Fear, Appeasement, and the Effectiveness of Deterrence[1]

Ron Gurantz[2] and Alexander V. Hirsch[3]

April 1, 2015

[1]We thank Robert Trager, Barry O'Neill, Tiberiu Dragu, Kristopher Ramsay, Robert Powell, Doug Arnold, Mattias Polborn, participants of the UCLA International Relations Reading Group, Princeton Q-APS International Relations Conference, and Cowbell working group, and two anonymous reviewers for helpful comments and advice.

[2]Corresponding Author. University of California, Los Angeles. Department of Political Science, 4289 Bunche Hall, Los Angeles, CA 90095-1472. Phone: (310) 825-4331. Email: rongurantz@ucla.edu.

[3]Associate Professor of Political Science, Division of the Humanities and Social Sciences, California Institute of Technology, MC 228-77, Pasadena, CA 91125. Email: avhirsch@hss.caltech.edu.

Abstract

Governments often fear the future intentions of their adversaries. In this paper we explore how this fear affects their willingness to fight, and thus ability to credibly deter, and show that it can make deterrent threats credible under seemingly incredible circumstances. We consider a model in which a defender seeks to deter a transgression that has both intrinsic and military value, and examine how the defender's fear that the challenger will exploit her gains in a future military confrontation affects the credibility of his deterrent threat. We derive conditions under which even a very minor transgression can effectively become a "test" of the challenger's future belligerence. As a result, the defender's deterrent threat of war becomes credible in equilibrium, even when the transgression does not immediately appear to be worth fighting over, and the ex-ante likelihood that the challenger is belligerent is infinitesimally small. We also show that the defender's fear can actually benefit her by allowing her to credibly deter. We demonstrate the robustness of our results to a variety of extensions, and apply the model to analyze the Turkish Straits Crisis of 1946.

Word Count: 13,870 (including Appendix)

1 Introduction

Early in the Cuban Missile Crisis, Robert Kennedy endorsed an invasion of Cuba despite the risk that it would lead to war with the Soviet Union. He argued that the Soviet Union had demonstrated its aggressiveness by placing missiles in Cuba *after* Kennedy's warning against it, and that if they responded to an invasion of Cuba with full-scale war, it would only demonstrate that such a war was inevitable in six months or a year anyway. Senator Richard Russell later made a very similar case for invasion, arguing that the Soviet Union had revealed their intention to challenge the United States around the globe, and in all likelihood provoke war, by their willingness to place missiles in Cuba after Kennedy had warned them that such an act would carry major consequences.[1]

Their reasoning highlights an important source of deterrence credibility that has been overlooked in the literature. Governments often fear the future intentions of an adversary, and will use their crisis behavior as a "test" of those intentions. If an adversary's defiant behavior reveals hostile intentions that go beyond the immediate stakes of a crisis, this can give the government a reason to respond with war. The adversary may then be deterred from taking those actions because they do not want to "fail the test" and trigger a conflict.

In this paper, we develop a model in which deterrence is sustained by an adversary's desire to avoid signaling hostility and provoking war. We show that this mechanism can be a powerful source of credible deterrence since, under the conditions we identify, even the most minor transgression can signal hostility and make major war a rational response. As a result, the model can account for a wide range of deterrence scenarios, including some Cold War crises where relatively minor transgressions were deterred by the threat of nuclear war.

The model also produces surprising results that previous theories have been unable to

[1] May and Zelikow 1997, 66-67, 172-173. See also Trachtenberg 2007.

identify. We link the study of deterrence with the literature on bargaining with endogenous power shifts, demonstrating that a condition that leads to conflict when the parties bargain under complete information also leads to deterrence when the defender is uncertain about the challenger's intentions. We also show that the defender can be better off fearing the challenger's intentions rather than knowing that they are benign, since it is the fear that the challenger may be hostile in the future that sustains the credibility of his deterrent.

Intuition and Example The model relies on two common features of international crises that are rarely incorporated into deterrence models: a defender's fear and uncertainty about a challenger's intentions, and endogenous power shifts.[2] It leaves out the more traditional deterrence mechanisms of reputation and commitment: the defender's preferences are commonly known, and he has no mechanism for tying his hands. In the model, there is a potential transgression that has both *direct* value to the challenger if there is peace, and *military* value if there is war; i.e., it results in an endogenous power shift. The defender would prefer to allow the transgression if it would lead to peace, but is uncertain about the challenger's intentions, and entertains the possibility that she is unappeasably belligerent.[3]

We first show that combining these two ingredients can produce credible deterrence under seemingly incredible circumstances: when the challenger is very unlikely to be unappeasably belligerent, the transgression is incredibly minor, and the threatened response is a major and costly war. Why does this happen? Intuition suggests that a peaceful defender would allow a minor transgression rather than fight if the challenger is unlikely to exploit it in a future war. But this intuition ignores a key element: that a *credible* threat of war affects what the defender can infer when the challenger transgresses.

[2]Fearon 1996; Powell 2006.

[3]Throughout the paper we refer to the defender as "he" and the challenger as "she."

Specifically, a challenger who transgresses *in the face of a credible threat of war* reveals that she prefers triggering a war to the status quo. This revelation can lead a fearful defender to infer that war is inevitable, and induce him to initiate it immediately rather than waiting to first be weakened. Finally, if the challenger believes the defender to be using this logic, then she will indeed be deterred from transgressing unless she actually desires war, fulfilling the defender's expectations. In our mechanism, the challenger's reaction to the deterrent threat is thus part of what sustains its credibility: "types" of challengers against whom a defender would *not* want to fight are "screened out."

An example helps to both clarify the logic, and demonstrate the mechanism's relevance to a historical deterrence scenario. During the early Cold War, the United States feared that the Soviet Union intended to launch a full-scale war against Western Europe and the United States. A 1952 National Security Council report on possible U.S. responses to Soviet aggression against West Berlin begins by asserting that "control of Berlin, in and of itself, is not so important to the Soviet rulers as to justify involving the Soviet Union in general war."[4] Thus, the report reasons that the Soviet Union will only attack West Berlin if they "decide for other reasons to provoke or initiate general war," and that the United States would therefore "have to act on the assumption that general war is imminent." In other words, an invasion of Berlin must imply that the Soviet Union both expects to trigger a wider war and affirmatively desires it, rather than implying that they think they can conquer Berlin without war. Since an invasion would imply that a wider war is imminent, the United States was to respond with a full mobilization and "full implementation of emergency war plans," thereby fulfilling the United States' commitment to fight in the event of an invasion.

