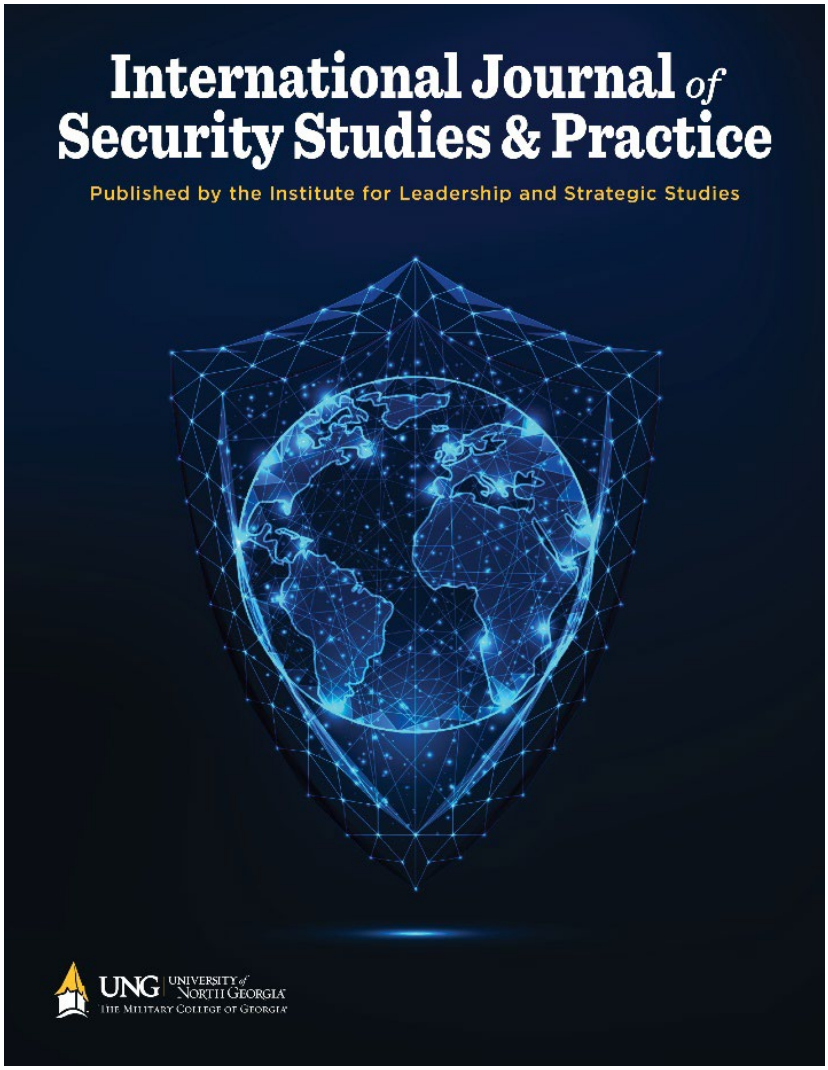
International Journal of Security Studies & Practice

Volume 3 | Issue 1 Article 6

2023

How States Provoke Adversaries to Solicit Information



Dmitriy Nurullayev, Ph.D.
University of Arizona
nurullayev@arizona.edu
ORCID: 0000-0003-2712-724X

Sydney Kerr
University of California, Riverside
ORCID: 0009-0008-53116062

Mathew Knudson, Ph.D.
Vanderbilt University

Recommended Citation

Nurullayev, D., Kerr, S., Knudson, M. (2023). How states provoke adversaries to solicit information. *International Journal of Security Studies & Practice*, 3(1), Article 6. http://ijssp.ung.edu

How States Provoke Adversaries to Solicit Information

Dmitriy Nurullayev
College of Applied Science & Technology
University of Arizona
Davis Center
Harvard University
ORCID: 0000-0003-2712-724X

Sydney Kerr
Department of Political Science
University of California, Riverside
ORCID: 0009-0008-53116062

Mathew Knudson
Department of Economics
Vanderbilt University

The authors declare that there is no potential conflict of interest related to this article.

Correspondence concerning this article should be addressed to Dmitriy Nurullayev, e-mail: nurullayev@arizona.edu

Abstract

In the context of conflict maneuvering, can state actors employ strategies of provocation designed to extract private information regarding the resolve of their adversaries? Existing literature has examined various ways in which state actors seek to increase their payoffs. Significant research, including scholarship on costly signaling, has studied how actors attempt to influence their opponents' perceptions. However, the current literature lacks an explanation of actor strategies aimed at extracting information rather than communicating or manipulating it. This work aims to fill this gap by considering the strategy of provocative probes. We argue that states can gain an informational advantage by provoking their opponents to solicit information, and we present a new bargaining model to demonstrate the theoretical soundness of this strategy.

How States Provoke Adversaries to Solicit Information

How states behave on the international stage, especially in the context of conflict bargaining, depends not only on their strength, motivation, and resolve but also on the strength and resolve of their adversaries and allies (Snidal, 1985). Mistrust is commonplace in the anarchic nature of international politics, and actors must carefully evaluate the stated resolve of their peers (Waltz, 2001). Due to the potential benefits of manipulating their adversaries through bluffing, promises of friendship, and threats of potential aggression, they are treated with equal skepticism. Making accurate interpretations and avoiding manipulation can lead to great rewards, while misunderstanding genuine signals of resolve or falling for well-crafted bluffs can result in significant losses or missed opportunities (Jervis, 2017).

Significant scholarly work has been done on the role of resolve and signaling in international relations (Schelling, 1980; Fearon, 1997; Kertzer et al., 2020). Information and misinformation are important variables in the study of interstate interactions (Jervis, 2017). As Thomas Schelling noted, there would be no conflicts in the world of perfect information. If all threats were credible and all information regarding capability was accurate, states would have no reason to fight (Schelling, 2008). Most of this existing work, however, is explored from the perspective of the signal sender (Schelling, 1980; Fearon, 1997). Far less is published on the strategic choices of actors who must receive and interpret the transmitted information. This leads to our main question: given the discussed incentives to misinform, can state actors induce their adversaries to disclose accurate information regarding resolve?[1]

[1] We define resolve as the willingness and capability to pursue a policy or decision "despite contrary inclinations or temptations to back down" (Kertzer, 2017). While the adversaries or even allies may not be certain of one's resolve, we assume that the actor is always aware of her own resolve. She may not communicate honestly (hide/bluff) or she may fail to communicate well despite her desire to send a genuine signal, but an actor cannot be unaware of her own level of commitment to an issue.

Suppose there was a tool with which a state could induce a genuine signal from its opponent. Envisage the following example: Actor A is interested in discovering the consequences of taking away territory X from Actor B. As they negotiate, A could ask B what B's resolve might be should negotiations fall through and fighting become necessary. However, this would be a subpar strategy since talk is cheap, and B does not have to possess a strong commitment or resources to respond with strong language. One possible solution in this scenario is for A to conduct a limited test or a probe. Instead of laying claim to the whole territory X, A could attempt to gain possession of just one small slice of it and observe the subsequent response issued by B. In theory, B should respond proportionally. In other words, B would dedicate, relative to the size of the slice, the same percentage of resources and resolve as it would have devoted to contesting the entire territory. For B, faking a strong response would be significantly more difficult, and by observing B's response, A can generate a better understanding of B's resolve. This refined understanding would lead to a potentially better bargaining strategy, allowing A to make a bold subsequent move or to avoid a devastating war. Note that "territory" is used to denote any strategic state interest, and actual examples can span a wide array of issues from nuclear proliferation to the creation of alliances.

