w_a$. Analogously, her war payoff against a strong opponent is $1-s_n-c_2$ without arming, and $1-s_a-c_2-k_2$ with arming. Hence, she will arm against the strong type when $k_2 < s_n - s_a$. To make the model interesting, make the following:

ASSUMPTION 1. The marginal effect of building arms on the probability of winning can only justify its cost if the opponent is strong: $w_n - w_a < k_2 < s_n - s_a$.

To ensure that this interval exists, we require that $s_n - w_n > s_a - w_a$. Although this specifies what player 2 would do if she knew her opponent's type, she is unsure about it. Player 1 knows whether he is weak or strong, but player 2 believes that he is strong with probability p and weak with probability $1 - p$, and this belief is common knowledge.

3 ANALYSIS

Under our assumptions, player 2 will certainly arm if she believes her opponent is strong, and will not if she believes he is weak. In between these certainties, her arming decision depends on her posterior belief that she acquires after player 1's ultimatum. Let q denote the (possibly updated) belief that player 1 is strong after his demand. Player 2's war payoffs are $W_2^n(q) = q(1-s_n) + (1-q)(1-w_n) - c_2$, and $W_2^a(q) = q(1-s_a) + (1-q)(1-w_a) - c_2 - k_2$, where the

superscript denotes her arming choice. Since player 2 will arm when $W_2^a(q) > W_2^n(q)$, it follows that she will arm when:

$$q > \frac{k_2 - (w_n - w_a)}{(s_n - s_a) - (w_n - w_a)} \equiv q_a. \tag{1}$$

Our assumptions ensure that $q_a \in (0, 1)$. We conclude that player 2 will arm if $q > q_a$ and will not arm otherwise.

We assumed that player 2's arming will reduce player 1's expected payoff from war, and we found that her decision to do so depends on her belief that player 1 is strong. In the "tactical game" that follows the rejection of the crisis ultimatum, player 1's incentives are clear: he wants player 2 to believe that he is weak. (As we shall see in an extension of the model, these incentives also arise in exactly the same way if we model the arming decisions of both sides explicitly.)

The question that we really wish to investigate is whether these incentives extend to the crisis game itself: after all, the only way to obtain better deals through bargaining is by convincing player 2 that one is strong. I will show that this game has *feint equilibria*. In these, player 1 always makes a low-value, low-risk demand if he is weak. If he is strong, however, he sometimes makes a high-value, high-risk demand (which credibly signals his strength) but sometimes pretends to be weak by making the low-value, low-risk demand. The risks and the intensity of fighting are endogenous: player 2 rejects the low-value demand with lower probability than the high-value demand, but arms only when rejecting the high-value demand. Hence, the strong player 1 foregoes some of the bargaining benefit that would arise from revealing his type in order to obtain some of the fighting benefit that would arise should negotiations fail and player 2 mistakenly believes he is weak.

3.1 THE FEINT EQUILIBRIA

The construction of feint equilibria proceeds in several steps. First, I show that the separating demand that only the strong type is willing to make must be larger, risker, and costlier than the demand that both he and the weak type are willing to pool on. Second, I show that player 2 would reject very large demands and accept very small demands regardless of her beliefs. This renders meaningless attempts to manipulate her beliefs (through feints or signaling) with demands in those ranges. Third, I specify intuitive beliefs for demands where player 2's reaction does depend on her beliefs: the more player 1 demands, the more player 2 is convinced that he is strong. Fourth, I show that when the fighting benefit from deceiving player 2 is not much greater than the low-value demand, the feint could be riskless (that is, the low-value demand might carry no risk of war). I then derive a sufficient condition—the fighting advantage from a feint is very large compared to the low-value demand—that guarantees that the feint must carry a strictly positive risk of war.

Let $\underline{x}$ denote the demand that both types are willing to make, and let $\overline{x}$ denote the demand that only the strong type is willing to make. Let $\underline{r}$ denote the probability with which player 2 rejects $\underline{x}$ without arming, and $\overline{r}$ denote the probability with which she rejects $\overline{x}$ with arming. Incentive-compatibility equilibrium conditions require that $\underline{x}$ is the low-value demand, and $\underline{r}$ is the low risk associated with it, as the following lemma demonstrates.

LEMMA 1. *In any feint equilibrium,* $\underline{x} < \overline{x}$ *and* $\underline{r} < \overline{r}$.

Proof. In a feint equilibrium,

$$\underline{r}W_s^n + (1-\underline{r})\underline{x} = \overline{r}W_s^a + (1-\overline{r})\overline{x} \qquad (\text{IC}_\text{s})$$

$$\overline{r}W_w^a + (1-\overline{r})\overline{x} \leq \underline{r}W_w^n + (1-\underline{r})\underline{x}. \qquad (\text{IC}_\text{w})$$

Adding these inequalities gives us $\underline{r}(s_n - w_n) \leq \overline{r}(s_a - w_a)$, but from our assumptions we know

that $s_n - s_a > w_n - w_a$, which implies that $s_n - w_n > s_a - w_a$, so this condition requires that $\underline{r} < \overline{r}$, as claimed. Furthermore, because $W_s^n > W_s^a$, this implies that if $\underline{x} \geq \overline{x}$, the indifference condition for the strong type cannot be satisfied. Therefore, $\underline{x} < \overline{x}$. □

Player 2 does not arm for any $q \leq q_a$, so the *best* war payoff (without arming) is $\overline{W}_2 = 1 - w_n - c_2$. She arms for any $q > q_a$, so the *worst* war payoff (with arming) is $\underline{W}_2 = 1 - s_a - c_2 - k_2$. Thus, *in any equilibrium,* if player 1 demands $x < x_1 = 1 - \overline{W}_2$, player 2 will accept, and if he demands $x > x_2 = 1 - \underline{W}_2$, she will reject. The only belief-contingent responses are to demands in $[x_1, x_2]$. Since player 2 must reject some offers with positive probability, we are interested in beliefs that leave her indifferent between accepting the demand, and rejecting it. Let q solve $x = 1 - W_2^n(q)$ if $q \leq q_a$, and solve $x = 1 - W_2^a(q)$ otherwise. This yields the cut-point demand $x_a = x_1 + q_a(s_n - w_n)$ where $x_a = 1 - W_2^n(q_a) = 1 - W_2^a(q_a)$. Define the posterior beliefs as follows:

$$q(x) = \begin{cases} 0 & \text{if } x < x_1 \\ \frac{x - w_n - c_2}{s_n - w_n} & \text{if } x_1 \leq x \leq x_a \\ \frac{x - w_a - c_2 - k_2}{s_a - w_a} & x_a < x \leq x_2 \\ 1 & \text{if } x > x_2. \end{cases} \tag{2}$$

It is clear from inspection that $q(x)$ is continuous because $W_2^n(q_a) = W_2^a(q_a)$, and strictly increasing (which implies the belief is unique). These beliefs are intuitively appealing: the more player 1 demands, the higher the probability that player 2 will assign to him being strong. The low-value demand is the largest demand player 2 would accept without arming: $\underline{x} = x_a$. The high-value demand is the largest demand she would accepting with arming: $\overline{x} = x_2$.

The weak type strictly prefers player 2 to accept even the low-value demand: $\underline{x} - W_w^n = C + q_a(s_n - w_n) > 0$, where $C = c_1 + c_2$. The strong type, on the other hand, might actually prefer

player 2 to reject the low-value demand and fight unprepared. In particular, if this demand is worse than fighting even a prepared opponent, then its risk must be strictly positive or else the strong type would not be willing to make it. Observe that (IC_s) gives us the risk of the high-value demand:

$$\overline{r} = \frac{\overline{x} - \underline{x} - \underline{r}\left(W_s^n - \underline{x}\right)}{\overline{x} - W_s^a}. \tag{3}$$

Since $\overline{r} < 1$ must be satisfied, $\underline{r} > 0$ will hold whenever $\underline{x} \leq W_s^a$, or:

$$q_a \leq \frac{s_a - w_n - C}{s_n - w_n} \equiv q_d. \tag{D}$$

In this situation, the strong type attempts to deceive player 2 into incorrectly rejecting the low-value demand and entering the war unprepared. It is always possible to construct a feint equilibrium with a riskless low-value demand when (D) is not satisfied. Although one can also construct equilibria with a strictly positive risk, these are all Pareto inferior. When (D) is satisfied, however, the low-value demand must be risky. In this case, the risk should not bee too high or the weak type would not be willing to run it, preferring to deviate to the largest possible riskless demand, x_1. The upper bound on the risk that makes such a deviation unprofitable is $\underline{r} \leq (\underline{x} - x_1) / \left(\underline{x} - W_w^n\right)$. Since $\underline{r}$ should be neither too large nor too small, the necessary condition that admits the existence of such values is that the upper bound is at least as large as the lower bound, or

$$q_a \geq \frac{q_d C}{C + s_n - s_a} \equiv \underline{q}. \tag{4}$$

It is worth emphasizing that (4) is not binding when (D) is not satisfied because in this case the lower bound is at 0, which trivially satisfies the requirements.