[4] *Foreign Relations of the United States (FRUS), 1952-54 VII Part 2*, 1268-1269.

Which Actions Sustain Deterrence? After clarifying the mechanism, our first main result is to derive a condition under which the combination of fear and endogenous power shifts can always produce credible deterrence. The key is to consider what sorts of transgressions most effectively "test" for the inevitability of war. It turns out that the objective magnitude of a transgression is *not* what makes it a good test. Instead, what makes a transgression a good test is the extent to which allowing it would *fail to appease an already belligerent challenger*. The reason is that the defender can already infer the challenger's initial belligerence from observing the transgression itself when the deterrent threat is credible.

What sort of transgression cannot appease an already-belligerent challenger? The literature on bargaining under complete information with endogenous power shifts already answers this question: it is one whose military value to the challenger *exceeds* its direct value.[5] The reason is that allowing such a transgression will only increase an already-belligerent challenger's appetite for war. The bargaining literature shows that when the challenger's initial belligerence under the status quo is *known*, similar conditions result in inefficient war or the gradual elimination of the defender. We show that when the challenger's belligerence is merely *feared* – even if only with infinitesimal probability – a transgression with this property can always be effectively deterred with a credible threat of war, regardless of how costly the war, or how minor the transgression.

Our more general insight is that it is not only the size of a transgression that matters for deterrence: exceedingly minor transgressions can be credibly deterred with the threat of major war, while larger transgressions may not meet the condition. Equally important is the *relationship* between a transgression's military and direct values. This relationship is what determines the effectiveness of appeasement, the ability to infer the inevitability of war, and ultimately the credibility of deterrence. Extending this insight yields our second

[5]Fearon 1996; Schwarz and Sonin 2008.

main result: that deterrence is more likely the greater is the *difference* between the military and direct value of the transgression to the challenger.

Together, our results suggest that empirical studies of deterrence that control for the "interests at stake" in a dispute are flawed because they fail to separately control for these values or the relationship between them.[6] For example, our results suggest that a challenger may treat a defender's threat to fight for a barren rock as credible, *if* the rock yields even a minor strategic advantage. The defender can reason that if the challenger is already belligerent, allowing her to occupy the rock will only make her (a bit) more so. Conversely, our results also suggest that the challenger may actually discount a defender's threat to fight for something much more substantial, like a valuable population center. The defender may reason that controlling that population center is of sufficient direct value to the challenger to appease her even if she is initially belligerent.

Can Fear Improve Welfare? The defender's fear is essential to the deterrence mechanism we present; without it, the challenger could not be influenced by the "signaling" implications of her actions. This raises an unusual possibility: might the defender actually benefit from being incompletely informed about the challenger's intentions? To answer this question, we ask what would happen if the defender were informed of the challenger's "type" at the start of the game, rather than attempting to infer it from her actions.

We first show that the probability of deterrence unambiguously decreases. In the baseline model, the defender's ability to deter a challenger who is peaceful or appeasable is predicated on his fear that she is actually unappeasably belligerent. If he already knows that this is not the case – and the challenger knows that he knows it – then the challenger will exploit his known preference for appeasement. A potential empirical implication of this result is that

[6]See Huth 1999 for summary of empirical literature.

events introducing uncertainty about a potential challenger's preferences, such as a sudden regime change, can actually increase the likelihood that deterrence succeeds by creating plausible fear on the part of a defender.

We last show that when the difference between the transgression's military and direct value is sufficiently large, the defender does indeed benefit from his fear and uncertainty in expectation. This fear induces the challenger to behave cautiously, which benefits the defender. While eliminating it could potentially prevent unnecessary wars, such wars rarely or never occur under the condition we identify.

The paper proceeds as follows. In the next section we discuss related literature. We then present a simple mathematical example. Next, we present the formal model and derive our main results. Following that, we briefly discuss the robustness of our main results to changes to the information structure, the game sequence, and the bargaining protocol. In particular, we embed the model in a game of "salami tactics," in which the challenger can attempt a series of transgressions, and wars occur because the challenger attempts a transgression against which the defender is willing to fight.[7] Next we present a case study of the crisis over the Turkish Straits in 1946 to demonstrate that our logic can shed light on the United States' decision to defend Turkey from Soviet invasion. Finally, we conclude with a summary and questions for future research.

2 Related Literature

Much of the crisis bargaining literature was developed to understand how a state could credibly threaten to use force – often nuclear force – to defend interests that do not immediately

[7]See Powell 1996.

appear to be worth fighting over.[8] This literature has naturally focused on characteristics or abilities of the *defender* – his resolve, his reputation, or his ability to commit – that can bolster his credibility. For example, some scholars have argued that defending states can bolster their credibility by making physical or verbal commitments, such as placing "trip-wire" forces in a threatened area, or giving public speeches before domestic audiences.[9] Others have argued that states can make "threats that leave something to chance."[10] Finally, many have focused on a defender's knowledge that failing to carry out a threat could damage its reputation for "resolve" in future interactions.[11]

Our model assumes away these previously-identified sources of deterrence credibility. The defender has no commitment mechanism to increase the cost of a concession. There are no probabilistic moves. And his intrinsic preference for appeasement is commonly known; there is no reputation for resolve that can preserved by fighting, as in "chain-store" games where firms resist new entrants to preserve the belief they may intrinsically prefer resisting.[12]

Instead, we look at characteristics of the *challenger* that can enhance the defender's credibility – her potential belligerence, and the defender's fear of it – and how these interact with characteristics of the *transgression* – its ability to appease a belligerent challenger. In so doing, we can account for a range of deterrence scenarios that remain puzzling in light of previous theories; specifically, ones in which objectively minor transgressions were deterred with the threat of a truly catastrophic war. After all, it is difficult to imagine a commitment device that is sufficiently strong, or consequences from a damaged reputation that are sufficiently great, to induce a state to affirmatively choose a nuclear war. But in

[8]An exception is "perfect deterrence theory" which considers the "inherent credibility" of deterrence based on the direct value of the interests at stake to the defender. See Zagare 2004.

[9]See Fearon 1994; O'Neill 1999; Schelling 1966; Slantchev 2011.

[10]Schelling 1966. See Nalebuff 1986 and Powell 1987 for formal models that rely on this mechanism.

[11]Alt, Calvert and Humes 1988; Sechser 2010.

[12]Milgrom and Roberts 1982.

our model, credible deterrence is always possible when allowing the transgression would not appease a belligerent challenger; even when war would be truly catastrophic.