In the well-known literary work "The Art of War," ancient Chinese philosopher and military general Sun Tzu (2014) writes: "So in war, the way is to avoid what is strong and to strike at what is weak." Guided by this principle, how does a state actor discriminate between times that call for an attack and times that necessitate a retreat? Understanding the relative power distribution between oneself and the adversary is imperative in answering this question. It does not strain credulity to assume that one is generally aware of their own strength and resolve, but how does one judge the relative power and resolve of an adversary? Since potential negotiated settlements

are directly dependent on parties' perceptions of mutual power distribution and resolve, states can benefit from obtaining accurate information on their opponents' capabilities and objectives. Information on state capabilities is often less private than information on resolve. Much has been said about signaling and communication of resolve that is done on a voluntary basis, yet there is a gap in our understanding of how states may solicit these signals. Our model aims to fill this void.

Generally, state actors tend to reach a negotiated settlement instead of resorting to armed conflict when there is an overlap between the perceived bargaining ranges of the adversaries (Powell, 2002). Therefore, ignorance is never a virtue, and states are always better off having complete and accurate information. Much academic work has been dedicated to examining states' efforts to manipulate the perceptions of their adversaries (Schelling, 2008 & 1980; Fearon, 1994). These manipulation efforts can take the form of genuine communication or deception. However, little research has been conducted to explore state actions aimed at extracting information, and we aim to fill this gap by exploring the strategy of provocation.[2]

This paper delves into the inherent logic behind the strategy of soliciting information through provocation. Following this introduction, Section 2 outlines the theoretical argument of probes and provides an illustrative example based on Beijing's military maneuvers during the Taiwan Strait Crisis of 1954. The formal model is presented in Section 3, while Section 4 explores the lessons that can be learned and presents implications for future research. The mathematical proofs for the model can be found in the Appendix.

[2] We are not the first to coin the concept of the probe strategy in international conflict. In their book "Deterrence in American Foreign Policy: Theory and Practice," George and Smoke (1974) explore the military actions of Beijing during the Taiwan Strait Crisis and refer to these actions as 'the probe strategy.' However, to our knowledge, this is the first attempt to demonstrate the rational soundness of the strategy.

Art of Provocation

The following is a brief theoretical argument of how states can benefit from executing probes and how they may go about engaging in such strategies. We start with the end goal—actor preferences and then go through the logical exercises that actors may follow to reach their preferred outcomes. Following the theoretical argument, we present a brief illustrative example based on a historical case.

Theoretical Planks

Since the cost of engaging in war is always positive, states seek to avoid conflicts through the bargaining process (Lake, 2010). This process of reaching a negotiated settlement is inherently complex because sides are not typically satisfied with just any outcome within their bargaining range. When deciding between war and acceptable negotiated settlements, state actors pursue the most advantageous option that the adversary will accept. At the same time, states recognize that demanding more than what the adversary is willing to accept risks pushing the adversary towards choosing war (Filson & Werner, 2002; Fearon, 1998). In simple terms, state actors seek negotiated settlements within their bargaining range that encroach as far as possible into the adversaries' interests without initiating war.

Now that we have established actor preferences, what strategies and maneuvers will actors employ to achieve these desired outcomes? In the context of bargaining, being misinformed is not advantageous, and state actors desire accurate information regarding their adversaries' capabilities and resolve. Obtaining information on resolve is particularly challenging since military budgets and specific capability developments are often publicly accessible, while the level of resolve remains elusive. As states can benefit from possessing accurate information on their adversaries'

resolve, it creates an incentive for targeted states to conceal this information or deliberately provide misinformation to their inquisitive opponents (Jervis, 2017).

Existing scholarship has extensively studied the strategy of costly signals as a means for state actors to effectively convey their resolve on specific issues (Gartzke et al., 2017; Fearon, 1994). In other words, actors who possess high levels of resolve but wish to avoid armed conflict often face the challenge of convincing their adversaries that their resolve is genuine. However, actors also have an incentive to bluff and manipulate their adversaries' perceptions. Due to the anarchic nature of the international system, signaling resolve becomes more of an art than a science (Schelling, 2008). Once signals of resolve are sent, the receiver must assess the credibility of these signals, which can be inconsistent and difficult to interpret.

The conduct of international politics is often complex and messy. While states can use signaling to communicate their resolve on specific issues, the consistency and credibility of these signals can vary. Priorities may shift over time, and states are not always transparent about the changing nature of their priorities. For instance, a new administration may perceive a particular issue as less significant than its predecessor did but may not want to acknowledge a decrease in their resolve level. To maintain appearances, they may continue using rhetoric that appears resolute without actually possessing the dedication to follow through.

Certain domestic and international events, such as elections, political uprisings, regime change, and alliance formation or collapse, can create new incentives to reassess power dynamics and renegotiate previously settled matters. For instance, even if the challenger is initially satisfied with the status quo, significant political changes, such as the election of a new leader in the target's system, may prompt the challenger to reevaluate the target's level of resolve on a particular issue. This dynamic is exemplified in the context of U.S.-North Korean relations.

Another source of uncertainty arises when an issue becomes highly salient for the state initiating the probe. Let's assume that states A and B have a disagreement over issue X. Initially, issue X is of low importance to state A, and given B's signals of strong resolve towards issue X, A is content with maintaining the status quo. However, if state A experiences a renewed understanding of the significance of issue X, whether due to a shift in strategic priorities or a change in administrations, for example, A may become less satisfied with the status quo and view B's existing signals of resolve as insufficiently convincing. In such circumstances, A would be motivated to verify B's level of resolve.

Given this discussion, there are several ways in which a challenger can gain better information on the resolve of a target. The challenger can pay attention to the rhetoric used by the target, request clarification on their commitment, or analyze how the changing circumstances may have influenced the target's resolve. However, these methods have their limitations as the target can manipulate events or provide disingenuous responses to gain a better bargaining position. In this context, actions speak louder than words. There is a clear distinction between merely threatening to play Russian roulette and actually pointing a pistol at oneself and pulling the trigger to demonstrate the resolve to follow through.

To address these challenges, we propose that states can engage in provocative acts, referred to as "probes," with the aim of eliciting valuable information about their opponents' resolve. These probes are designed to prompt a response from the target. Even if the target chooses not to respond, this silence itself conveys meaningful information. When a well-designed probe is executed, the target's response can be strong, weak, or non-existent. Since a response to a probe must be in the form of action, the possibility of bluffing a "strong response" is significantly reduced. Once the response is issued, the challenger can utilize this information to update their beliefs regarding the

target's overall resolve. It is important to note that we adopt rationalist assumptions and analyze states as unitary actors. Based on this updated information, the challenger can exercise greater caution or assertiveness in their future interactions with the target.

Illustrative Example: The Taiwan Strait Crisis

The Taiwan Strait Crisis of 1954 demonstrates the logic of this strategy. In September of that year, the armed forces of the People Republic of China (PRC) began shelling Republic of China (ROC) positions in Quemoy, and a few days later, PRC planes began daily aerial bombardments of the Tachen Islands (Huei, 2019). This prompted the U.S. Joint Chiefs of Staff to recommend to the U.S. President that he authorize tactical nuclear strikes on mainland China to deter their further aggression. In December of 1954, the U.S. signed a Mutual Defense Treaty with Chiang Kai Shek putting Taiwan under U.S. protection. This, however, did not include the outlying islands which had already been under attack.