PROPOSITION 1. *If $p > q_a \geq \underline{q}$, there are perfect Bayesian equilibria in which the weak player 1 demands $\underline{x} = x_a$, and the strong player 1 demands $\underline{x}$ with probability $\phi = (1-p)q_a/[p(1-q_a)]$ and $\overline{x} = x_2$ with probability $1 - \phi$. Player 2 accepts any $x \leq x_1$, rejects any $x \in (x_1, \underline{x}]$ with*

probability $\underline{r}$*, rejects any* $x \in (\underline{x}, \overline{x}]$ *with probability* $\overline{r}$*, and rejects any* $x > \overline{x}$ *for sure. The rejection probabilities are* $\underline{r} = 0$ *if* (D) *is not satisfied, and* $\underline{r} \in \left(1 - \frac{s_n - s_a}{W_s^n - \underline{x}}, \frac{\underline{x} - x_1}{\underline{x} - W_w^n}\right)$ *otherwise, and* $\overline{r}$ *is as defined in* (3)*. On and off the path, beliefs are defined in* (2)*.*

The intuition for this result is as follows. If player 1 is strong he can credibly reveal this provided he is willing to run higher risks of war in which the opponent is prepared. The mechanism is the same as in the standard costly signaling models. To prevent bluffing from weak types, the strong types must incur costs and risks that the weak ones would not be willing to incur even if doing so would convince the opponent they are strong. Bluffing, however, is not the only strategic problem player 2 faces: sandbagging is another.

Player 2 reacts to the low-value demand by accepting it with a higher probability. On one hand, this is unattractive to the strong type: the terms are worse than the separating high-value demand, and there is a good chance that it will be accepted. On the other hand, this is attractive: the risk of war is lower, and even when it is positive the war that follows will be against an unprepared opponent. In equilibrium, the strong type balances these trade-offs and sometimes feigns weakness.

When (D) is satisfied, the low-value demand is too unattractive to the strong type: he will only feign weakness if there is a chance that it will be rejected. In this situation, minimizing the risk associated with this demand has ambiguous social welfare implications, which is why Proposition 1 specifies the range of risks that can be supported in equilibrium.

3.2 Selection of a Signaling Equilibrium

Like most signaling games, this one has many equilibria. Of particular interest are ones in which the strong type either fully or partially reveals his type. The claim of Lemma 1 holds for any fully or partially separating equilibrium where the weak type demands some $\underline{x}$ and the strong type demands some $\overline{x}$, with player 2 rejecting the former with probability $\underline{r}$ and the latter with probability $\overline{r}$.[14]

SEPARATING EQUILIBRIA. In a separating equilibrium, $q(\underline{x}) = 0$ and $q(\overline{x}) = 1$, which immediately implies that the high-value demand will be exactly the same as in the feint equilibria, or $\overline{x} = x_2$. The low-value demand is the highest demand the weak player 1 can make provided that making it leads player 2 to infer that he is weak, or $\underline{x} = x_1$. Because preventing deviations would require positive probabilities of rejection of demands in $[\underline{x}, \overline{x}]$, it follows that the equilibrium beliefs over that range would have to be exactly the same as in the feint equilibria as well. Thus, one substantive difference between the separating and feint equilibria is that in the former the weak type gets a strictly lower payoff because the low-value demand is smaller.

The intuition is that because player 2 would conclude that the opponent is weak after seeing this demand in the separating equilibrium, her expected payoff from rejecting it and fighting without arming will be much higher than the corresponding payoff in a feint equilibrium where she believes there might be a chance that her opponent is strong. This implies that her expected payoff from rejection is strictly larger in the separating equilibrium, so the acceptable low-value demand will be correspondingly smaller. This is particularly evident in the case where the low-value demand is riskless in the feint equilibria as well. Observe that in a separating equilibrium, the low-value demand would reflect the most player 2 would be willing to concede when she is certain that her opponent is weak. In a feint equilibrium, on the other hand, the corresponding low-value demand reflects what she would be willing to concede when is suspects her opponent might actually be strong.

Recall that the high-value demand is the same in both types of equilibrium. The fact that the a riskless low-value demand is strictly better in a feint equilibrium where the strong type is indifferent between the two demands implies that the risk of the high-value demand must be lower in the feint equilibrium. This is so because the strong type's expected payoff from the high-value demand is strictly decreasing in the risk of war, so if the low-value demand increases, the risk of the high-

value demand must decrease if he is to remain indifferent. This implies that the expected payoff for player 1 is *strictly higher in the feint equilibria*, which gives one possible reason for selecting them in situations where both equilibrium types exist.

When the conditions stated in Proposition 1 are not satisfied, feint equilibria will not exist. In particular, when $q_a < \underline{q}$, the weak type would want to make the highest possible riskless demand x_1 that would reveal his weakness. In other words, this is where the separating equilibrium would still exist. In fact, separating equilibria can be supported with the assessment used in the proof of Proposition 1 with appropriate minor adjustments.

SEMI-SEPARATING (BLUFFING) EQUILIBRIA. In a semi-separating equilibrium, the weak type sometimes demands $\underline{x}$ but occasionally bluffs by demanding $\overline{x}$, and the strong type demands $\overline{x}$ for sure. Because the weak type is the only one demanding $\underline{x}$ with positive probability, it follows that in such equilibria, $\underline{x} = x_1$. The high-value demand has to be such that the strong type would be unwilling to separate by making a larger demand. The most straightforward way to accomplish that is to use the same belief system as in the feint equilibria, but require that player 2 certainly reject any $x > \overline{x}$. (Since player 2 is indifferent for any $x \in [x_1, x_2]$ and rejects any $x > x_2$ regardless of beliefs, this is clearly possible.)

For instance, we can support $\overline{x} = x_a$ in a bluffing equilibrium as follows. Let $\overline{r} > \underline{r} = 0$ be such that the weak type is indifferent between $\overline{x}$ and $\underline{x} = x_1$, or $\overline{r} = (x_a - x_1)/(x_a - W_w^n)$. Consider a strategy for player 2 such that she accepts any $x \leq x_1$, rejects any $x \in (x_1, \overline{x}]$ with probability $\overline{r}$, and rejects any $x > \overline{x}$ with certainty. This strategy is sequentially rational with the assessment in (2). As in the feint equilibrium, deviation to $x \in (x_1, \overline{x})$ merely produces peace terms that are worse than $\overline{x}$ with the same risks and same type of war (without player 2 arming), so it cannot be profitable. Any deviation to $x > \overline{x}$ results in a certain war. The strong type cannot profit if $\overline{r} W_s^n + (1 - \overline{r}) x_a \geq W_s^a$. Since $W_s^n > W_s^a$, the sufficient condition for this is $x_a \geq W_s^a$, or (D) not being satisfied. Recall

that this means that the peace terms are at least as good for the strong type as fighting an armed opponent. When this is not the case, the peace terms are so bad that the strong type's only incentive to demand them is in the possibility that player 2 might reject them and fight a war unprepared. This means the risk of war should be sufficiently high, or $\overline{r} \geq (W_s^a - x_a)/(W_s^n - x_a)$. There are ranges for the parameters that satisfy this requirement.

It is always possible to satisfy the weak type's indifference condition for a sufficiently low risk for $\overline{x}$. This risk will also deter deviations that cause certain war when (D) is not satisfied, and for some parameter configurations even when (D) is satisfied. In either case, the binding condition for the existence of bluffing equilibria is in the high risk associated with making large demands. This risk dampens the strong type's ability to separate and keeps him locked into making a demand so low that even the weak type is willing to mimic it. If we are willing to preserve the substantively more appealing monotonicity exhibited by the rejection probability in the feint equilibria, then this artificial constraint will disappear, and so will the bluffing equilibria. In other words, there are strong substantive reasons to select the feint equilibria over the fully revealing or bluffing equilibria when these types coexist.

3.3 The Likelihood of Feints

The probability with which the strong type feigns weakness is $\phi = (1-p)q_a/[p(1-q_a)]$, so:

$$\frac{\partial \phi}{\partial p} = \frac{-q_a}{p^2(1-q_a)} < 0, \quad \text{and} \quad \frac{\partial \phi}{\partial q_a} = \frac{1-p}{p(1-q_a)^2} > 0.$$

The more pessimistic player 2 is, the more likely is the strong player 1 to feign weakness. The second comparative static is more interesting: since the feint probability is strictly increasing in the critical belief q_a, we can conduct additional comparative statics on this belief as defined in (1). It is immediate that the higher the marginal cost of arming to player 2, the more likely is player 1

to feign weakness (because even relatively low probabilities of him being weak can discourage her from arming when doing so is costly).

THE BENEFIT OF ARMING. Player 2's marginal benefit from arming depends on player 1's type and the technology of fighting implicit in the definition of the probabilities of winning. Let $b_w = w_n - w_a < k_2$ be her benefit from arming against a weak opponent, and $b_s = s_n - s_a > k_2$ be the benefit from arming against a strong opponent. Since $q_a = (k_2 - b_w)/(b_s - b_w)$, we now obtain:

$$\frac{\partial q_a}{\partial b_w} = \frac{k_2 - b_s}{(b_s - b_w)^2} < 0 \quad \text{and} \quad \frac{\partial q_a}{\partial b_s} = \frac{b_w - k_2}{(b_s - b_w)^2} < 0.$$

As player 1's benefit from player 2's failure to arm (e.g., b_s) goes up, the probability of a feint goes down. This is surprising because it says that as the benefit of successful deception *increases*, the likelihood that player 1 will attempt to deceive player 2 *decreases*.