The characteristics we examine are prominent in the international relations literature outside of the study of deterrence. First, recent work has examined the sustainability of peace when parties fear that their adversaries are unappeasably belligerent. Several authors analyze how this can induce a "spiral" of fear that causes peace to unravel,[13] while Acharya and Grillo focus on a state's incentive to mimic a "crazy type."[14]

Second, many works study endogenous power shifts. An early example is Powell, who focuses on the vulnerability of states to "salami tactics."[15] Several later authors study bargaining under complete information, and produce the common insight that it is difficult or impossible to maintain a stable settlement if the available peaceful arrangements shift the military balance toward the challenger more quickly than they increase her payoff from peace.[16] This condition closely resembles the unappeasability condition in our model for successful deterrence; thus, one interpretation of our result is that a condition generating war or unstable settlements in a complete information context helps sustain deterrence in an incomplete information context.

Finally, more recent works explore the consequences of endogenous power shifts for other behaviors; Slantchev examines states' incentives to make costly unilateral decisions to shift power in their direction by investing in military capacity,[17] while Kydd shows how endogenous power shifts create a rationale for states to make costly commitments in the form of assurances that they won't exploit them.[18]

[13] See Baliga and Sjostrom 2009; Chassang and i Miquel 2010.

[14] Acharya and Grillo 2014.

[15] Powell 1996. See also Schelling 1966, 66.

[16] Fearon 1996; Schwarz and Sonin 2008.

[17] Slantchev 2011.

[18] Kydd and McManus 2014.

3 Example

To illustrate the intuition underlying the model, we present a simple example. Suppose a challenger (C) and a defender (D) initially posses equal shares of a landmass of size and value equal to 1. The challenger may attempt to forcibly seize an additional sliver of size $\delta = \frac{1}{100}$; that is, to increase her holdings to $\frac{1}{2} + \delta$. The defender can only prevent the seizure with a war, in which the victor takes control of the entire landmass. If the defender allows the seizure, the challenger may then attempt to initiate war in an attempt to take control of the rest of the landmass. For the sake of intuition, imagine each state's share is its respective homeland, and the sliver desired by the challenger is a mostly barren hilltop on the defender's border with some strategic value in a war.

The direct value of possessing a given share of the landmass is common and equal to its size. The probability of victory in a war also depends on the prevailing division, but the challenger begins with a small edge. Specifically, the challenger's initial probability of victory is $p_q = \frac{1}{2} + \frac{1}{1{,}000}$, and should she successfully seize the hilltop becomes $p_r = \left(\frac{1}{2} + \delta\right) + \frac{1}{1{,}000}$. Since the probability of victory in a war increases by δ with seizure of the hilltop, this quantity is both the hilltop's *direct* and *military* value.

The defender's cost of war is commonly known to be $c = \frac{1}{4}$. The challenger's cost of war is private information, and can either be "normal" ($\theta_C = c = \frac{1}{4}$) with probability $\frac{9{,}999}{10{,}000}$ or low ($\theta_C = 0$) with probability $\frac{1}{10{,}000}$. The low-cost type is willing to wage war to correct any imbalance between her share of the landmass and her military strength, and would therefore be willing to wage war after a concession in an attempt to seize the entire landmass. Ex-ante, however, the defender is almost certain that the challenger is not the low-cost belligerent type, i.e. $P(\theta_C = 0) = \frac{1}{10{,}000}$.

Suppose that the challenger had the option to immediately seize the entire landmass.

Clearly the defender would be willing to respond with war rather than appeasement. Such a seizure would leave the defender with nothing. A war, although costly, would still have a positive payoff in expectation.

The central issue, however, is whether it is reasonable for the defender to respond to an attempted seizure of the barren hilltop with full-scale war, which would deter any challenger who is a normal type and wishes to avoid war. Intuition suggests that he would not; were the defender to respond with war he would suffer an immediate cost of $\frac{1}{4}$, but only prevent a trivial change of $\delta = \frac{1}{100}$ in direct value and in the military balance. This tiny change in the military balance appears particularly irrelevant because the probability that the challenger is belligerent is $\frac{1}{10{,}000}$. If this intuition were true, the challenger would always seize the hilltop regardless of her type because she expects to make a gain without provoking war.

Nevertheless, this intuition is false; the mere presence of fear that the challenger is a low-cost type is sufficient to sustain credible deterrence. By attempting to seize the hilltop while anticipating that this action will result in war, the challenger reveals herself to be the low-cost belligerent type that will eventually go to war over the entire landmass. To see this, suppose that both sides believe that an attempted seizure of the hilltop will trigger war. Then the normal type of challenger will be deterred, since the tiny imbalance of $\frac{1}{1{,}000}$ between her initial holdings and her initial probability of victory is dwarfed by the cost of war $\frac{1}{4}$. Only the low-cost type will seize since she prefers war to the status quo.

Understanding this, the defender will infer that a challenger who seizes is a low-cost belligerent type. Although he'd still prefer to allow seizure if it would successfully appease the challenger, he also knows that appeasement is impossible because gaining the hilltop will increase the challenger's payoff for war as much as it increases her payoff for peace. He will thus infer the inevitability of war from an attempted seizure and will respond preemptively. Anticipating war, it is strictly better for him to go to war before the hilltop is seized rather than after, since otherwise the challenger would gain a very slight military advantage δ. Since responding to a seizure with war is optimal, the normal-cost type of challenger will indeed be deterred, choosing to end the game peacefully rather than provoke a war.

Notice that the defender's ability to infer that the challenger is belligerent is conditional on his own strategy. If the challenger believed that the defender would permit seizure of the hilltop, then all types of challenger would opportunistically seize, and the defender would learn nothing about the challenger's type. This points to the fact that the example has multiple equilibria. When all challengers seize, permitting seizure is optimal given the low ex-ante probability that the challenger is belligerent. The challenger's decision to transgress only reveals her belligerence when she believes that the defender will respond with war, and war is an optimal response only when this belligerence is revealed.

4 The Model

The model is a simple two-period game of aggression and deterrence played between a potential challenger (C) and a defender (D).

Sequence In the first period, the challenger chooses whether or not to attempt a transgression $x^1 \in \{a, \emptyset\}$ that has both *direct value* to her in the event that peace prevails, and *military value* in the event that war breaks out. The transgression could represent any number of prohibited actions that would shift the military balance toward the challenger, but also benefit her if her intentions vis-a-vis the defender were ultimately peaceful; it therefore presents the defender with an inference problem about the challenger's true intentions. Such actions could include occupying territory belonging to the defender or a protégé, enacting sanctions, or developing valuable scientific technology that could be weaponized like nuclear capability.[19] The challenger's attempt to transgress is observable to the defender, and thus could also be interpreted as making a demand of the defender to allow it.