In his memoir titled *Inside the Story of the Decision Making during the Shelling of Jinmen*, Wu Lengxi, a member of the Chinese Communist Party Central Committee, paints a compelling description of factors which motivated PRC's actions. Lengxi explains that PRC was motivated to attack Taiwan but was unsure how the United States would respond should such an attack occur (Lengxi, 1994). Lengxi cites personal conversations with Chairman Mao and recalls the logic behind Mao's strategy of attacking the smaller islands in the strait:

> Chairman Mao said that the bombardment of Jinmen, frankly speaking, was our turn to create international tension for a purpose. For the present, we should first wait and see what international responses, especially American responses, there were to our shelling, and then we could decide on our next move. (p. 3)

Lengxi notes that Mao wished to better understand the level of resolve the United States possessed to protect Taiwan and he caveats that the PRC was not willing to attack at all costs. In other words, the response which was to be issued by the Americans was of great importance to Mao's subsequent decision-making.

> According to their responses during the past days, Mao said that Americans were worried not only by our possible landing at Jinmen and Matsu, but also our preparation to liberate Taiwan. In fact, our bombardment of Jinmen was a probe. We did not say if we were or were not going to land. We were acting as circumstances dictated. (p. 3).

Lengxi further explains that the United States' prior illustrations of commitment did not make it clear to the PRC what level of resolve the U.S. possessed to protect various islands which lay between the mainland and Taiwan. He writes:

> The treaty [between the United States and Taiwan], however, did not clearly indicate whether the U.S. defense perimeter included Jinmen and Mazu. Thus, we needed to see if the Americans wanted to carry these two burdens on their backs... Our bombardment was a test... Our talks were a test through diplomacy in order to get to the bottom of the American reaction. (p. 4).

Formal Model

In this section, we consider a formal model of two states that may engage in an exchange of information about their resolve. Specifically, we consider an environment in which a challenger is uncertain about the resolve of the defender and may attempt to resolve this uncertainty by engaging in a small-scale conflict or a probe. The defender's ability to respond is correlated to its resolve, and consequently, these small-scale conflicts can be informative about the likely outcome of a full-scale war.

Unlike many extant signaling models of war, we show that an equilibrium exists in which states deliberately misrepresent their type to gain at the bargaining table. This equilibrium exists despite the awareness of the challenger that defenders sometimes misrepresent, rationality of both parties, and the use of Bayes' rule to update beliefs. It is driven by a simple mechanism; low resolve states mimic the behavior of high resolve states. While the challenger is aware that this deception is going on, the costs of accidentally making aggressive demands from a high resolve defender are high enough to make the challenger take these signals seriously. We further show that this result is robust to the possibility of exogenous distrust of the signals which may lead misrepresentations to backfire.

Setup

We consider a challenger (C) and a defender (D) who are contesting ownership of a piece of territory, worth $2V \in R+$. For simplicity, we assume the probability of victory if the challenger and the defender fight over the territory is 1/2 for each.[3] Suppose that the challenger does not know the resolve of the defender.[4] With a probability $b_0 \in (0, 1)$, the defender has high resolve (HR), meaning that the defender incurs a low cost (C_{HR}^D) of fighting for the territory, and the challenger will incur a high cost of fighting for the territory, (C_{HR}^C) . Conversely, with probability $1 - b_0$, the defender has low resolve (LR), so that his costs are high (C_{LR}^D) for fighting while the challenger will incur a low cost of fighting (C_{LR}^C). We assume that $(C_{HR}^D) < V < (C_{HR}^C)$ and that $(C_{LR}^D) < V < (C_{LR}^C)$. An immediate implication is that highly resolved defenders are willing to fight for the territory, and low resolved defenders are not. The challenger is only willing to fight a

[3] Territory is used as an abstract concept here. Naturally, territory can be substituted with other tangible or even intangible goals that states pursue.

[4] Resolve can be substituted with capability or the combination of resolve and capability.

low resolve defender.

Rather than a standard bargaining model of war, we assume that the challenger has only two options: she can either back down (BD), and let the defender keep the territory, or double down (DD), and demand the territory from the defender. Because low resolve defenders are not willing to fight, doubling down on a low resolve defender results in the surrender of the territory to the challenger, while doubling down on a high resolve defender results in a war.

Probes and Responses

Prior to the decision to double down or back down, the challenger may pay a cost of C_p to probe the defender in order to learn more about the type of the defender. The defender can either send a strong response or a weak response. A strong response uses a large amount of military resources, while a weak response uses only a few.[5] The costs of responding to the probe are naturally correlated with the costs of responding to an invasion of the territory. A highly resolved state will generally have low costs of responding strongly to the probe, while a low resolve state will generally have high costs of responding strongly to the probe. However, the small-scale nature of the probe implies it may be possible for some deception to occur. For example, a state with high resolve may choose to move resources out of the probed area to appear to have low resolve, while a state with low resources may concentrate them in the probed area to give the appearance of resolve. Choosing to use more resources means bearing a greater cost, as those resources may be destroyed or diverted from more valuable duties, but this cost may vary. For example, the cost of diverting some resources could be abnormally low at some times because there is nothing else for them to be doing, or because they happen to be close by already, or because the terrain allows a

[5] As with previous simplifications, the use of military resources can be substituted with a wide range of alternative resources depending on the context of the probe.

clever defensive strategy. The costs could sometimes be abnormally high at other times because there are other important duties for the defender to attend to, or they are far away, or the terrain is difficult to defend.

Formally, high resolve defenders have a low cost of responding strongly to the probe on average, while low resolve defenders have a high cost of responding strongly to the probe. However, this is only true on average; sometimes costs are low even for a low resolve defender, or high even for a high resolve defender. For simplicity, we assume that the costs of a strong response may be either c or c + Δ, where Δ is so large that it is prohibitively expensive for a defender to send a strong response. This is equivalent to assuming it is simply not possible to respond strongly, which could be the case if it is simply impossible to move troops to the relevant area in time for it to matter. For high resolve defenders, low costs occur with a probability of π_{HR}, and prohibitively high costs occur with a probability of $(1 - \pi_{HR})$. Similarly, low resolve defenders face low costs with probability π_{LR} and high costs with probability $(1 - \pi_{LR})$. Naturally, $\pi_{HR} > \pi_{LR}$. Upon observing the results of the probe, the defender forms beliefs b_s or b_w according to Baye's rule. The challenger is fully aware that the defender may be responding strategically and incorporates this fact when deriving her beliefs.