At first glance, it would appear that the converse should be true: after all, the strong type can benefit from deception most when his war payoff against an unprepared opponent is much higher than his payoff from an armed one. This logic, however, does not consider player 2's response. If the marginal benefit from arming increases (wether against a weak or a strong opponent), then player 2 would arm even if she is less convinced that the opponent is strong. In other words, when player 2 expects to get a significantly worse payoff if she fails to arm and can mitigate this disaster by arming, she will arm as a precaution even though she might not assign a great probability to her opponent being strong. Because larger demands cause her to revise her beliefs upwards, this implies that the largest demand player 1 can make without provoking arming upon rejection decreases. This reduces the strong type's incentives to feign weakness.

RELATIVE POWER. We can think of $s_n - w_n$ as the strong type's power relative to the weak type's when player 2 is unprepared, and $s_a - w_a$ as the analogous relative power when she is prepared. We

now have:

$$\frac{\partial q_a}{\partial s_n} = \frac{b_w - k_2}{(b_s - b_w)^2} < 0, \quad \frac{\partial^2 q_a}{\partial s_n \partial w_n} = \frac{b_s + b_w - 2k_2}{(b_s - b_w)^3} \gtrless 0 \Leftrightarrow b_s - k_2 \gtrless k_2 - b_w.$$

The interpretation of the partial derivative is straightforward: an increase in the strong type's probability of winning against a disarmed opponent increases the risks from rejecting the low-value offer, and increases player 2's propensity to arm. This reduces the value of the feint to the strong type, and he feints less often. The cross-partial shows that the magnitude of this effect depends on the expected war-time performance of the weak type as mediated by the marginal cost of arming for player 2. The cross-partial is positive when the marginal cost of arming is relatively small ($b_s - k_2 > k_2 - b_w$). In that case, an increase in relative power of the strong type due to a drop in w_n magnifies the detrimental effect of s_n and leads to a sharp decline in the desirability of the low-value offer. In other words, because player 2 finds it cheaper to arm, she counters this increase in the strong type's relative power more readily. If, on the other hand, the cross-partial is positive because her marginal costs of arming are high, then an increase in relative power due to a drop in w_n, although unpleasant, does not lead to very drastic revisions of player 2's arming propensity. Even though she still arms more readily in response to an increase in relative power, the effect is muted because of the high costs of doing so.

Turning now to the effect of relative power against an armed opponent, we have:

$$\frac{\partial q_a}{\partial s_a} = \frac{k_2 - b_w}{(b_s - b_w)^2} > 0, \quad \frac{\partial^2 q_a}{\partial s_a \partial w_a} = \frac{b_s + b_w - 2k_2}{(b_s - b_w)^3} \gtrless 0 \Leftrightarrow b_s - k_2 \gtrless k_2 - b_w.$$

The direct effect of an increase in the strong type's probability of winning against a prepared opponent is perhaps surprising: the better this type expects to do in such a war, the *more* likely is he to feign weakness! To understand this, we must consider how player 2 responds to such an increase. Her benefit from war, even when fully prepared, decreases, which means that the terms she is prepared to concede in the high-value demand become more attractive to player 1. The strong

type will thus be willing to feign weakness either because the risk of making this demand increases or because the terms of the low-value offer improve considerably. Because player 2 expects to do rather poorly in a war against the strong type, the relative value of arming in effect declines and she becomes more willing to make concessions. The improvement in the terms of the low-value demand make it more attractive to the strong type, and makes him more likely to attempt a feint.

The cross-partial is the same as for the relative power against an unarmed opponent but because the effect of s_a is different, so is the overall interpretation. The cross-partial is positive when player 2's marginal cost of arming is small. With such costs, player 2 tends to arm even while relatively optimistic; that is, while she still believes with a relatively high probability that her opponent is weak. Thus, a decline in her expected war payoff due to an increase in the weak type's strength, w_a, affects her propensity to arm and she becomes less willing to do so. This increases the threshold belief for arming and magnifies the effect of increase in the strong type's relative power. In other words, even though the strong type's power relative to the weak type is not that pronounced when the latter is only moderately weak, the increase in his relative strength has a disproportionately large effect on player 2's incentive to arm when her costs are low. Conversely, when her cost of arming is high (and the cross-partial is negative), player 2 only arms when relatively pessimistic. This means that increases in the weak type's strength have a smaller marginal impact on her expected war payoff, and so her incentive to arm does not increase nearly as dramatically. This dampens the overall effect of an increase in the strong type's relative power.

LURING INTO WAR. When (D) is satisfied, the strong type prefers fighting an unprepared opponent to the peace terms from the low-value demand. The feint under these conditions can be interpreted as an attempt to lure the opponent into fighting by lulling him into a false sense of optimism. Not surprisingly, decreasing the costs of war makes this condition easier to satisfy. Somewhat less predictably, a decrease in player 2's marginal cost of arming does so as well. To see why this

should be so, observe that lowering k_2 effectively lowers the barrier to preventive arming, which in turn makes player 2 less willing to make concessions. This reduces the peace benefit from the low-value demand, and if the decline is sufficiently large, makes fighting an unprepared opponent more attractive to the strong type. Gauging the effect of s_n in this context is slightly more involved because q_a, q_d, and $\underline{q}$ are all decreasing in s_n. However, it can be shown that q_a decreases at a faster rate than $\underline{q}$, which means that for high enough values of s_n that satisfy (D), the necessary condition for the existence of the feint equilibria, $q_a \geq \underline{q}$, will be violated. The logic is as follows. As we have seen, increasing s_n lowers the arming threshold for player 2, which in turn lowers the terms of the low-value demand. If the strong type is to feign weakness, the risk associated with this demand must increase (so he can reap the benefits of war against an unprepared opponent). However, this makes the demand less attractive to the weak type, and when the risk is sufficiently high, (4) will fail, and he will not be willing to make the low-value demand, opting instead for a riskless x_1. In other words, as the advantages of deceiving player 2 increase for the strong type, he becomes less able to mislead her successfully.

4 Endogenous Tactical Incentives

The model I analyzed is tractable and transparent, which makes the exposition easier to follow. It is also generic because it leaves the functional form of the technology of war unspecified. However, player 1 does not have an opportunity to react to player 2's expected behavior once a demand is rejected. To study the problem with fully endogenous tactical incentives, we must model the technology of war explicitly. Although this limits the results somewhat, the importance of the question justifies the cost.

4.1 THE EXTENDED MODEL

The crisis game is the same as in the original model, the difference is what happens when players go to war. As before, player 1 makes an ultimatum demand x. If player 2 accepts, players receive $(x, 1-x)$, if she rejects a costly contest (war) occurs. The contest is a simultaneous-move game in which each player chooses a level of effort $m_i \geq 0$ at cost $c_i > 0$. The probability of winning is determined probabilistically by the ratio contest-success function $\pi_i(m_1, m_2) = m_i/(m_1 + m_2)$ if $m_1 + m_2 > 0$ and $\pi_i = 1/2$ otherwise.[15] The winner obtains the entire benefit, so player i's expected payoff from a contest is $\pi_i(m_1, m_2) - m_i/c_i$.

The game has one-sided incomplete information.[16] Player 2 knows her own cost of effort, c_2, but is unsure about the player 1's cost. Specifically, player 2 believes that player 1 is strong, $\overline{c}_1$ with probability p and weak, $\underline{c}_1 < \overline{c}_1$, with probability $1-p$. These beliefs are common knowledge. If the costs of effort are too high, then war is prohibitively costly and the game will carry no risk of bargaining breakdown. We thus make the following

ASSUMPTION 2. The uninformed player's costs are not too high: $c_2 > \sqrt{\underline{c}_1 \overline{c}_1}$.

Since the strategies for the crisis bargaining game would have to form an equilibrium in the contest continuation game, I analyze that first.

4.2 THE CONTEST ENDGAME

There are only two possibilities in the continuation game following player 1's demand: either player 2 will infer his type or not. If she infers the type, as she would after the separating high-value high-risk demand that only the strong type makes, the contest is one of complete information. If she can only partially infer it, as she would do after the low-value low-risk demand that the strong and the weak type pool on, the contest is one of asymmetric information where her posterior belief that

player 1 is strong is $q \in (0,1)$. I derive the expected equilibrium war payoffs for both situations, and then show that the more convinced player 2 becomes that player 1 is strong, the more intense her fighting effort gets. This worsens the strong type's war payoff, and gives him an incentive to mislead player 2 that he is weak. That is, I show that the incentive to feign weakness can arise fully endogenously.