If the challenger does not attempt to transgress ($x^1 = \emptyset$), then the game ends with peace. If she does ($x^1 = a$), then the defender may either allow the transgression ($y^1 = n$) or resist it ($y^1 = w$). To make credible deterrence as difficult as possible, we assume that the challenger's act presents the defender with a fait accompli; to resist the transgression means war.

If the defender allows the challenger to transgress, then the game proceeds to a second period. In the second period, the challenger's payoffs are assumed to be higher in the event of either peace or war as a result of having successfully transgressed, and the defender's are assumed to be lower. The challenger then decides whether to enjoy her direct gains and end the game peacefully ($x^2 = n$), or herself initiate war under the more favorable military balance ($x^2 = w$). The sequence of the game is depicted in Figure 1.

[Figure 1 about here.]

Defender's Payoffs Unlike many analyses of deterrence credibility, where the defender's payoffs are uncertain to the challenger, we assume that the defender's payoffs are common knowledge. Moreover, he has a known preference for appeasement. To capture that preference we denote the defender's payoff as n_D^t if the game ends with peace in period t and w_D^t if the game ends with war in period t, and assume that:

1. allowing the transgression makes him worse off in both peace ($n_D^2 < n_D^1$) and war

[19] According to Slantchev 2011, military investments are intrinsically costly rather than beneficial to a challenger with peaceful intentions, and therefore do not satisfy our assumptions. (An additional distinction with our model is that the defender has no opportunity to preempt a shift in military power). However, military investments could also satisfy our assumptions if they yielded benefits in interactions with internal political rivals or states other than the defender, as in Baliga and Sjostrom 2008.

$(w_D^2 < w_D^1)$,

2. allowing the transgression is strictly better than responding with war if the challenger will subsequently choose peace $(n_D^2 > w_D^1)$.

Given these assumptions, the defender's optimal response to a transgression depends on his interim assessment of the probability β that a challenger who has attempted to transgress will initiate war even after being allowed to do so. If war is truly inevitable, then he prefers to avoid the cost $w_D^1 - w_D^2 > 0$ of allowing an unappeasably belligerent challenger to transgress; this cost captures the (potentially small) endogenous shift in military power that results. However, if allowing the transgression would actually appease the challenger, then he prefers to do so and avoid the cost $n_D^2 - w_D^1$ of a preventable war. It is simple to show that the defender will prefer to respond to the transgression with war whenever his interim belief β exceeds a threshold $\bar{\beta}$, where

$$\bar{\beta} = \frac{n_D^2 - w_D^1}{(n_D^2 - w_D^1) + (w_D^1 - w_D^2)} \in (0, 1)\,. \tag{1}$$

Crucially for our argument, $\bar{\beta} < 1$ – that is, if war is truly inevitable, then the defender prefers war sooner to war later regardless of its cost.

The defender's dilemma in our model is thus closely related to Powell's analysis of "salami tactics," a relationship we further explore in Section 6.[20] The defender is vulnerable to exploitation by the challenger because a small transgression is below his known threshold for war. However, his fear that the challenger's intentions may in fact be far reaching, and his preference for war sooner rather than war later if it is to be inevitable, may sometimes lead him to respond with war.

[20]Powell 1996.

Challenger's Payoffs Because the defender never intrinsically prefers to fight to prevent the transgression, the key factor sustaining his willingness to do so must be the possibility that the challenger seeks to transgress to strengthen herself for a future war. To model this possibility, we assume that the challenger has fixed and known payoffs n_C^t for peace in each period, but her payoffs from war $w_C^t(\theta_C)$ depend on a *type* $\theta_C \in \Theta \subset \mathbb{R}$ drawn by "nature" at the start of the game, where Θ is an interval. The probability distribution over the challenger's type is described by a continuous distribution $f(\theta_C)$ with full support over Θ, and the challenger's payoffs n_C^t and $w_C^t(\theta_C)$ satisfy the following:

1. Successfully transgressing has a *direct value* if the game ends in peace ($n_C^2 - n_C^1 > 0$) and a *military value* if the game ends in war ($w_C^2(\theta_C) - w_C^1(\theta_C) > 0 \ \forall \theta_C \in \Theta$),

2. In each period t the challenger's war payoff $w_C^t(\theta_C)$ is continuous and strictly increasing in θ_C. In addition, there exists a unique challenger type $\bar{\theta}_C^t$ strictly interior to Θ that is *indifferent* between peace and war (i.e. $w_C^t\left(\bar{\theta}_C^t\right) = n_C^t$).

Our assumptions imply the following. First, all challenger types intrinsically value the transgression, i.e., even absent a war. Second, a challenger's type θ_C indexes her willingness to fight. Third and most importantly, in each period t there is positive probability that the challenger prefers peace to war $\left(\theta_C < \bar{\theta}_C^t\right)$ and war to peace $\left(\theta_C > \bar{\theta}_C^t\right)$. Thus, there is always the possibility (however unlikely) that she prefers war to the status quo $\left(w_C^1(\theta_C) \geq n_C^1 \iff \theta_C \geq \bar{\theta}_C^1\right)$. In addition, once the challenger has successfully transgressed, there is always the possibility that the challenger is a type against whom war is inevitable; formally, these are types $\theta_C \geq \bar{\theta}_C^2$ who would unilaterally initiate war even after being allowed to transgress.

Although challenger types $\theta_C > \bar{\theta}_C^2$ are modeled as unilaterally initiating war, this outcome could also represent an unmodeled continuation game where the challenger makes an additional demand against which the defender is willing to fight. Interpreted as such, a number of rationales for the defender's willingness to fight are possible; it could once again be driven by fear that war is inevitable in a future unmodeled period, his threat over the subsequent demand could be "intrinsically" credible as in perfect deterrence theory,[21] he could fail to fully internalize the cost of war,[22] or war could result from a commitment problem.[23] Our baseline model is agnostic about the rationale so as not to confuse the main points. However, in the "salami tactics" extension of the model considered in Section 6, wars result from a mixture of fear and commitment problems.

Fear and Uncertainty To isolate how the defender's fear and uncertainty affects his ability to credibly deter, we solve and compare two variants of the model. In the first variant, the challenger's type θ_C is publicly revealed at the start; the game thus proceeds without any uncertainty about her intentions. In the second variant the challenger's type θ_C is her private information, $f(\theta_C)$ represents the defender's *prior beliefs* over the challenger's type, and the defender attempts to update those beliefs from the challenger's actions.