Beliefs and Distrust

The challenger will have to determine whether or not to believe the response of the defender. She may either trust the response of the defender or distrust the response. In the latter case, she does not update her beliefs according to the defender's response, believing it to be uninformative. For example, a challenger may observe that the defender had mobilized a disproportionate number of troops from other areas in responding to the probe, leading her to think the strong response she observed may not be genuine. On the other hand, the challenger could

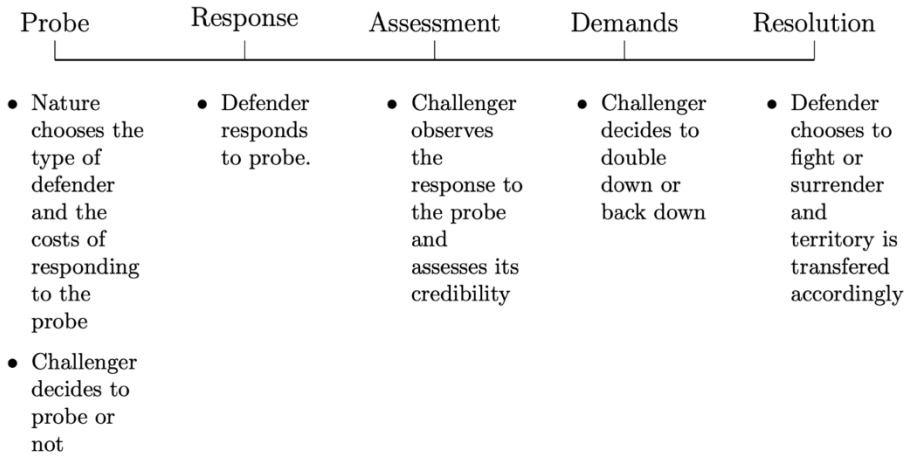
determine that while the defender moved troops away from the probe, indicating a lack of resolve, she may learn that this was in response to some other disturbance that was more pressing. The challenger may also mistakenly distrust a signal. For example, it may be the case that the highly resolved defender's strategy depends upon moving troops quickly to the territory rather than keeping them there, but the challenger might interpret this as a sign that she is overreacting as a deterrent. Because this is not a model of intelligence in the sense of reconnaissance and espionage, we do not explicitly model the sources of this information, but we allow them to affect the beliefs of the challenger.

Formally, we model this possibility of distrust as the possibility that the challenger, for the sake of constructing her beliefs about the defender, acts as though she did not observe the response to her probe. That is, her beliefs after observing the response remain b0. A micro foundation for this result may be an ability to observe a noisy signal of whether costs are high or low; If there is a reason for high costs (another state is attacking them, for instance) uncorrelated with the defender's type, then the challenger knows she should not infer much from the challenger's weak response. However, we do not model this process explicitly. This distrust of the response occurs with exogenous probability γ.

Summary of the Model

Figure 1

Sequence of Events for the Limited Probe Model



For convenience, all variables that will be used in this section are collected in Table 1. We also list the assumptions placed on the parameter values. The first was delineated in the introduction to this section.

Table 1

Meaning of Variables Key

Variable	Meaning
V	The value of the territory
C_{HR}^{C}	The cost to the challenger of fighting a high resolve defender
C_{LR}^{C}	The cost to the challenger of fighting a high resolve defender
C_{HR}^{D}	The cost of fighting to a high resolve defender
C_{LR}^{D}	The cost of fighting to a low resolve defender
C_{S}^{D}	The cost of a strong response to the defender
Δ	A value by which response cost may either rise or fall.
c_p	The cost to the challenger of probing
b_0	The prior belief of the challenger that the defender is highly resolved
b	A generic belief of the challenger that the defender is highly resolved
b_S	The belief of the challenger after a strong response
b_W	The belief of the challenger after a weak response
γ	The probability that the challenger distrusts the response to her probe
π_{HR}	The probability of a low response cost for a high resolve defender
π_{LR}	The probability of a low response cost for a low resolve defender

Assumption 1 (Decisive Resolve). Let the costs to the challenger and defender of war be such that $C_{HR}^{D} < V < C_{HR}^{C}$ and $C_{LR}^{D} > V > C_{LR}^{C}$.

The assumption of relevance says that the prior beliefs of the challenger should show 'substantial' uncertainty about the defender, in the sense that information from the probe could convince them to change their desired course of action. In the absence of this assumption, probing may not generate a strong enough signal to change the challenger's decision, in which case, it is irrelevant. As we are interested in examining why nations do probe, we proceed assuming relevance.

Assumption 2 (Relevance). *The prior belief must satisfy:*

$$b_0 \in \left[\frac{V}{(C_{HR}^{C} - V)\,\frac{\pi_{HR}}{\pi_{LR}} + V}, \frac{V}{[C_{HR}^{C} - V]\,\frac{(1-\pi_{HR})}{(1-\pi_{LR})} + V}\right]$$

Lastly, we require that the frequency of distrust be sufficiently low, so that there is enough opportunity for probe responses to matter in equilibrium.

Assumption 3 (Sufficient Trust). *The probability of distrust must satisfy* $\gamma \leq min\{\frac{V+C_{HR}^{D}-c}{V+C_{HR}^{D}}, \frac{2V-c}{2V}\}$

The equilibrium concept used in this model is Perfect Bayesian Equilibrium. Because this is a dynamic game, we solve using backward induction. Hence, we begin by considering the incentives of the defender to put up a fight for the territory in the event that the challenger doubles down. Table 2 lists the net payoffs for both the defender and the challenger based on the defender type and the challenger's decision to either double down or back down.

Table 2

Payoffs Net of Probing or Response Costs

C	D Type	Outcome	C Payoff	D Payoff
Double Down	Low Resolve	D Surrenders	2V	0
Back Down	Low Resolve	No transfer	0	2V
Double Down	High Resolve	War	$V - C_{HR}^{C}$	$V - C_{HR}^{D}$
Back Down	High Resolve	No Transfer	0	2V

Fighting or Surrendering

Recall that the defender faces a cost of C_{HR}^{D} of fighting if she has high resolve, and of C_{LR}^{D} of fighting if she has low resolve. Naturally, $C_{HR}^{D} < C_{LR}^{D}$. Recall further that the expected value of fighting for the territory, V , is such that $C_{HR}^{D} < \mathrm{V} < C_{LR}^{D}$, so that only highly resolved defenders find fighting to be worthwhile. Consequently, whenever a challenger doubles down against the low resolve defender, the defender surrenders the territory without a fight, giving the challenger a payoff of V and the defender a payoff of 0. Whenever a challenger doubles down against a high resolve defender, the defender fights and receives a payoff of $V - C_{HR}^{D}$, while the challenger receives $V - C_{HR}^{C}$

Lemma 1. When the challenger doubles down, the defender fights if she has high resolve, and surrenders if she has low resolve.

To simplify the remaining exposition, we will make use of the payoffs in Table 2, without further concern for what could happen if the defender made a nonoptimal choice at the last stage.

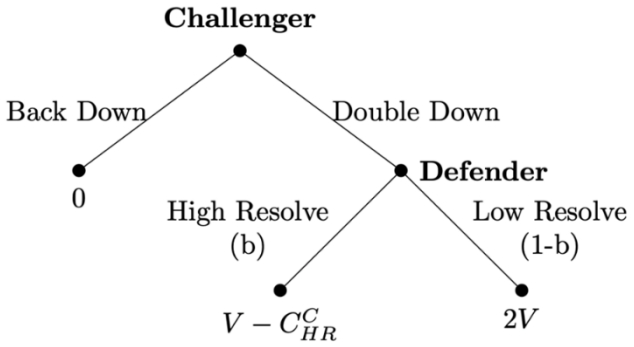
Backing Down or Doubling Down

Because the challenger only wants to double down when she suspects the defender has low resolve, we now determine the critical belief for which she is willing to double down. Because the payoff of doubling down depends upon the type of the defender, the challenger's strategies depend

upon her beliefs about the type of the defender.