4.2.1 Complete Information

Players optimize $\max_{m_i} \{m_i/(m_i+m_j) - m_i/c_i\}$, which yield the best responses $m_1^*(m_2) = \sqrt{c_1 m_2} - m_2$ and $m_2^*(m_1) = \sqrt{c_2 m_1} - m_1$ in an interior equilibrium. Solving the system of equations then gives us the equilibrium effort levels: $m_1^* = c_2\left(\frac{c_1}{c_1+c_2}\right)^2$ and $m_2^* = c_1\left(\frac{c_2}{c_1+c_2}\right)^2$. The equilibrium expected payoffs are:

$$W_1 = \left(\frac{c_1}{c_1+c_2}\right)^2 \quad \text{and} \quad W_2 = \left(\frac{c_2}{c_1+c_2}\right)^2. \tag{5}$$

Fighting is inefficient: $W_1 + W_2 < 1 \Leftrightarrow 0 < 2c_1c_2$. Players always have an incentive to negotiate a division of the good instead of fighting to win it all. Moreover, a mutually-acceptable peaceful division always exists. The rationalist puzzle that arises from war's inefficiency remains intact.[17]

4.2.2 One-Sided Asymmetric Information

Player 2, whose cost c_2 is common knowledge, believes that player 1 is strong with probability q and weak with probability $1-q$. Player 1 knows his own cost, and optimizes as he would under complete information, which yields:

$$m_1(m_2; c_1) = \max\left(\sqrt{c_1 m_2} - m_2, 0\right), \tag{6}$$

which eliminates some contests from consideration.

LEMMA 2. *In equilibrium, either both types of player 1 exert positive effort in the contest, or only the strong type does.*

This means that there are only two possibilities to consider: either both types of player 1 spend strictly positive effort (skirmish), or only the strong type does (war). The fanciful names are meant as reminders that contests in which the weak type participates are lower in intensity than conflicts in which only the strong type participates.

THE SKIRMISH EQUILIBRIUM. Let $\underline{m}_1 = m_1(m_2;\underline{c}_1)$ $\overline{m}_1 = m_1(m_2;\overline{c}_1)$ denote the effort levels of the weak and strong types, respectively. Because player 2 is unsure about player 1's type, her optimization problem is $\max_{m_2} \{q m_2/(\overline{m}_1 + m_2) + (1-q)m_2/(\underline{m}_1 + m_2) - m_2/c_2\}$. Her equilibrium effort level is

$$m_2^* = \underline{c}_1\overline{c}_1\left[\frac{f(q)}{g(q)}\right]^2, \tag{7}$$

where $f(q) = q\sqrt{\underline{c}_1} + (1-q)\sqrt{\overline{c}_1} > 0$ and $g(q) = \underline{c}_1\overline{c}_1/c_2 + q\underline{c}_1 + (1-q)\overline{c}_1 > 0$. The expected equilibrium war payoffs are:

$$W_1(q;c_1) = \left(1 - \frac{f(q)}{g(q)}\sqrt{\frac{\underline{c}_1\overline{c}_1}{c_1}}\right)^2 \quad \text{and} \quad W_2(q) = \left(q\underline{c}_1 + (1-q)\overline{c}_1\right)\left[\frac{f(q)}{g(q)}\right]^2.$$

In the skirmish equilibrium, $\underline{m}_1 > 0$, which means that $m_2^* < \underline{c}_1$, or:

$$q < \frac{\overline{c}_1\sqrt{\underline{c}_1}}{c_2\left(\sqrt{\overline{c}_1} - \sqrt{\underline{c}_1}\right)} \equiv q_s \tag{8}$$

is the necessary condition for this equilibrium to exist.

THE WAR EQUILIBRIUM. In this case, the weak type does not exert any effort in equilibrium, so $\underline{m}_1 = 0$. The strong type's optimal effort is still defined by (6). Player 2's maximization problem, $\max_{m_2} \{q m_2/(\overline{m}_1 + m_2) + (1-q) - m_2/c_2\}$, is simpler because whatever positive effort she expends, she will win outright if her opponent happens to be the weak type. The solution is:

$$m_2^* = \overline{c}_1\left(\frac{qc_2}{\overline{c}_1 + qc_2}\right)^2. \tag{9}$$

The expected equilibrium war payoffs are:

$$W_1(q;\overline{c}_1) = \left(\frac{\overline{c}_1}{\overline{c}_1 + qc_2}\right)^2 \quad \text{and} \quad W_2(q) = 1 - q + q\left(\frac{qc_2}{\overline{c}_1 + qc_2}\right)^2.$$

It is not difficult to verify that $q \geq q_s$ is the condition for the war equilibrium to exist. The two cases characterize the complete solution to the one-sided incomplete information contest.

4.3 The Sun Tzu Principle of Feigning Weakness

LEMMA 3 (Sun Tzu). *When player 1 exerts positive effort in the contest, his equilibrium payoff is decreasing in player 2's belief that he is strong.*

The logic behind the principle is straightforward. Player 2's equilibrium effort level is increasing in q: the more pessimistic she is, the higher the effort she will exert. This leads player 1 to compensate by increasing his own effort, leading to an overall decrease in his expected payoff because of the higher costs he incurs in the process. This parallels Sun Tzu's principle of feigning weakness which he stated as follows: "If your opponent is of choleric temper, seek to irritate him. Pretend to be weak, that he may grow arrogant" (6).

This result provides microfoundations for Assumption 1 in the original model. It is worth noting that Sun Tzu's principle is here derived as the result of optimal rational behavior in a contest under uncertainty. The upshot of this analysis is that the strong type's incentive to mislead player 2 in the strategic game arises in this model as well.

4.4 The Crisis Ultimatum

As it turns out, the method for constructing feint equilibria in this model is analogous to what we did in the simple one. I will only sketch the steps here.

EQUILIBRIUM BELIEFS. The belief-contingent responses are limited by the best and worst war payoffs that player 2 can expect.

LEMMA 4. *Let $x_1 = 1 - \overline{W}_2$ and $x_2 = 1 - \underline{W}_2$, where $\underline{W}_2 = W_2(\overline{c}_1, c_2)$ is player 2's expected payoff from a full information contest against a strong opponent and $\overline{W}_2 = W_2(\underline{c}_1, c_2)$ is her analogous payoff against a weak opponent. In any equilibrium, player 2 will accept any $x \leq x_1$ and reject any $x > x_2$ regardless of her beliefs.*

Since only the strong type ever demands $\overline{x}$ in equilibrium, rejection leads to a complete-information war against a strong opponent. With such a belief, she accepts any x such that $1 - x \geq \underline{W}_2$, and because player 1 has no incentive to demand less than what she is willing to accept, it follows that in equilibrium,

$$\overline{x} = 1 - \underline{W}_2 = x_2, \tag{10}$$

which is exactly the same as in the simple model.

Since both types make the low-value demand with positive probability, rejection leads to war with incomplete information with a posterior belief $q(\underline{x})$. Player 2's optimal effort is then given by (7) if the contest admits the skirmish equilibrium and by (9) otherwise. I shall use $W_2(q(x))$ to denote the expected payoff with the understanding that this notation refers to the appropriate payoff.[18] With such a belief, player 2 will accept any demand such that $1 - x \geq W_2(q(x))$. Because player 1 has no incentive to offer more than the absolute minimum necessary to obtain acceptance, it follows that in equilibrium,

$$\underline{x} = 1 - W_2(q(\underline{x})). \tag{11}$$

Because the low-value demand results in a belief-contingent response, $\underline{x} \in [x_1, x_2]$ with $q(\underline{x})$ satisfying (11). The following lemma proves that it is always possible to find such a belief.

LEMMA 5. *For any $x \in [x_1, x_2]$, there exists a unique $q(x) \in [0,1]$ that satisfies* (11). *Moreover, $q(x)$ is strictly increasing in x.*

We conclude that in any equilibrium, player 2 will accept any $x < x_1$, will reject any $x > x_2$, and can randomize between accepting and rejecting any $x \in [x_1, x_2]$ when her posterior beliefs are defined by Lemma 5. This is the exact analogue to the (on and off the path) beliefs we constructed in the simple model.

THE FEINT EQUILIBRIA. It is not difficult to verify that the analogue to the incentive-compatibility conditions in Lemma 1 obtains in this model as well. Letting $\underline{r}$ and $\overline{r}$ be the probabilities with which player 2 rejects $\underline{x}$ and $\overline{x}$, respectively, we know that $\underline{r} < \overline{r}$ and $\underline{x} < \overline{x}$ in any feint equilibrium.

As before, there are conditions that permit $\underline{r} = 0$. The low-value demand can be riskless only when the incentives of the weak and the strong types are aligned given these peace terms: $W_1(q(\underline{x}); \overline{c}_1) \leq \underline{x}$. This implies that the low-value demand cannot be smaller than $\widetilde{x} = W_1(q(\widetilde{x}); \overline{c}_1)$. Hence, the lower bound on the low-value demand is $x^* = \max[\widetilde{x}, x_1]$.

Finally, Bayes rule yields the feint probability: $\phi = q(\underline{x})(1-p)/[p(1-q(\underline{x}))]$, which requires $p > q(\underline{x})$. Because $q(x)$ is increasing and $q(\overline{x}) = 1 > p$, this puts an upper bound on the low-value demand. In particular, there exists $x^{**} < \overline{x}$ such that $q(x^{**}) = p$, so that only $x < x^{**}$ can be supported as a low-value demand in a feint equilibrium. Observe in particular that $x^* = x_1$ ensures that $x^{**} > x^*$.