5 Results

We begin by characterizing equilibria of the two model variants. We then more closely examine the conditions for deterrence when the defender is uncertain about the challenger's type. Finally, we directly compare equilibria to isolate the effect of uncertainty on the probability of deterrence and the defender's welfare. All proofs are in the main Appendix.

[21]Zagare 2004.

[22]Chiozza and Goemans 2004; Jackson and Morelli 2007.

[23]Powell 2006.

Complete Information

When the challenger's type θ_C is revealed at the start of the game, the defender knows whether she would initiate war if allowed to transgress. The challenger also knows that the defender knows it. For each challenger type $\theta_C \in \Theta$, the game then yields a unique Subgame Perfect Equilibrium, which we solve by backward induction. Note that the challenger's threshold for preferring peace to war in the first period $\bar{\theta}_C^1$ could be both below or above her threshold for preferring peace to war in the second period $\bar{\theta}_C^2$, depending on the payoffs.

In the second period, the challenger will initiate war after successfully transgressing if and only if $\theta_C \geq \bar{\theta}_C^2$. Anticipating this, the defender will respond to an attempted transgression with war only when $\theta_C \geq \bar{\theta}_C^2$, understanding that war is inevitable. Otherwise, he will allow the transgression given his known preference for appeasement. This yields the following equilibrium behavior.

Lemma 1. *When the challenger's type θ_C is known to the defender, she is deterred from transgressing if and only if the payoffs are such that $\bar{\theta}_C^2 < \bar{\theta}_C^1$, and in addition $\theta_C \in [\bar{\theta}_C^2, \bar{\theta}_C^1)$. Otherwise the challenger transgresses, and either the defender allows it $\left(\theta_C < \bar{\theta}_C^2\right)$, or responds with war $\left(\theta_C \geq \bar{\theta}_C^2\right)$. The probability of deterrence is $P\left(\theta_1 < \bar{\theta}_C^1, \theta_1 \geq \bar{\theta}_C^2\right)$.*

When the challenger's type is known to the defender, the conditions for successful deterrence are thus narrow. Deterrence only occurs when the challenger initially prefers peace to war ($\theta_C < \bar{\theta}_C^1$), but would prefer war to peace *after* being allowed to transgress ($\theta_C \geq \bar{\theta}_C^2$). Only in this case is the initial threat of war both credible, and sufficient to deter the challenger. In all other cases, the challenger transgresses – either because (i) she is peaceful but knows she can get away with it ($\theta_C < \min\left\{\bar{\theta}_C^1, \bar{\theta}_C^2\right\}$), (ii) because she initially prefers war but transgressing will appease her ($\bar{\theta}_C^2 < \bar{\theta}_C^1$ and $\theta_C \in [\bar{\theta}_C^1, \bar{\theta}_C^2)$), or (iii) because she prefers war to peace in both periods ($\theta_C \geq \max\left\{\bar{\theta}_C^1, \bar{\theta}_C^2\right\}$).

Incomplete Information

When the challenger's type is unknown to the defender, he *fears* that she is a type against whom war is inevitable, and attempts to infer this through her actions. The solution concept is Perfect Bayesian Equilibrium, and equilibria are as follows.

Proposition 1. *When the challenger's type θ_C is unknown to the defender, there always exists a pure strategy equilibrium.*

1. *There exists a* ***no deterrence equilibrium,*** *in which the challenger always transgresses, and she is always permitted to do so, i.f.f.*

$$\bar{\beta} \geq P\left(\theta_C \geq \bar{\theta}_C^2\right)$$

2. *There exists a* ***deterrence equilibrium****, in which (i) the defender always responds to the transgression with war, (ii) all types $\theta_C < \bar{\theta}_C^1$ who do not initially prefer war are deterred, and (iii) the probability of deterrence is $P\left(\theta_1 < \bar{\theta}_C^1\right)$, i.f.f.*

$$\bar{\beta} \leq P\left(\theta_C \geq \bar{\theta}_C^2 \mid \theta_C \geq \bar{\theta}_C^1\right)$$

When both pure strategy equilibria exist, there also exists a mixed strategy equilibrium, but the defender is best off in the deterrence equilibrium.

A pure strategy equilibrium of the model thus always exists when the defender is uncertain about the challenger's type. Moreover, whenever a mixed strategy equilibrium exists, the defender is always better off in the pure strategy deterrence equilibrium. Because our focus is on the conditions under which the defender can achieve credible deterrence in equilibrium, we henceforth restrict attention to pure strategy equilibria.

Pure strategy equilibria of the model are of two types. The first is a "no deterrence equilibrium." The challenger always attempts to transgress, and consequently the defender can infer nothing about the challenger's type simply from observing the transgression itself. He therefore decides how to respond on the basis of his prior $P\left(\theta_C \geq \bar{\theta}_C^2\right)$ that the challenger is sufficiently belligerent to initiate war after transgressing. If that prior $P\left(\theta_C \geq \bar{\theta}_C^2\right)$ is low and/or the defender's belief threshold $\bar{\beta}$ for responding with war is high, then this equilibrium will exist. Recall that $\bar{\beta}$ is determined by the cost $n_D^2 - w_D^1$ of an avoidable war relative to the cost $w_D^1 - w_D^2$ of allowing an unappeasably belligerent challenger to transgress. These conditions accord with the conventional wisdom for when deterrence should fail – when the benefit of avoiding war is high relative to the cost of appeasement, and the challenger is very unlikely ex-ante to be belligerent.

The second type of pure strategy equilibrium is a "deterrence equilibrium." In this equilibrium the defender always responds to the transgression with war. Consequently, the challenger is deterred from transgressing unless she is initially belligerent, in the sense of preferring war to the status quo (i.e. $\theta_C \geq \bar{\theta}_C^1$). Crucially, this deterrence allows the defender to draw an inference from observing the transgression itself even if it is objectively minor – precisely that the challenger is an initially belligerent type. As a result, he decides whether to respond with war *not* on the basis of his prior $P\left(\theta_C \geq \bar{\theta}_C^2\right)$ that the challenger will initiate war after transgressing, but his *posterior* $P\left(\theta_C \geq \bar{\theta}_C^2 \mid \theta_C \geq \bar{\theta}_C^1\right)$ that allowing an *already belligerent* challenger to transgress will fail to appease her.