Figure 2

Possible Outcomes of Each Offer from the Perspective of the Challenger



Recall that b ∈ [0, 1] is the belief of the challenger that the defender is highly resolved. We now derive the expected payoff of doubling down as a function of this belief to be:

$$EU_{DD}(b) = b(V - C_{HR}^{C}) + (1 - b)(2V) = (2 - b)V - bC_{HR}^{C}$$

To get a sense of which offer is preferred and when, imagine that the challenger came to the bargaining table feeling almost certain that the defender had high resolve, or that $b = 1$. In that case, $EU_{DD}(1) = V - C_{HR}^{C} < 0$. She would know that if she doubles down, she is going to have a costly war that is not worth fighting from her perspective, because $C_{HR}^{C} > V$. *S*he would do better to back down. Conversely, imagine that the challenger is almost certain that the defender is of low resolve, or that $b = 0$. In that case, $EU_{DD}(0) = 2V > 0$. While the challenger could back down and ensure herself a payoff of 0, she knows that she could double down and get the territory without much cost; even if a war were to result, the costs are low enough to be worth fighting. As b increases from 0 toward 1, the odds of a costly war increase, and the odds of getting the territory without much cost fall. At some point between 0 and 1, there is a level of probability for which

the preference "flips" from doubling down and backing down. We call this level of probability $b*$, and it is equal to $\frac{2V}{V+C_{HR}^{C}}$

Lemma 2. Let $b* = \frac{2V}{V+C_{HR}^{C}}$. *The challenger prefers to back down whenever* $b > b*$, *and prefers to double down whenever* $b \leq b*$

Responding to the Probe

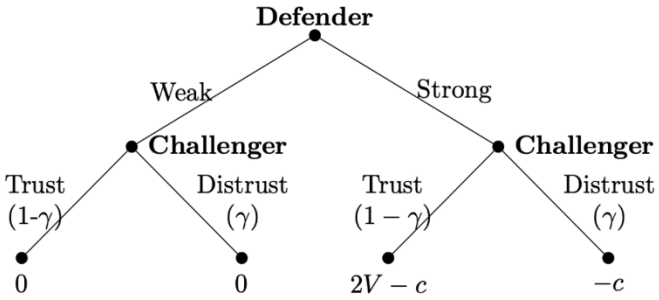
The challenger forms her belief about the resolve of the defender by observing the response of the defender to her probe. We now establish the equilibrium of the subgame in which the probe has been sent and verify formally that the beliefs of the challenger will follow the response of the defender.

On the part of the challenger, there are four possible outcomes to consider. She could respond strongly and either be trusted or distrusted, and she could respond weakly and either be trusted or distrusted. What happens in the event of a trusted response will depend on the beliefs formed by the challenger; what happens in the event of a distrusted response will depend upon the challenger's prior.

To begin our analysis, suppose that $b_0 < b*$, that the challenger backs down after a trusted strong response, and that the challenger doubles down after a trusted weak response. In the event of distrust, or a trusted weak response, the challenger will double down, and she will surrender her territory. In the event of a trusted strong response, the defender will keep her territory without fighting. The payoffs faced by a low resolve defender are presented in Figure 3.

Figure 3

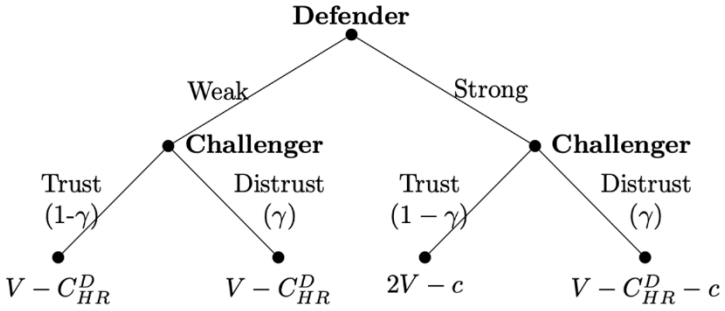
Low Costs, Low Resolve, Low Prior



The defender prefers to send a strong response if and only if (1−γ)(2V −c)+γ(−c) ≥ 0. Straightforward algebra implies that this holds if $\gamma \leq \frac{2V-c}{2V}$. For a high resolve defender, doubling down on the part of the challenger incites war, so the payoffs change to those in Figure 4.

Figure 4

Low Costs, High Resolve, Low Prior



Consequently, the high resolve defender wishes to send a strong signal if

$$V - C_{HR}^{D} \leq (1-\gamma)(2V-c) + \gamma(V - C_{HR}^{D} - c).$$

Algebraic manipulation implies that this condition holds if

$$\gamma \leq \frac{V+C_{HR}^{D}-c}{V+C_{HR}^{D}}.$$

In the event that b0 > b∗ so that the challenger backs down in the event of distrust, the payouts are modified as in Figures 5 and 6.

Figure 5

Low Costs, Low Resolve, High Prior

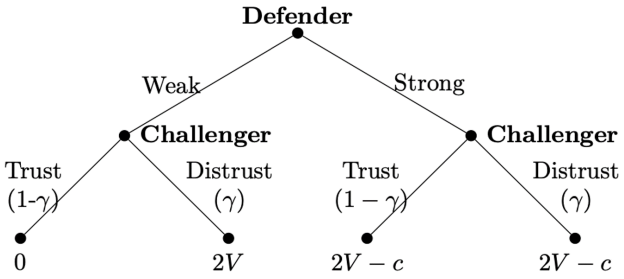
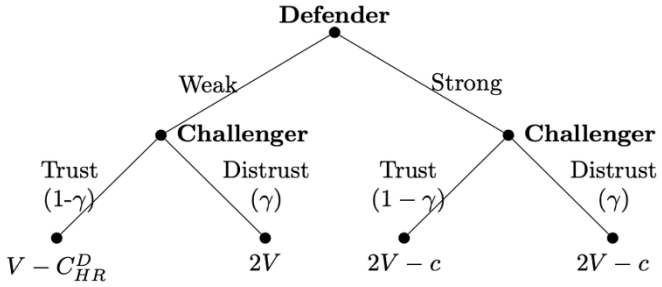


Figure 6

Low Costs, High Resolve, High Prior



The defender prefers to send a strong response if and only if $\gamma(2V) \leq 2V - c$, or if $\gamma \leq \frac{2V-c}{2V}$. For a high resolve defender, doubling, doubling down on the part of the challenger incites war, so the

payoffs change to those in Figure 6. Consequently, the high resolve defender prefers to send the strong response whenever

$$(1-\gamma)(V-C_{HR}^{D})+\gamma 2V \leq 2V-c.$$

This holds whenever

$$\gamma \leq \frac{V+C_{HR-c}^{D}}{V+C_{HR}^{D}}.$$

Consequently, given the conjectured response of the challenger, it is indeed rational for the defender to send a strong response whenever it is not prohibitively expensive, as long as the probability that it is distrusted, γ, is sufficiently small. Specifically, it must be the case that $\gamma \leq min\{\frac{V+C_{HR-c}^{D}}{V+C_{HR}^{D}}, \frac{2V-c}{2V}\}$, which is ensured by the assumption of sufficient trust.

We now complete our analysis by showing that the proposed behavior by the challenger is consistent with her beliefs. In order for the challenger's behavior to be rational upon observation of the response, it must be that $b_s > b*$ and $b_w < b*$. In order for this to be true, high resolve defenders must choose high resources sufficiently often, while low resolve defenders do not choose high resources too often. Recalling that c + Δ is high enough to make it irrational to choose to send a strong response, we need only consider the probability that the defender experiences high costs given her type. Holding the properties of the probe constant, the beliefs will be consistent with this equilibrium as long as the relevance assumption holds. Essentially, it must be the case that states are substantially uncertain about the type of the defender; enough so to actually change their actions based on the results of the probe.