PROPOSITION 2. *Any $\underline{x} \in [x^*, x^{**}]$ can be supported in a feint equilibrium with a riskless low-value demand and $\overline{x} = x_2$. Player 2 accepts any $x \leq \underline{x}$, rejects any $x \in (\underline{x}; \overline{x}]$ with probability $r(x) = \frac{x - \underline{x}}{x - W_1(q(x); \overline{c}_1)}$, and rejects any $x > \overline{x}$ with certainty. On and off the path, her beliefs are defined in Lemma 5.*

Although it is possible construct feint equilibria with $\underline{r} > 0$ when the low-value demand can

be riskless, a social welfare argument would select the Pareto-optimal equilibrium with $\underline{r} = 0$. There is, however, a major difference between this model and the original one. When the conditions that permit $\underline{r} = 0$ are not met (e.g., (D) is satisfied), the original model admits feint equilibria with a risky low-value demand. This is not the case here: with this particular technology of war it is not possible to induce the weak type to run a risk of war under conditions that make fighting more beneficial than the peace terms for the strong type. (The proof of this is a bit involved and is omitted.) I conjecture that this is an artifact of the particular functional form chosen for the technology of conflict. This is why Proposition 2 restricts attention to feint equilibria with a riskless low-value demand. Substantively, these equilibria are equivalent to the ones in the original model.

5 Discussion

Although the framing of the model might make it look like the feint mechanism applies only narrowly to situations where a player might derive a tactical fighting advantage, the substance of the claim is more general. At the most abstract level, the mechanism applies to any setting where an attempt to influence a player's behavior with a threat might trigger a counter-response that would diminish the effectiveness of executing the threat if the attempt fails. As described, this is a very generic phenomenon and it is somewhat surprising that the formal study of coercion has neglected it. If I were to venture a guess as to the reason, it would have to be that we have only recently begun to study the distribution of power as an endogenous variable rather than something fixed by observable capabilities. As a result, we have only recently become aware that some of our general conclusions depend on the assumption of a fixed distribution of power.[19]

It is possible to use this mechanism to study the puzzle of secret defensive alliances.[20] One prominent vein in the alliance literature explains them as valuable signalling and commitment de-

vices (Morrow 2000). A defensive alliance, by its very nature, is supposed to enhance state A's capability against state C by adding the capabilities of state B to A's. This should improve A's defensive posture against C, and deter C from attacking. Abstracting away from how credible B's commitment to A is, concluding such an alliance in secret cannot increase A's deterrent threat for the simple reason that C is unaware of B's promise to aid A in war. So what is the point of concluding such an alliance?

The feint mechanism offers one possible answer: since a defensive alliance increases A's strength, making it public would alert C that she would have to be better prepared if she wants to coerce A. This would impel C to increase her capabilities, either by arming or by searching for allies of her own. If C succeeds, the overall benefit of the alliance might actually decrease. Hence, A might take his chances with a secret alliance: although C is less likely to agree to terms beneficial to A, if war occurs A will fight with B's help against an opponent who did not have the opportunity to prepare.

5.1 The Disadvantages of Democracy in Crisis Bargaining

There is an ongoing debate about the advantages democracies enjoy over other political systems when it comes to crisis bargaining or war fighting. One especially prominent argument is that democracies are better able to signal the resolve of their leaders in crises, perhaps because of audience costs, the interaction of opposition and incumbent parties, or other institutional features. The (somewhat simplified) core of these arguments is that democracies constrain the leaders' ability to bluff because open public debate and reselection incentives force them to issue threats only when they are resolved to follow through on them. In other words, it might be much more difficult for a democratic leader to conceal his resolve than for an autocrat. This makes threats more credible, which is held to be a good thing.[21]

The problem with the exclusive focus on credibility is that it neglects the consequences a believ-

able threat might have for the threatener if the target fails to comply. The tactical game here shows one possible reaction a target might have to a threat that is more credible: she might start preparing for a fight. In other words, enhancing credibility might actually diminish capability. The trade-off between communicating one's resolve without provoking a countervailing response is a difficult one. As such, even if one grants the argument that democracies can communicate resolve better than non-democracies, it is not at all clear that this will enable them to obtain better peace terms or enjoy lower risks of war. In fact, the present model suggests that the opposite might well be true.

It is generally the case that military capabilities are much more readily observable than the will to use them. This means that a country with a well-trained and well-supplied army that it is unwilling to commit to a fight is "weaker" than an opponent whose objective capabilities are not as great but who is ready to use them all in that fight. This is why indicators based on observable capabilities might not be very good predictors of how a crisis will end: the driving force behind the outcome is the contest of will rather than of brute numerical strength.

Consider now a democracy whose leader cannot feign weakness because the interaction of domestic political groups reveals the political will to use the observable capabilities. In the context of our model, this leader will either make the high-value demand when he is resolved or the low-value demand when he is not. This means that a democratic leader is more likely to be forced into a separating equilibrium than a non-democratic leader who can conceal his resolve. As we have seen, in a separating equilibrium the weak type's peace terms are worse than the peace terms he can obtain in a feint equilibrium with a riskless low-value demand. And while the terms of the high-value demand are the same for the resolved type in both cases, the risk he has to run to obtain them is strictly greater in the separating equilibrium. In other words, an unresolved democratic leader will obtain worse peace terms than an unresolved nondemocratic leader, and a resolved democratic leader must run higher risks of war to obtain the same peace terms as a resolved nondemocratic leader. This

suggests that the openness of democracies might put them at a disadvantage in crisis bargaining precisely because it communicates resolve better.

5.2 Shows of Strength and the Fostering of False Optimism

One of the most prominent causal mechanisms that explains war as the result of bargaining failure due to asymmetric information is the *risk-return trade-off* (Fearon 1995, Powell 1999). The essence of the mechanism is a screening logic: a player who is uncertain about his opponent's expected payoff from war makes a demand which balances the risk of rejection should its terms prove unacceptable to the opponent with the extra gain from peace these terms represent should they prove acceptable. Although one can always ensure peace by making a demand that even the strongest type of opponent would accept, this strategy is generally suboptimal because it involves large concessions that might well be unnecessary if the opponent is actually weak. The optimal strategy trades the gain from making a demand that is slightly less favorable to the opponent against the slightly higher risk that such a demand entails. The risk of war therefore arises from not knowing precisely what kind of demand that opponent would find agreeable.

The mechanism that explains war in the present model is different even though the basic ultimatum game is the same. In contrast to the traditional screening setting in which player 1 is uncertain about player 2's expected payoff from war, our crisis is a signaling setting in which it is player 1 who has private information. In fact, we did not need to assume any sort of uncertainty about player 2's type at all. The interaction is dominated by the informed player's attempt to signal his type in a credible manner: when the strong type succeeds in separating from the weak type, player 2 becomes much more amenable to concessions. The risk of war is a necessary feature of a separating strategy that aims to achieve credible communication.

The feint equilibria exhibit this costly signaling dynamics common to crisis bargaining behavior.

The strong player 1 can only obtain the high-value demand $\overline{x}$ at the cost of a high risk of a costly war with a fully prepared strong player 2. This discourages the weak type from attempting to bluff with the same demand. Endogenizing the war contest does not alter the basic logic of costly signaling. The only way a strong player can obtain a better deal is by revealing credibly that he is strong, which requires him to engage in behavior that the weak type would not want to mimic.

The interesting new feature of the feint equilibria is that the strong type of player 1 might mimic the behavior of the weak instead. One reason for this comes from the incentives the strong player 1 has to keep private his information about his own strength in the event of war. In the exogenous specification of the distribution of power, a player's expected war payoff may depend on his opponent's private information but not on her *beliefs* about the information that he knows but she does not. This means that with exogenous war payoffs, it does not matter to the player whether he fights an adversary that is fully informed or one that is uncertain about his strength. There is no reason for the player to manipulate the belief with which his opponent would enter the war, only the belief she has when deciding what to do about his demands. In these cases, the strong player is better off whenever his opponent knows that he is strong.

With endogenous war payoffs, the player does care about the beliefs with which his opponent begins the war. The informed strong type's expected payoff under uncertainty is strictly better than his payoff when his opponent is fully informed. (As $q \to 1$ the payoff under uncertainty converges to the complete-information payoff but by Lemma 3, it is strictly decreasing in q.)

This gives the strong type a potent reason not to reveal his strength during the crisis itself. He may deliberately leave his opponent in a state of *false optimism* in order to exploit the advantages of surprise in case war breaks out. Unlike the usual scenario in which strong types always attempt to overcome the optimism of the opponent with costly or risky shows of strength, the feint equilibrium dynamic suggests that they may not be willing to do so even if such actions are potentially available

to them. This creates a serious problem for peaceful crisis resolution because mutual optimism is regularly blamed as a major cause of war.[22]

In the classic formulation of the mutual optimism argument, "war is usually the outcome of a diplomatic crisis which cannot be solved because both sides have conflicting estimates of their bargaining power."[23] One problem is overconfidence about the likely development of the war: its duration (short), outcome (victory), and costs (low). In the model with endogenous war effort, the expected outcome depends on how hard the actors fight. Their joint efforts determine the probability of victory, and their uncertainty about the behavior of the opponent induces uncertainty in these estimates.