This exceedingly simple observation is in fact our key insight. In the presence of *fear* that war may be inevitable, the primary factor determining the defender's ability to credibly deter *in equilibrium* is not the cost of war, the severity of the transgression, or the initial probability that the challenge is belligerent. The reason is that when deterrence is actually effective, the defender can already infer the challenger's initial belligerence from the transgression itself. Instead, the primary factor is actually the *effectiveness of appeasement against an already belligerent challenger.*

The implications of this simple insight are surprisingly strong, as illustrated in the following corollary.

Corollary 1. *When allowing the transgression cannot appease an already belligerent challenger, i.e.* $\bar{\theta}_C^2 \leq \bar{\theta}_C^1 \iff P\left(\theta_C \geq \bar{\theta}_C^2 \mid \theta_C \geq \bar{\theta}_C^1\right) = 1$*, then the deterrence equilibrium exists for all defender payoffs and probability distributions satisfying the initial assumptions.*

Thus, when appeasement is impossible against a belligerent challenger, the deterrence equilibrium always exists. This is true even when the "no deterrence equilibrium" also exists because of a high cost of war $(n_D^1 - w_D^1)$, a low cost of allowing the transgression in both direct $(n_D^1 - n_D^2)$ and military $(w_D^1 - w_D^2)$ terms, and/or a sufficiently low probability that the challenger is belligerent $P\left(\theta_C \geq \bar{\theta}_C^t\right)$ in both periods.[24] The deterrence equilibrium remains because the defender can use the transgression (however minor) as a *test* of the challenger's initial belligerence, knows that initial belligerence ensures future belligerence because appeasement is ineffective, and therefore prefers to respond with war upon observing the transgression. The challenger is thereby deterred unless she affirmatively prefers immediate war, fulfilling the defender's expectations.[25]

Figure 2 depicts the equilibrium correspondence for an example in which the defender's belief threshold $\bar{\beta}$ for responding with war is very high. The defender's prior $P\left(\theta_C \geq \bar{\theta}_C^1\right)$

[24]The irrelevance of the cost of war for Corollary 1 partially depends on a literal interpretation of the second period. If it is instead interpreted as an unmodeled continuation game where the challenger attempts a transgression against which the defender is willing to fight, then an additional weak condition on the cost of war is necessary. We address this point more fully in the Robustness section.

[25]Baliga and Sjostrom 2008 can also exhibit a qualitatively similar separating equilibrium when their assumption (3) fails and the "crazy type" values nuclear weapons sufficiently highly. However, there are also important differences. Because we micro-found the defender's willingness to attack a crazy type in the preference for war sooner vs. war later, she will do so in the equilibrium regardless of her war payoffs; our model also yields new comparative statics on when such an equilibrium will prevail.

that the challenger prefers immediate war to the status quo is on the x-axis, while the prior $P\left(\theta_C \geq \bar{\theta}_C^2\right)$ that she would initiate war after transgressing is on the y-axis; both quantities are derived from the challenger's underlying payoffs and the distribution over her type. The figure demonstrates that the deterrence equilibrium can remain even when the probabilities that the challenger would be belligerent in either period are arbitrarily low, which can be seen by observing that the hatched triangle extends to the origin. Moreover, this property would persist even if the defender's threshold for war $\bar{\beta}$ were made arbitrarily high.

[Figure 2 about here.]

The (In)effectiveness of Appeasement

The preceding analysis demonstrates that in the presence of fear and uncertainty about the challenger's intentions, the effectiveness of appeasement and the credibility of deterrence are really two sides of the same coin. Deterrence can be credible if appeasement would be ineffective against a belligerent challenger, even if the ex-ante probability of that belligerence is very low. Conversely, if appeasement is effective then deterrence can be undermined even if the ex-ante probability that the challenger is belligerent is high.

We now consider the question of what makes appeasement less effective, and consequently deterrence more effective. To answer this question we examine the payoff properties of the transgression itself. Recall that transgressing has both a military value $w_C^2(\theta_C) - w_C^1(\theta_C)$, which is the challenger's gain in the event of war, and a direct value $n_C^2 - n_C^1$, which is her gain in the event of peace. These quantities determine how the challenger's preference for war changes as a result of successfully transgressing. We henceforth denote them using $\delta_C^m(\theta_C)$ and δ_C^d, respectively.

We begin with a simple condition that does not require any additional assumptions.

Corollary 2. *Appeasement is ineffective, and thus the deterrence equilibrium exists when the defender is uncertain about the challenger's type, if and only if* $\delta_C^m\left(\bar{\theta}_C^1\right) \geq \delta_C^d$.

Thus, a sufficient condition for the deterrence equilibrium to exist is that military value of the transgression $\delta_C^m(\cdot)$ exceed its direct value δ_C^d to a challenger of type $\bar{\theta}_C^1$ who is initially indifferent between peace and war. The logic is straightforward. For such a challenger type, $\delta_C^m\left(\bar{\theta}_C^1\right) \geq \delta_C^m$ means that the military gains from successfully transgressing increase her net benefit from war as much as the direct gains from transgressing reduce it. Since she initially weakly preferred war to peace, she and all types more belligerent than her must continue to prefer war to peace after transgressing. Allowing the transgression therefore cannot appease any type of challenger who was initially belligerent (i.e. $P\left(\theta_C \geq \bar{\theta}_C^2 \mid \theta_C \geq \bar{\theta}_C^1\right) = 1$), which by Corollary 1 implies that the deterrence equilibrium exists.

The condition in Corollary 2 turns out to be familiar from the literature examining complete information bargaining in the presence of endogenous shifts in military power.[26] To our knowledge, however, it is absent from the literature (either empirical or theoretical) on deterrence. The bargaining literature finds that similar conditions generally result in wars or the gradual elimination of one player. In contrast, we find that this same condition can lead to a fearful peace with deterrence of even a very minor transgression with very high probability. Both predictions are rooted in the same property; allowing the transgression cannot appease a belligerent challenger. However, the distinction arises from the difference in assumptions about whether the challenger is initially belligerent. In the complete information setting, the challenger's belligerence at the outset is assumed. In our model, the defender can believe that the challenger is very likely to be peaceful ex-ante; however, his *fear* that the challenger is unappeasably belligerent – however small – allows him to credibly deter.

[26] Fearon 1996; Schwarz and Sonin 2008.

Comparative Statics When credible deterrence is possible but difficult – either because the defender's threshold $\bar{\beta}$ for war is high and/or the probability the challenger will be belligerent $P\left(\theta_C \geq \bar{\theta}_C^2\right)$ is low – the model generally exhibits multiple equilibria. This complicates extracting testable comparative statics because the model cannot speak to how the players will form expectations about whether the transgression is being treated as a "test" – it can only say when doing so would be an equilibrium.