Theorem 1. The following set of strategies, with accompanying beliefs derived by Bayes' rule, is an equilibrium of the post-probe sub-game. The defender sends a strong response whenever her response costs are low, and a weak response otherwise. The challenger doubles down after a trusted weak response and backs down after a trusted strong response. She doubles down after a distrusted response if $b_0 < b*$ *and backs down if* $b_0 \geq b*$.

The Decision to Probe

What is the value to the challenger of additional information about the type of the defender? It comes down to how it will affect her decision at the bargaining table. There are two cases to consider.

Case 1: Decreased Caution. One possibility is that, while the challenger would have backed down in the absence of information because $b_0 > b*$, the probe may convince her otherwise. If the challenger sees a strong response to the probe she is even more certain that the challenger is of higher resolve than before because $b_s > b_0$, and so she continues to back down. However, if a weak response is observed, she may now feel emboldened to demand more from the defender, because $b_w < b*$. In this case, a weak response to the probe contains enough information about the defender that she updates her beliefs sufficiently to change her mind and behave less cautiously. She previously backed down and always avoided war, whereas now, she responds to information that the defender may be of low resolve by doubling down, potentially leading to war.

Case 2: Increased Caution. It is also possible that while the challenger doubled down in the absence of information because $b_0 < b*$, she may decide to be more cautious after the probe. If the challenger sees a weak response to the probe, then she is even more convinced that the defender is of low resolve, because $b_w < b_0$, and therefore continues to double down. However, in the event that the challenger responds strongly, she may be confident enough that the defender is

actually strong to back down. That is, $b_s > b*$. This means that the challenger is behaving more cautiously than before. She was previously doubling down no matter what, but now, she is responding to information that the defender may be strong by backing down.

The rational self-interest of the state implies that these behavior changes strictly increase challenger welfare. Because the challenger does not know the outcome of the probe when she launches it, she must consider her expected utility across one more dimension of uncertainty, the outcome of the probe. Let EU_{probe} be her expected utility from launching a probe, and let $EU_{noprobe}$ be her expected utility in the absence of the probe. Consider the case of increased caution. The challenger is choosing, on the basis of her information, to back down, when she could have continued to choose to double down. That is, she has the option of simply ignoring the probe's information and acting accordingly. This would ensure her an identical payoff to when she did not probe at all. But, the challenger in this case prefers to change her strategy, so it must be that making this change raises her utility. Hence, when she probes, she can always achieve at least the same expected utility as if she did not probe, which is true in any case when she does not change her behavior, but does gain expected utility in any case where she changes her behavior. This occurs whenever she receives a strong signal and trusts it. Because trusted strong signals occur with positive probability, $EU_{probe} > EU_{noprobe}$.

Conversely, consider the case of decreased caution. In this case, the challenger is choosing to double down when she could have continued to choose to back down. If the state is rational, it must be the case that doubling down gives a higher level of expected utility than choosing to back down. Further, choosing to back down no matter the outcome of the probe is equivalent to simply not launching a probe. Hence, whenever she switches to doubling down, she is doing so because the expected utility of doing so exceeds the expected utility of not probing. This happens whenever

she observes a weak signal and trusts it. Because this happens with positive probability in the case of decreased caution, $EU_{probe} > EU_{noprobe}$.

In Lemma 3, we verify this logic algebraically. Given that there is some benefit to the probe in expectation, it is in the self interest of the challenger to send a probe as long as the cost of doing so, cp, is sufficiently small.

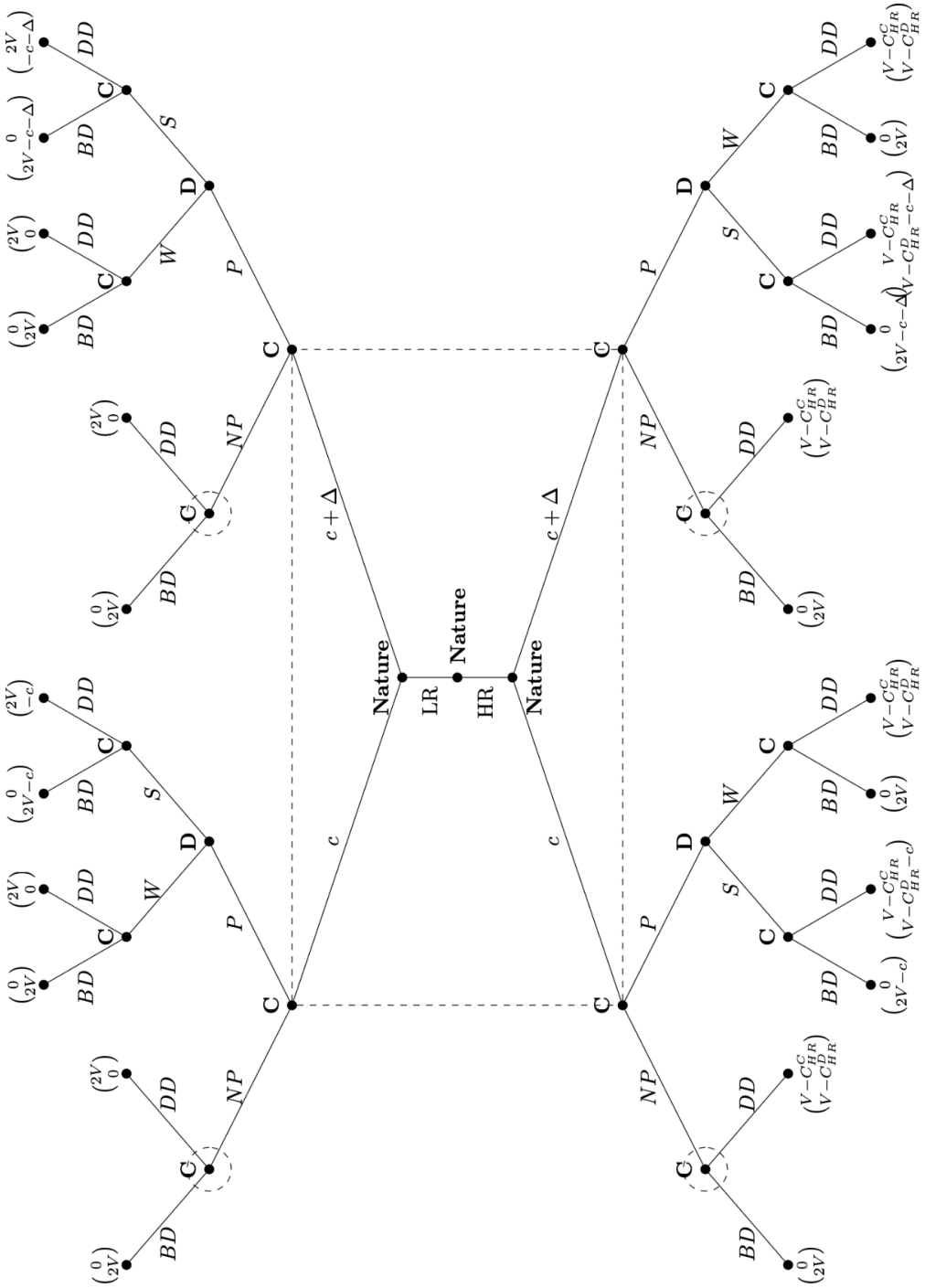
Lemma 3. It is the case that $EU_{probe} > EU_{noprobe}$, for any $b_0 \in (0,1)$.