The resulting expectations about the war may well be incompatible. In the skirmish equilibrium, the strong type expects to win with probability $\pi_1(q;\overline{c}_1) = 1 - f(q)\sqrt{\underline{c}_1}/g(q;c_2)$, and player 2 expects to win with probability $\pi_2(q) = f(q)^2/g(q;c_2)$. These players are too optimistic because $\pi_1(q;\overline{c}_1) + \pi_2(q) > 1$. Similarly, in the war equilibrium the strong type expects to win with probability $\pi_1(q;\overline{c}_1) = \overline{c}_1/(\overline{c}_1 + qc_2)$, whereas player 2 expects to win with probability $\pi_2(q) = [(1-q)\overline{c}_1 + qc_2]/(\overline{c}_1 + qc_2)$. As in the skirmish equilibrium, these expectations are incompatible: it is easily verified that $\pi_1(q;\overline{c}_1) + \pi_2(q) > 1$. These optimistic expectations about victory translate into optimistic estimates about the expected payoffs from war.

It is crucial to understand that these disagreements are not about some fundamental underlying "true" probability of winning. Instead, they are disagreements about how war will "play out," and this, of course, depends to a large extent on the opponent's likely behavior. That behavior in turn depends on what the opponent expects the player to do, and these expectations are profoundly influenced by the opponent's belief about some aspect that is privately known by the player. This is where deliberate falsification enters the picture.

When mutual optimism is a possible cause of war, credible signaling might be some sort of

imperfect cure. When players have exaggeratedly optimistic expectations about their chances in war because they are not aware of private information the opponents posses, the only way to arrive at a peaceful settlement is to reduce this mutual optimism. As we know from our crisis bargaining studies, the only way to do so is through costly signaling. The cure is imperfect because the attempt to impart credibility to one's message forces the actor to behave in ways that increase the probability of war. Scholars are well aware of this paradox inherent in crisis bargaining, and it is perhaps best summarized by Schelling: "Flexing of muscles is probably unimpressive unless it is costly or risky. [. . .] Impressive demonstrations are probably the dangerous ones. We cannot have it both ways."[24]

The results here suggest that the difficulty with settling peacefully may go beyond the risk generated by signaling efforts. When unwarranted optimism arises from lack of information to which the opponent has access, it can be dispelled only when the opponent chooses to reveal it. Unfortunately, the logic of feigning weakness suggests that an actor may choose instead of obfuscate inferences in order to gain advantage in the war that follows. In other words, the actor *may deliberately foster false optimism* even though this may make it very unlikely that his opponent would concede enough to make that actor willing to forego fighting.[25]

Private information can remain private not for lack of means to reveal it but because the only type who can afford to send the credible signal may have no incentive to do so. It is this intentional and strategic concealment of information that is so troubling for resolving crises peacefully. To see how matters can come to a head, consider a crisis in which side A has deliberately fostered optimism in side B. Because side B (incorrectly) believes herself strong, she engages in very risky actions designed to cause side A to revise his war expectation downward. Unfortunately, side A cannot use side B's willingness to run risks as evidence that side B is strong, not when he misled B into believing that she is strong. In other words, when you have gone to great lengths to convince the opponent to be optimistic, you cannot very well use that optimism as evidence that your own

assessment is faulty. Side B's signaling behavior then will be more likely to cause war because A is essentially dismissing it, because B is unwilling to offer the necessary concessions, and because B's exaggerated optimism is prompting her to take very large risks. In this situation, mutually incompatible crisis expectations cannot be reconciled without the actual resort to arms. As Blainey puts it, "The start of war is... marked by conflicting expectations of what that war will be like. War itself then provides the stinging ice of reality."[26]

6 Conclusion

Consider the Chinese options in the fall of 1950. On one hand, they could openly threaten with intervention and demand that the U.N. forces remain south of the 38th parallel. If this works, the outcome is excellent. However, making this high demand is also very risky: if the U.S. happens to be resolved to unify Korea, this demand would simply alert it to prepare better for fighting the PRC. The resulting war would be of very high intensity and the Chinese would certainly lose the tactical advantage that would secure a first morale-boosting victory. On the other hand, the Chinese could demur and ask that only U.S. troops desist from crossing the parallel. Although permitting the occupation of North Korea by South Korean troops is not as good as keeping it free of U.N. forces, there is some chance that the U.S. would agree to this and war would be averted. Should the U.S. prove to be bent on unification, the absence of a credible signal can be expected to increase American confidence and possibly cause the U.S. to march into a war without the type of preparation it would have engaged in knowing the Chinese were going to intervene in strength. These are unpalatable choices, certainly, and no wonder Mao vacillated for so long before making up his mind on the strategy to pursue.

This stylized description of the situation seriously abstracts from the complex domestic dynamics

in both countries, and it may well have been the case that by the time Mao resolved to intervene, the United States had become undeterrable by the Chinese without open Soviet support. In November, war may have been already unavoidable (Slantchev N.d.). However, the logic of feigning weakness developed in this article can help explain why the Chinese did not pursue more vigorous signaling actions when they were resolved not to permit unification.

The crisis bargaining literature focuses on how strong actors can signal their strength and reduce the possibility of bluffing. When weak types can mimic their actions, messages will not be believed, and when threats are not credible, they are unlikely to influence the behavior of the opponent. This basic mechanism also obtains in the model presented here. This article, however, also points out some perverse incentives that strong types may face that may make them unwilling to send costly signals even when they could have done so.

One implication of this result is that it is not safe to infer that one's opponent is weak when he fails to engage in some costly action that is available to him and that could persuade one that he is strong. One should carefully consider the incentive to feign weakness for tactical purposes. This, of course, may be harder than it sounds because, after all, it could be the case that the opponent is not signaling because he really is weak.

The logic of the feint also suggests that overcoming mutual optimism in crises may be very difficult for two reasons. First, when a strong opponent who could reveal his strength to reduce an actor's optimism decides to feign weakness, then that actor may persist in her incorrect beliefs and blunder into disaster. Second, the possibilities for peaceful resolution of the crisis may diminish because the feigning opponent himself may be unable to correct his optimistic beliefs. Because he has purposefully misled the other actor, he cannot take her costly signals as evidence that he should revise his expectations: after all, she is signaling precisely because she believes that she is strong, which is the false belief he has taken great care to induce. In this rather unfortunate scenario, war

may be the only way to inject a dose of reality into these beliefs.

A PROOFS

Proof of Proposition 1. Since the strong type is mixing, $\underline{r}W_s^n + (1-\underline{r})\underline{x} = \overline{r}W_s^a + (1-\overline{r})\overline{x}$, which gives $\overline{r}$ in (3). Note that $\underline{x} < W_s^n$ yields:

$$q_a < \frac{s_n - w_n - C}{s_n - w_n}. \tag{12}$$

If (12) is not satisfied, $\overline{r} \in (0,1)$ regardless of the value of $\underline{r}$, so we can take $\underline{r} = 0$. Suppose now that (12) is satisfied. Then $\overline{r} < 1$ yields $\underline{r} > 1 - (s_n - s_a)/\left(W_s^n - \underline{x}\right) \equiv \underline{r}'$. Taking $\underline{r}' \geq 0$ yields (D), which ensures that the low-value demand must be risky, otherwise $\underline{r} = 0$ can work. Also, $\overline{r} > 0$ yields $\underline{r} < 1 - (s_n - s_a - k_2 - C)/\left(W_s^n - \underline{x}\right) \equiv \underline{r}''$. Since $\underline{r}'' - \underline{r}' > 0 \Leftrightarrow k_2 + C > 0$, such $\underline{r}$ exist.

Since the weak type should not have an incentive to demand $\overline{x}$, it follows that $\underline{r}W_w^n + (1-\underline{r})\underline{x} \geq \overline{r}W_w^a + (1-\overline{r})\overline{x}$, which simplifies to:

$$\overline{r} \geq \frac{\underline{r}\left(\underline{x} - W_w^n\right) + \overline{x} - \underline{x}}{\overline{x} - W_w^a}. \tag{13}$$

It is readily verifiable that $\overline{r} \in (0,1)$ regardless of $\underline{r}$. Since both (3) and (13) must hold, we require that:

$$\frac{\underline{r}\left(\underline{x} - W_s^n\right) + \overline{x} - \underline{x}}{\overline{x} - W_s^a} \geq \frac{\underline{r}\left(\underline{x} - W_w^n\right) + \overline{x} - \underline{x}}{\overline{x} - W_w^a}.$$

At $\underline{r} = 0$, the inequality reduces to $s_a \geq w_a$, which holds. Recall that if (D) is not satisfied, there are no lower-bound restrictions on $\underline{r}$ to guarantee valid $\overline{r}$ values, which means that in this case we may use $\underline{r} = 0$.

We now derive the range of low risks that can be supported in equilibrium when (D) is satisfied. The weak type should not deviate to the best possible riskless demand, x_1, so $\underline{r} \leq (\underline{x} -$

$x_1)/\left(\underline{x} - W_w^n\right) \equiv \hat{\underline{r}} < \underline{r}''$. $\underline{r}' \leq \hat{\underline{r}}$ reduces to (4).