Nevertheless, the consequences of our basic insight – that the effectiveness of appeasement influences the credibility of deterrence – can be examined by deriving comparative statics on the probability of deterrence under the assumption that the deterrence equilibrium prevails whenever it exists. With this assumption, the probability that deterrence is successful is 0 when the deterrence equilibrium does not exist, and is $P\left(\theta_C < \bar{\theta}_C^1\right)$ when it does. The following Proposition considers this comparative static as a function of the challenger's payoffs, while holding the defender's payoffs fixed.

Proposition 2. *Suppose that,*

- *the deterrence equilibrium prevails whenever it exists,*
- *the transgression's military value is equal to δ_C^m for all challenger types,*
- *the challenger's first period payoffs are held fixed.*

Then appeasement is less effective (i.e. $P\left(\theta_C \geq \bar{\theta}_C^2 \mid \theta_C \geq \bar{\theta}_C^1\right)$ is increasing) the greater is the difference $\delta_C^m - \delta_C^d$ between the challenger's military and direct value for the transgression. Consequently, the probability that deterrence is successful is increasing in $\delta_C^m - \delta_C^d$.

With our additional assumptions, the probability of deterrence is increasing in the difference $\delta_C^m - \delta_C^d$ between the military and direct value of the transgression to the challenger. The intuition is similar to Corollary 2; the greater is $\delta_C^m - \delta_C^d$, the more likely it is that appeasement will fail against a belligerent challenger, the more willing is the defender to respond with war conditional on deterrence failing, and the better able he is to deter. This effect is depicted in Figure 3; the left panel shows the probability of deterrence when the defender's payoffs are fixed, while the right panel depicts the probability when the defender's payoffs are initially drawn from a distribution.[27]

[Figure 3 about here.]

Proposition 2 has surprising and counterintuitive implications for empirical studies of deterrence failure: it predicts that deterrence should be most effective against transgressions that carry a high military value *relative to* their direct value for a challenger. This result helps to explain cases of deterrence success when the immediate stakes do not appear to be worth fighting for; the low intrinsic value of the stakes may actually contribute to the effectiveness of deterrence. For example, in 1958 the United States successfully deterred a Chinese invasion of the tiny island of Quemoy with threats of war.[28] Precisely because the island had marginal direct value, it could not possibly appease a China intent on war. Therefore, an invasion could have easily been perceived as an informative signal of both the present and future belligerence of the Chinese Communists.

Interpreted in the context of Proposition 2, the Allies' difficulty in deterring Hitler from annexing Austria and invading the Sudetenland prior to World War II is also easier to under-

[27]Specifically, the right panel depicts $G\left(\frac{1-F\left(\bar{\theta}_C^2\right)}{1-F\left(\bar{\theta}_C^1\right)}\right) \cdot F\left(\bar{\theta}_C^1\right)$, where $G\left(\bar{\beta}\right)$ is the induced probability distribution over $\bar{\beta}$. To generate both figures we assume that $w_C^1\left(\theta_C\right) = \theta_C$, the transgression's military and direct cost to the defender's are equal to .1, and the challenger's type is uniformly distributed on $[0, 1]$. The left panel assumes that the defender's benefit from peace is $n_D^1 - w_D^1 = .55$, while the right panel assumes it is uniformly distributed on $[-.1, 1]$.

[28]Soman 2000, 181.

stand. While these territories were presumably of significantly greater intrinsic value to the Allies than Quemoy was to the United States, they were also densely populated by German co-ethnics. Consequently, the notion that occupying them might satisfy a belligerent Germany was actually quite plausible, resulting in exploitation of the Allies' known preference for appeasement and deterrence failure.

An additional complication is that the comparative static in Proposition 2 varies the challenger's values for transgressing while holding those of the defender fixed. However, in many applications it is reasonable to suppose that a transgression with greater direct or military value for the challenger is also one that imposes greater direct or military costs on the defender. This relationship is important for making empirical predictions about deterrence success because, while allowing a transgression with a higher direct value might more effectively appease, it is also more intrinsically worth fighting over. Figure 4 illustrates the probability of deterrence in a numerical example where the values to the challenger for transgressing are equal to the costs imposed on the defender. In the example, the probability of deterrence is always increasing in the transgression's military value (on the x-axis). However, increasing the transgression's direct value (on the y-axis) has a non-monotonic effect; the probability of deterrence first decreases due to the logic of Proposition 2, and then increases as the defender's intrinsic willingness to fight dominates.[29]

[Figure 4 about here.]

Proposition 2 and the preceding example jointly demonstrate the importance of distinguishing the military from the direct value of a transgression in empirical analyses of deterrence, a distinction that has been largely ignored.[30] Such studies are generally moti-

[29]The figure is generated by assuming that $w_C^1(\theta_C) = \theta_C$ with $\theta_C \sim U[0, 1.1]$, and both the defender's benefit $n_D^1 - w_D^1$ and lowest type of challenger's benefit $n_C^1 - w_C^1(0)$ for avoiding war is .5.

[30]See Huth 1999 for a summary.

vated by the premise that the absolute magnitude of intrinsic values alone determine states' willingness to carry out their threats. However, our model demonstrates that an equally important factor is states' expectations and inferences about their adversaries' future behavior – for this question, the *relationship* between the military and direct value of a transgression is essential.

The Benefits of Fear

The preceding analysis demonstrates that counterintuitively, it can be the defender's *fear* – rather than an intrinsic willingness to fight – that allows him to credibly deter. This property suggests that the defender may actually benefit in expectation from his fear and uncertainty. Our last result examines this claim by directly comparing equilibrium outcomes and payoffs in the game with uncertainty to the game with complete information.

Proposition 3. *Suppose that the deterrence equilibrium prevails whenever it exists. Then,*

1. *the probability of deterrence would decrease if the defender knew the challenger's type.*
2. *when the probability* $P\left(\theta_C < \bar{\theta}_C^2 \,|\, \theta_C \geq \bar{\theta}_C^1\right)$ *that appeasement is effective is below*

$$\left(\frac{P\left(\theta_C < \bar{\theta}_C^1\right)}{P\left(\theta_C \geq \bar{\theta}_C^1\right)}\right) \cdot \left(\frac{n_D^1 - n_D^2}{n_D^2 - w_D^1}\right),$$

the defender is better off in expectation not knowing the challenger's type.