With this, we may now present the equilibrium of the entire game.

Theorem 2. The following strategies on the part of the challenger and the defender, with beliefs defined by Bayes' rule, constitute a Perfect Bayesian Equilibrium: The challenger sends a probe if $EU_{probe}(b) - EU_{noprobe}(b) \geq c_p$, and does not otherwise. The defender sends a strong response whenever her response costs are low, and a weak response otherwise. The challenger doubles down after a trusted weak response and backs down after a trusted strong response. She doubles down after a distrusted response if $b_0 < b*$ and backs down if $b_0 \geq b*$.

Figure 7

A Game Tree Presenting the Case in Which $\gamma = 0$



Conclusion

This article argued that state actors can employ strategies of provocation to extract private information about the resolve of their adversaries. While previous scholarship has focused on how state actors use signaling to increase their individual payoffs, there is a gap in the current literature when it comes to explaining strategies aimed at extracting information rather than communicating or manipulating it. Our model suggests that the use of provocative probes is a valid approach, and state actors can benefit from employing provocations to test the resolve of their adversaries.

Based on these findings, several theoretical implications arise. If we conclude that state actors can successfully provoke their opponents to extract valuable private information about their resolve without risking unnecessarily costly warfare, it implies that there must be potential countermeasures to mitigate the effectiveness of such strategies. Further research is needed to explore these counterstrategies. Additionally, these arguments suggest that there is an additional layer of complexity in the bargaining process. Minor conflicts preceding a settlement or more significant hostile actions can be seen as rational actions aimed at gaining a better understanding of the adversary.

References

Avenhaus, R., & Zartman, I. W. (2007). Introduction: Formal models of, in, and for international negotiations. In R. Avenhaus & I. W. Zartman (Eds.) *Diplomacy games: Formal models and international negotiations.* (pp. 1-22) http://dx.doi.org/10.1007/978-3-540-68304-9_1

Fearon, J. D. (1994). Signaling versus the balance of power and interests: An empirical test of a crisis bargaining model. *Journal of Conflict Resolution*, 38(2), 236-269. https://doi.org/10.1177/0022002794038002004

Fearon, J. D. (1997). Signaling foreign policy interests: Tying hands versus sinking costs. *Journal of Conflict Resolution*, *41*(1), 68-90. https://www.jstor.org/stable/174487

Fearon, J. D. (1998). Bargaining, enforcement, and international cooperation. *International Organization*, *52*(2), 269–305. http://www.jstor.org/stable/2601276

Filson, D., & Werner, S. (2002). A bargaining model of war and peace: Anticipating the onset, duration, and outcome of war. *American Journal of Political Science*, *46*(4), 819–837. https://doi.org/10.2307/3088436

Gartzke, E. A., Carcelli, S., Gannon, J. A., & Zhang, J. J. (2017, August). Signaling in foreign policy. *Oxford Research Encyclopedia of Politics.* https://doi.org/10.1093/acrefore/9780190228637.013.481

George, A. L., & Smoke, R. (1974). Deterrence in American foreign policy: Theory and practice. Columbia University Press. https://hdl.handle.net/2027/heb00720.0001.001

Huei, P. Y. (2019, June 25). Strait rituals: China, Taiwan, and the United States in the Taiwan Strait crises, 1954-1958. *Hong Kong Scholarship Online*. https://doi.org/10.5790/hongkong/9789888208302.001.0001

Jervis, R. (2017). *Perception and misperception in international politics: New edition*. *Princeton University Press*. https://doi.org/10.1515/9781400885114

Kertzer, J., Rathbun, B., & Rathbun, N. (2020). The price of peace: Motivated reasoning and costly signaling in international relations. *International Organization*, *74*(1), 95-118. https://doi:10.1017/S0020818319000328

Kertzer, J. (2017, April). Resolve, time, and risk. *International Organization*, *71*(S1), S109-S136. https://doi.org/10.1017/S0020818316000424

Lake, D. A. (2010). Two cheers for bargaining theory: Assessing rationalist explanations of the Iraq war. *International Security*, *35*(3), 7-52. https://doi.org/10.1162/ISEC_a_00029

Powell, R. (2002). Bargaining theory and international conflict. *Annual Review of Political Science*, 5(1), 1-30. https://doi.org/10.1146/annurev.polisci.5.092601.141138

Schelling, T. C. (1980). The strategy of conflict: With a new preface by the author. *Harvard University Press*. https://www.google.com/books/edition/The_Strategy_of_Conflict/7RkL4Z8Yg5AC?hl=en&gbpv=0

Schelling, T. C. (2008). Arms and influence. In Strategic Studies (pp. 96-114). *Routledge*. https://www.taylorfrancis.com/chapters/edit/10.4324/9780203928462-13/arms-in%EF%AC%82uence-thomas-schelling.

Snidal, D. (1985). The game theory of international politics. *World Politics*, *38*(1), 25-57. https://doi.org/10.2307/2010350

Tzu, S. (2022). The Art of War. Germany. *Golden Classic Audiobooks.* (Original work published between 475 and 221 B.C.E.)

Wagner, R. H. (2000). Bargaining and war. *American Journal of Political Science*, *44*(3), 469–484. https://doi.org/10.2307/2669259

Waltz, K. (2018). Man, the state, and war: A theoretical analysis. *Columbia University Press*. https://doi.org/10.7312/walt18804

Lengxi, W. (1994). Inside story of the decision making during the shelling of Jinmen. *Zhuanji wenxue [Biographical Literature]*, (1), 11. https://digitalarchive.wilsoncenter.org/document/memoir-wu-lengxi-inside-story-decision-making-during-shelling-jinmen

Appendix

Mathematical Proofs

Proof of Lemma 1. The proof of this Lemma is given in the text preceding it.

Proof of Lemma 2. Note that $EU_{DD}(0) = V >$ and $EU_{DD}(1) = V - C_{HR}^{C} < 0$. It is clear that EU_{DD} is continuous, and further, $EU'_{DD}(b) = -V - C_{HR}^{C} < 0$. Hence, by the Intermediate Value Theorem, there exists a value $b* \in (0,1)$ such that $EU_{DD}(b*) = 0$, namely, $b* = \frac{2V}{V+C_{HR}^{C}}$. Because $EU_{DD}(b)$ is strictly decreasing, it must be the case that $EU_{DD}(b) < 0$ for all and $b > b*$. Given that the payoff of backing down is 0, this establishes that the challenger prefers to double down whenever $b > b*$ and to back down whenever $b > b*$.

Proof of Theorem 1. We begin this proof by considering the beliefs at each information set. Let b_s be the belief after a strong response, b_w be the belief after a weak response. Formally,

$$b_s = \frac{\pi_{HR} b}{\pi_{HR} b + \pi_{LR}(1-b)}$$

$$b_w = \frac{(1-\pi_{HR}) b}{(1-\pi_{HR}) b + (1-\pi_{LR})(1-b)}$$

By Lemma 2, it is optimal for the challenger to double down if and only if $b < b*$.