Let $r(x)$ denote the player 2's rejection probability. Since $r(x) = 0$ for any $x \leq x_1$, if $\underline{r} = 0$, then player 1 cannot profit from deviating to a riskless x regardless of type. If $\underline{r} > 0$, the derivation ensures that the weak type cannot profit by deviating to $x_1 \geq x$, which also means that the strong cannot profit either. Since $r(x) = \underline{r}$ for any $x \in (x_1, \underline{x})$, such a demand only produces peace terms worse than $\underline{x}$, so a deviation cannot be profitable. Since $r(x) = \overline{r}$ for any $x \in (\underline{x}, \overline{x})$, such demands result peace terms worse than $\overline{x}$ and same risk of war against armed player 2. The strong type cannot profit from deviating and since our construction ensures that the weak cannot profit from $\overline{x}$, he will not deviate either. Since $r(x) = 1$ for any $x > \overline{x}$, the strong type cannot profit from deviations to certain war because $W_s^a < \overline{x}$. Neither can the weak type: $W_w^a < W_w^n < \underline{x}$. □

Proof of Lemma 2. Let $m_2^* \geq 0$ denote player 2's equilibrium effort, and $m_1^*(c_1) = m_1(m_2^*; c_1)$ player 1's effort. There can be no equilibrium in which player 1 makes no effort regardless of type. Suppose, to the contrary, that $m_1^*(\overline{c}_1) = m_1^*(\underline{c}_1) = 0$ in some equilibrium. Since $m_1(c_1) > 0$ whenever $c_1 > m_2^*$, this implies that $m_2^* \geq \overline{c}_1 > 0$. This cannot be optimal because she can deviate to a lower effort and still win for sure. Therefore, in any equilibrium at least one type of player 1 must be exerting a strictly positive effort. This cannot be the weak type by himself. Suppose, to the contrary, that $m_1^*(\underline{c}_1) > 0$ and $m_1^*(\overline{c}_1) = 0$ in some equilibrium. Since $m_1^*(\underline{c}_1) > 0$ implies that $m_2^* < \underline{c}_1$, it follows from $\underline{c}_1 < \overline{c}_1$ that $m_2^* < \overline{c}_1$, and so $m_1^*(\overline{c}_1) > 0$ as well, a contradiction. □

Proof of Lemma 3. Note that m_2^* is increasing in q when Assumption 2 is satisfied: in the skirmish equilibrium, $\text{sign}\,\frac{\partial m_2^*}{\partial q} = \text{sign}\left(c_2 - \sqrt{\underline{c}_1 \overline{c}_1}\right) > 0$, where the inequality follows from Assumption 2; in the war equilibrium, $\frac{\partial m_2^*}{\partial q} = \frac{2q\overline{c}_1^2 c_2^2}{(\overline{c}_1 + qc_2)^3} > 0$. Turning now to the claim of the lemma, observe that in the skirmish equilibrium, $\frac{\partial W_1(c_1)}{\partial q} = -\left(\frac{\sqrt{c_1} - \sqrt{m_2^*}}{c_1\sqrt{m_2^*}}\right)\frac{\partial m_2^*}{\partial q} < 0$. because the bracketed term is positive by (8) and because m_2^* is increasing in q. Since only the strong type participates

in the war equilibrium, inspection of his payoff is sufficient to establish the claim. □

LEMMA 6. *$W_2(q)$ is continuous and strictly decreasing.*

Proof. (Continuity.) Since $W_2(q)$ is continuous for each equilibrium, it is enough to show that it is continuous at q_s where the equilibrium switch occurs: $W_2^s(q_s) = 1 - \frac{\underline{c}_1 + \sqrt{\underline{c}_1 \overline{c}_1}}{c_2} = W_2^w(q_s)$.

(Monotonicity.) In the war equilibrium, $\frac{dW_2^w(q)}{dq} = -\frac{\overline{c}_1^2(\overline{c}_1 + 3qc_2)}{(\overline{c}_1 + qc_2)^3} < 0$. In the skirmish equilibrium, $\frac{dW_2^s(q)}{dq} = \frac{g'f^2}{g^2} + \frac{2f(q\underline{c}_1 + (1-q)\overline{c}_1)}{g^3}\Big(f'g - g'f\Big) < 0$. To see this, note that $f > 0$, $g > 0$, $f' < 0$, $g' < 0$, and $f'g - g'f = \sqrt{\underline{c}_1 \overline{c}_1}\left(\sqrt{\overline{c}_1} - \sqrt{\underline{c}_1}\right)\left(1 - \frac{\sqrt{\underline{c}_1 \overline{c}_1}}{c_2}\right) > 0$ by Assumption 2. The last requirement is that $gg'f + 2(q\underline{c}_1 + (1-q)\overline{c}_1)(f'g - g'f) < 0$, which can be shown but it takes three pages of algebra. □

Proof of Lemma 4. Lemma 6 implies that to get the best and worst payoffs for player 2, we only need to consider $q = 0$ and $q = 1$. So, $\lim_{q\to 0} W_2^s(q) = \left(\frac{c_2}{\underline{c}_1 + c_2}\right)^2 = \overline{W}_2$, and $\lim_{q\to 1} W_2^s(q) = \lim_{q\to 1} W_2^w(q) = \left(\frac{c_2}{\overline{c}_1 + c_2}\right)^2 = \underline{W}_2$, with $\underline{W}_2 < \overline{W}_2$. □

Proof of Lemma 5. By Lemma 6, $W_2(q)$ is continuous, and the intermediate value theorem implies that for any $y \in [\underline{W}_2, \overline{W}_2]$, there exists q such that $W_2(q) = y$. By Lemma 6, $W_2(q)$ is strictly decreasing, so $q(y) = W_2^{-1}(y)$ is unique and strictly decreasing in y. Letting $x = 1 - y$ establishes the claim. □

Proof of Proposition 2. Let $\underline{W}_1 = W_1(1; \overline{c}_1) = W_1(\overline{c}_1, c_2)$. Since the strong player 1 is willing to mix, $U_1(\underline{x}; \overline{c}_1) = U_1(\overline{x}; \overline{c}_1)$, or

$$\underline{r} W_1(q(\underline{x}); \overline{c}_1) + (1 - \underline{r})\underline{x} = \overline{r}\,\underline{W}_1 + (1 - \overline{r})\overline{x}. \tag{14}$$

Using the definitions of $\underline{x}$ and $\overline{x}$, we can rewrite this as:

$$\overline{r} = \frac{\underline{r}\,[1 - W_1(q(\underline{x}); \overline{c}_1) - W_2(q(\underline{x}))] - \underline{W}_2 + W_2(q(\underline{x}))}{1 - \underline{W}_1 - \underline{W}_2} \tag{15}$$

Since $\overline{r} > 0$ and $1 > \underline{W}_1 + \underline{W}_2$, it follows that:

$$\underline{r}\,[1 - W_1(q(\underline{x}); \overline{c}_1) - W_2(q(\underline{x}))] > \underline{W}_2 - W_2(q(\underline{x}))$$

must hold. Since $W_2(q(x)) > \underline{W}_2$ for any $q(x) < 1$, the right-hand side is negative, so $\underline{r} = 0$ certainly satisfies this condition. Since $\overline{r} < 1$ as well,

$$\underline{r}\,[1 - W_1(q(\underline{x}); \overline{c}_1) - W_2(q(\underline{x}))] < 1 - \underline{W}_1 - W_2(q(\underline{x})) \tag{16}$$

must hold. There are two cases to consider. First, suppose

$$1 - \underline{W}_1 - W_2(q(\underline{x})) \geq 0. \tag{Z}$$

Since $W_1(q(x); \overline{c}) > \underline{W}_1$ for any $q(x) < 1$, it follows that (16) is satisfied for any $\underline{r}$, so for $\underline{r} = 0$ in particular. Now suppose that (Z) is not satisfied. Both sides of (16) are negative, which implies that only $\underline{r} > 0$ can possibly satisfy it. Thus, if (Z) is not satisfied, the low-value demand cannot be riskless. Because we are looking for equilibria with such a demand, assume that (Z) holds for the rest of the proof. I labeled this condition to indicate the zero-risk associated with the low-value demand, and it is the analogue to the converse of (D) in the simple models.

Consider now the rejection probability specified in the proposition. For any $x \in (\underline{x}, \overline{x}]$, $r(x) = (x - \underline{x})/\,[x - W_1(q(x); \overline{c}_1)]$ solves $U_1(x; \overline{c}_1) = \underline{x}$. That is, player 2's rejection probability leaves the strong type indifferent between any demand in that range and the equilibrium *riskless* low-value demand. Note in particular that $x > \underline{x} \geq \widetilde{x}$ implies that $r(x)$ is a valid probability. Moreover, $r(\overline{x}) = \overline{r}$ from (15) because $\underline{r} = 0$ and $q(\overline{x}) = 1$. Since $\underline{x} \geq x_1$, taking $x^* = \max(\widetilde{x}, x_1)$ yields the lower bound on the riskless demand that can be supported in equilibrium. The upper bound x^{**} follows from Bayes rule and is derived in the text.