Note that $b_s > b*$ if

$$\frac{\pi_{HR} b}{\pi_{HR} b + \pi_{LR}(1-b)} > \frac{2V}{V+C_{HR}^{C}},$$

Which is equivalent to

$$\pi_{HR} b > \frac{2V}{V+C_{HR}^{C}} \pi_{HR} b - \frac{2V}{V+C_{HR}^{C}} \pi_{LR} b + \frac{2V}{V+C_{HR}^{C}} \pi_{LR} \quad .$$

Moving all terms with b to the left hand side and factoring gives

$$b\left[\pi_{HR} - \frac{2V}{V+C_{HR}^{C}}\pi_{HR} + \frac{2V}{V+C_{HR}^{C}}\pi_{LR}\right] > \frac{2V}{V+C_{HR}^{C}}\pi_{LR} \ ,$$

which is equivalent to

$$b > \frac{\frac{2V}{V+C_{HR}^{C}}\pi_{LR}}{\pi_{HR} - \frac{2V}{V+C_{HR}^{C}}\pi_{HR} + \frac{2V}{V+C_{HR}^{C}}\pi_{LR}} \ .$$

Simplifying,

$$b > \frac{V\pi_{LR}}{\frac{1}{2}(V+C_{HR})\pi_{HR} - V\pi_{HR} + V\pi_{LR}} \ .$$

$$b > \frac{V}{\frac{1}{2}(C_{HR}-V)\frac{\pi_{HR}}{\pi_{LR}} + V} \ .$$

This final inequality must hold by the relevance assumption. Consequently, it is rational for a state to back down after a strong signal. Further, $b_w > b*$ if

$$\frac{(1-\pi_{HR})b}{(1-\pi_{HR})b + (1-\pi_{LR})(1-b)} < \frac{2V}{V+C_{HR}^{C}}$$

$$(1-\pi_{HR})b < \frac{2V}{V+C_{HR}^{C}}(1-\pi_{HR})b - \frac{2V}{V+C_{HR}^{C}}(1-\pi_{LR})b + \frac{2V}{V+C_{HR}^{C}}(1-\pi_{LR})$$

$$(1-\pi_{HR} - \frac{2V}{V+C_{HR}^{C}}(1-\pi_{HR}) + \frac{2V}{V+C_{HR}^{C}}(1-\pi_{LR}))b < \frac{2V}{V+C_{HR}^{C}}(1-\pi_{LR})$$

$$([C_{HR}^{C} - V](1-\pi_{HR}) + 2V(1-\pi_{LR}))b < 2V(1-\pi_{LR})$$

$$(\frac{1}{2}\,[(C_{HR} - V]\,\frac{(1-\pi_{HR})}{(1-\pi_{LR})} + V)b < V$$

$$b < \frac{V}{\frac{1}{2}[(C_{HR}-V]\frac{(1-\pi_{HR})}{(1-\pi_{LR})}+V}\ .$$

This final inequality must hold by the relevance assumption. Consequently, it is rational for a state to double down after a weak signal.

Consequently, given the beliefs derived by Bayes' rule under the assumption of these strategies, the challenger finds it optimal to double down after weak responses or after distrusting strong responses, and to back down after strong responses or distrusting weak responses.

We now turn to the decisions of the defender. By the assumption that high costs are prohibitively expensive, we need only to show that it is worthwhile to send a strong response when costs are low for both the high and low resolve defenders.

Case 1: $b_0 < b*$. The low resolve defender prefers to send a strong response if and only if $(1-\gamma)(2V-c) + \gamma(-c) \geq 0$, or if $\frac{2V-c}{2V} \geq \gamma$. The high resolve defender prefers to send a strong response if $(1-\gamma)(V - C_{HR}^{D}) + \gamma 2V \leq 2V - c$. This implies the following inequalities:

$$V - \gamma V - C_{HR}^{D} + \gamma C_{HR}^{D} + \gamma 2V \leq 2V - c$$

$$\gamma C_{HR}^{D} + \gamma V \leq V + C_{HR}^{D} - c$$

$$\gamma \leq \frac{V+C_{HR}^{D}-c}{V+C_{HR}^{D}}\,.$$

Hence, the defender's actions are consistent with the proposed equilibrium as long as $\gamma \leq min\{\frac{V+C_{HR}^{D}-c}{V+C_{HR}^{D}}, \frac{2V-c}{2V}\}$. Given Assumption 3, this inequality is satisfied.

Case 2: $b_0 \geq b*$. The low resolve defender prefers to send a strong response if and only if $(1-\gamma)(2V-c)+\gamma(-c) \geq 0$, or if $\frac{2V-c}{2V} \geq \gamma$. The high resolve defender prefers to send a strong response if $(1-\gamma)(V-C_{HR}^{D})+\gamma 2V \leq 2V-c$. This implies the following inequalities:

$$V-\gamma V-C_{HR}^{D}+\gamma C_{HR}^{D}+\gamma 2V \leq 2V-c$$

$$\gamma C_{HR}^{D}+\gamma V \leq V+C_{HR}^{D}-c$$

$$\gamma \leq \frac{V+C_{HR}^{D}-c}{V+C_{HR}^{D}}.$$

Hence, the defender's actions are consistent with the proposed equilibrium as long as $\gamma \leq min\{\frac{V+C_{HR}^{D}-c}{V+C_{HR}^{D}}, \frac{2V-c}{2V}\}$. Given Assumption 3, this inequality is satisfied.

Hence, the defender wishes to send a strong response whenever costs are low. As the beliefs are consistent with the strategies and no player has an incentive to deviate, we have shown that these strategies are a Perfect Bayesian Nash Equilibrium.

Proof of Lemma 3. First suppose $b_0 < b*$. In the absence of a probe, the challenger doubles down. With the probe, the challenger instead backs down when there is a strong signal which she trusts. The difference in expected utility from this change is

$$b_0\pi_{HR}(1-\gamma)(C_{HR}^{C}-V)+(1-b_0)\pi_{LR}(1-\gamma)(-2V)$$

The first term is the probability of a trusted strong signal from a high resolve defender, times the increase in utility from avoiding a war against a high resolve defender. The second term is the probability of a strong response from a low resolve defender, times the utility lost by backing down. This difference is positive if and only if

$$b_0\pi_{HR}(C_{HR}^C - V) + b_0\pi_{LR}2V > 2V$$

or, equivalently, if

$$b_0 > \frac{2V}{\pi_{HR}(C_{HR}^C - V) + \pi_{LR}2V}.$$

This is ensured by the relevance assumptions.

Now suppose $b_0 > b*$. In the absence of a probe, the challenger backs down. With the probe, the challenger instead doubles down when there is a weak signal which he trusts. The difference in expected utility from this change is

$$b_0(1 - \pi_{HR})(1 - \gamma)(V - C_{HR}^C) + (1 - b_0)(1 - \pi_{LR})(1 - \gamma)(2V)$$

The first term is the probability of a trusted weak signal from a high resolve defender, times the decrease in utility from starting a war against a high resolve defender. The second term is the probability of a weak response from a low resolve defender, times the utility gained by demanding his territory. This difference is positive if and only if

$$-b_0(1 - \pi_{HR})(C_{HR}^C - V) - b_0(1 - \pi_{LR})(2V) > -2V$$

which is equivalent to

$$b_0 < \frac{2V}{(1-\pi_{HR})(C_{HR}-V)+(1-\pi_{LR})(2V)}.$$

This is ensured by the relevance assumptions.

Proof of Theorem 2. Lemma 3 implies the desirability of probing given a low enough cost. The remainder follows directly from Theorem 1.

A game tree presenting the case in which $\gamma = 0$ is presented in Figure 8. This case is chosen as it dramatically simplifies the tree to where it may be presented in a single page.