We now check that deviations are unprofitable. Since player 2 accepts any $x < \underline{x}$, such deviation from $\underline{x}$ is not profitable. Any $x \in (\underline{x}; \overline{x}]$ is rejected with probability that leaves the strong type

indifferent between x and $\underline{x}$. But since $\underline{x}$ is also the weak type's payoff and the weak type's payoff from war is strictly worse than the strong type's, this deviation is strictly worse for the weak type. Finally, any $x > \overline{x}$ is rejected for sure, and the resulting war is one in which player believes she is facing the strong type. This is clearly worse for the strong type (at $\overline{x}$ he fights such a war with positive probability but also obtains $\overline{x} > \underline{W}_1$ with positive probability), and this implies it is also worse for the weak type. □

NOTES

[1] Appleman (1961, 763,768), Whiting (1960, 122).

[2] Appleman (1961, 65).

[3] Schelling (1966, 55, fn. 11).

[4] The debate about the causes of U.S. failure to understand the seriousness of Chinese threats is quite intense. The literature on the subject is intricate and it is well beyond the scope of this article to delve in details on that issue. Many studies assert that the Chinese threat *was* credible but that the U.S. administration mistakenly dismissed it (Lebow 1981). The opposite assertion is that the Chinese were spoiling for a fight (Chen 1994, 40). Slantchev (N.d.) counters both in detail.

[5] Results similar in spirit can be obtained in other settings such as jump-bidding in auctions (Hörner and Sahuguet 2007), and repeated contests (Münster 2007).

[6] Fearon (1994); Schultz (1998); Sartori (2005); Guisinger and Smith (2002); Schelling (1966).

[7] Banks (1990).

[8] Powell (1996); Wagner (2000).

[9] Powell (1993).

[10] Stueck (2002, 89).

[11] The vulnerability to aerial attacks and inferiority of equipment and (supposedly) morale led MacArthur to assure President Truman at the Wake Island Conference that should the Chinese attempt to intervene, "there would be the greatest slaughter" (United States Department of State 1976, 953).

[12] Powell (1993); Slantchev (2005).

[13] For ease of exposition, I will refer to player 1 as "he" and player 2 as "she."

[14] The result can be immediately obtained by replacing (IC_s) with a weak inequality such that the high-value demand is weakly preferable for the strong type. If $\underline{x} \geq \overline{x}$, the payoff from demanding $\underline{x}$ will always be strictly greater than the payoff from $\overline{x}$, which means that the strong type would not want to demand $\overline{x}$, a contradiction.

[15] This one is the classic contest success function from economics (Hirshleifer 1989). In the economics literature, surveyed by Garfinkel and Skaperdas (2007), the interest in the rent dissipation and the inability to create a contract that would avoid it, not so much in the signaling properties of arming or taking advantage of informational asymmetries.

[16] In a previous version of this article, I derived the results for the two-sided incomplete information case. Aside from making the algebra more involved, the analysis adds nothing of significance.

[17] Fearon (1995).

[18] When it is necessary to be explicit about which equilibrium I am referring to, I shall use $W_2^s(q(x))$ for the skirmish equilibrium, and $W_2^w(q(x))$ for the war equilibrium.

[19] Even non-formal studies that highlight the importance of resolve asymmetry and the desirability of being non-provocative tend to treat the distribution of power as fixed (George and Simons 1994).

[20] I thank Jeff Ritter for suggesting this. See his dissertation for an extended study of secret alliances (Ritter 2004).

[21] Fearon (1994); Schultz (2001); Bueno de Mesquita, Morrow, Siverson and Smith (2003). Slantchev (2006) provides a dissenting view on the audience cost mechanism.

[22] Blainey (1988, 53). Wittman (1979) offers the first rationalist account. Fey and Ramsay (2007) attempt to show that the mutual optimism explanation cannot be sustained as a result of rational behavior. Slantchev and Tarar (2007) counter their argument.

[23] Blainey (1988, 114).

[24] Schelling (1966, 238–39). See also Fearon (1995, 397); Schultz (1998, 829).

[25] Misleading the opponent is not the only reason a strong type might not wish to separate himself from the weak type. Kurizaki (2007) analyzes a model in which player 1 can decide whether to make his threat public (so whoever backs down incurs audience costs) or keep it private (so backing down is costless). In the private threat equilibrium, the strong player 1 is indifferent between going public and staying private, whereas the weak type always threatens in private. The strong type is indifferent because he always fights when resisted and player 2 resists with the same probability after private and public threats. She does so because capitulation is costlier after a public threat, in which case she needs to be fairly certain her opponent is strong. In private, the costs of capitulation are much lower, so she can concede even if she thinks player

1 might be bluffing. There is no benefit to the strong player 1 in getting player 2 to think that he is weak.

[26] Blainey (1988, 56).

REFERENCES

Appleman, Roy E. 1961. *South to the Naktong, North to the Yalu*. Washington, DC: U.S. Government Printing Office.

Banks, Jeffrey S. 1990. "Equilibrium Behavior in Crisis Bargaining Games." *American Journal of Political Science* 34 (August): 599–614.

Blainey, Geoffrey. 1988. *The Causes of War*. 3rd ed. New York: The Free Press.

Bueno de Mesquita, Bruce, James Morrow, Randolph Siverson, and Alastair Smith. 2003. *The Logic of Political Survival*. Cambridge: The M.I.T. Press.

Chen, Jian. 1994. *China's Road to the Korean War: The Making of the Sino-American Confrontation*. New York: Columbia University Press.

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

Fearon, James D. 1995. "Rationalist Explanations for War." *International Organization* 49 (Summer): 379–414.

Fey, Mark, and Kristopher W. Ramsay. 2007. "Mutual Optimism and War." *American Journal of Political Science* 51 (October): 738–754.

Garfinkel, Michelle R., and Stergios Skaperdas. 2007. Economics of Conflict: An Overview. In *Handbook of Defense Economics, Volume 2: Defense in a Globalized World*, ed. Todd Sandler, and Keith Hartley. Amsterdam: North Holland pp. 649–710.

George, Alexander L., and William E. Simons, eds. 1994. *The Limits of Coercive Diplomacy*. 2nd ed. Boulder: Westview Press.

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

Hirshleifer, Jack. 1989. "Conflict and rent-seeking success functions: Ratio vs. difference models of relative success." *Public Choice* 63: 101–112.

Hörner, Johannes, and Nicholas Sahuguet. 2007. "Costly Signalling in Auctions." *The Review of Economic Studies* 74 (1): 173–206.

Kurizaki, Shuhei. 2007. "Efficient Secrecy: Public versus Private Threats in Crisis Diplomacy." *American Political Science Review* 101 (August): 543–558.

Lebow, Richard Ned. 1981. *Between Peace and War: The Nature of International Crisis*. Baltimore: The Johns Hopkins University Press.

Morrow, James D. 2000. "Alliances: Why Write Them Down?" *Annuual Reviews of Political Science* 3: 63–83.

Münster, Johannes. 2007. "Repeated Contests with Asymmetric Information." Manuscript, Free University, Berlin.

Powell, Robert. 1993. "Guns, Butter, and Anarchy." *American Political Science Review* 87 (March): 115–32.

Powell, Robert. 1996. "Stability and the Distribution of Power." *World Politics* 48 (January): 239–67.

Powell, Robert. 1999. *In the Shadow of Power*. Princeton: Princeton University Press.

Ritter, Jeffrey M. 2004. *Silent Partners and Other Essays on Alliance Politics*. Cambridge: Harvard University. Doctoral dissertation.

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

Schelling, Thomas C. 1966. *Arms and Influence*. New Haven: Yale University Press.

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

Schultz, Kenneth A. 2001. *Democracy and Coercive Diplomacy*. Cambridge: Cambridge University Press.

Slantchev, Branislav L. 2005. "Military Coercion in Interstate Crises." *American Political Science Review* 99 (November): 533–547.

Slantchev, Branislav L. 2006. "Politicians, the Media, and Domestic Audience Costs." *International Studies Quarterly* 50 (June): 445–477.

Slantchev, Branislav L. N.d. *Military Threats: The Costs of Coercion and the Price of Peace*. Cambridge: Cambridge University Press.

Slantchev, Branislav L., and Ahmer Tarar. 2007. "War Is Not a Bet: Mutual Optimism as a Cause of War." Manuscript, Department of Political Science, University of California, San Diego.

Stueck, William. 2002. *Rethinking the Korean War*. Princeton: Princeton University Press.

Sun Tzu. 2005. *The Art of War*. El Paso Norte Press. Tr. Lionel Giles.

United States Department of State. 1976. *Foreign Relations of the United States, 1950. Volume VII: Korea*. Washington, DC: Government Printing Office.

Wagner, R. Harrison. 2000. “Bargaining and War.” *American Journal of Political Science* 44 (July): 469–84.

Whiting, Allen S. 1960. *China Crosses the Yalu: The Decision to Enter the Korean War*. New York: The Macmillan Company.

Wittman, Donald. 1979. “How a War Ends: A Rational Model Approach.” *The Journal of Conflict Resolution* 23 (December): 743–63.