Proposition 3 first shows that the probability of deterrence decreases when the challenger's type is known to the defender. Thus, it is specifically the defender's fear, and not just the possibility that the challenger may be unappeasably belligerent, that generates credible deterrence. To see why this is the case, suppose for simplicity that appeasement is completely ineffective, i.e., $\bar{\theta}_C^2 \leq \bar{\theta}_C^1$. If the defender knew the challenger's type, then she would be unable to deter her from transgressing whenever that type was outside of $\left[\bar{\theta}_C^2, \bar{\theta}_C^1\right)$. When $\theta_C \geq \bar{\theta}_C^1$ the challenger would be unappeasably belligerent and transgress, while when $\theta_C < \bar{\theta}_C^2$ it would be commonly known that the challenger would be peaceful after transgressing, and she would exploit the defender's known preference for appeasement. However, if the defender is uncertain of the challenger's type, then he can credibly maintain fear that a challenger who is in fact appeasable $\left(\theta_C < \bar{\theta}_C^2\right)$ may be unappeasably belligerent $\left(\theta_C \geq \bar{\theta}_C^1\right)$, infer that unappeasable belligerence from the transgression in equilibrium, be willing to respond with war, and deter all types $\theta_C < \bar{\theta}_C^1$ from transgressing.

This insight suggests that events creating uncertainty about the intentions of a challenger can have important and surprising effects on the potential for credible deterrence. For example, a sudden regime change in a potential adversary may create fear on the part of a defender about that adversary's intentions that did not previously exist; understanding that fear and the potential inference the defender could draw from a transgression, the adversary could paradoxically be easier to deter.

Finally, it is indeed the case that the defender's uncertainty can sometimes benefit her in expectation. The second part of the Proposition states that this will be the case when the probability $P\left(\theta_C < \bar{\theta}_C^2 \,|\, \theta_C \geq \bar{\theta}_C^1\right)$ that appeasement would work against a belligerent challenger is sufficiently low. The reason is that it is then unlikely that the defender's uncertainty will result in a preventable war. Moreover, when a belligerent challenger cannot be appeased $\left(\bar{\theta}_C^2 < \bar{\theta}_C^1\right)$, the defender is unambiguously better off not knowing her type.

Counterintuitively then, the defender's fear can actually be source of strength. Thus, it is not surprising that states often claim to fear the unappeasable belligerence of their adversaries. For example, in 2003, North Korea warned that a U.S. air strike against their nuclear plant would lead to "total war." They accompanied this warning by explicitly stating that such an attack would be viewed as a precursor to an invasion.[31] This insight suggests interesting avenues for future work to which we return in the conclusion.

6 Robustness

We now briefly discuss the robustness of our basic insight that fear can generate credible deterrence to a variety of common complexities studied in the international relations literature. We examine the following modifications to our model: 1) additional potential transgressions by the challenger that allow for "salami tactics," 2) an endogenous transgression by the challenger, 3) an option for the challenger to back down, 4) interdependent values for war, and 5) uncertainty about the defender's willingness to fight. The robustness checks are formally conducted in a Supplemental Appendix and briefly described here.

Salami Tactics The defender's dilemma in the first period resembles Powell's game of "salami tactics"; the defender is vulnerable to exploitation because a small transgression is below his known threshold for war, but his fear of what the challenger may do after transgressing can lead him to fight.[32] However, our second period is distinct; it is simply a unilateral decision by the challenger to fight or end the game. We now explore the robustness of our insights to the possibility that this stage represents an unmodeled continuation game of salami tactics, in which the challenger continues attempting transgressions, and in each period the defender makes the final decision of whether or not to fight.

Salami Tactics Game Suppose a challenger and a defender jointly occupy a landmass of size and value equal to 1. Say the **advantaged** party at time t is that which holds a

[31]KCNA News Agency 2003.

[32]Powell 1996.

majority of the landmass, and let δ_t denote the **excess** holdings of the advantaged party in period t above $\frac{1}{2}$. If a war occurs in period t, the probability the advantaged party wins is:

$$p(\delta_t) = \left(\frac{1}{2} + \delta_t\right) + \phi(\delta_t)$$

where $\phi(\delta_t) = \frac{2\delta_t(1-2\delta_t)}{Z}$ and Z is very large. So the advantaged party has a military strength that **exceeds** her share of the landmass by a very small amount $\phi(\delta_t)$.[33] Also suppose that the defender's cost of war is commonly known to be $c_D \geq \frac{1}{4}$. The challenger's type θ_C is unknown and uniformly distributed over $\theta_C \sim U\left[-\frac{1}{4}, 0\right]$, and her cost of war is $c_C = -\theta_C$.

Now consider a $T \geq 3$ period game, and a $T+1$-length series of increasing values $0 < \delta_1 ... < \delta_T = \frac{1}{2}$. Each δ_t represents the challenger's excess holdings above $\frac{1}{2}$ if the game advances to period t. Assume the challenger is initially advantaged ($\delta_1 > 0$), and that the increments of advancement $\delta_t - \delta_{t-1}$ are less than the defender's cost of war c_D for all $t < T$. In each period t, the challenger decides whether or not to attempt to advance from δ_t to δ_{t+1}. If she doesn't attempt it, the game ends. If she does, the defender chooses whether or not to respond with war, and the challenger's probability of victory is $p(\delta_t)$. ■

The challenger's probability of victory in a war as a function of her position is depicted in Figure 5. The salami tactics game captures the essential features of the baseline model. First, in each period the challenger has a military advantage $\phi(\delta_t) = \frac{2\delta_t(1-2\delta_t)}{Z}$ that exceeds (by a tiny amount) her share of the landmass; there is thus some very small chance that she prefers war to peace $\left(c_C \leq \frac{2\delta_t(1-2\delta_T)}{Z}\right)$. Second, the probability of victory increases faster than the challenger's excess holdings in the interval $\delta_t \in \left[0, \frac{1}{4}\right]$, so there is a region of advancement where the military value of each transgression exceeds its direct value.

[[Figure 5 about here.]]

[33] We require at least $Z > 6$ for $p(\delta_t)$ to be strictly increasing in δ_t.

In the Supplemental Appendix we prove the following result.

Proposition 4. *If* $c_D \in \left[\frac{1}{4}, \frac{1}{2}\right)$*, then for any* t^* *such that* $c_D < \frac{1}{2} - \delta_{t^*} \iff \delta_{t^*} < \frac{1}{2} - c_D$*, there exists an equilibrium in which all types of challengers advance to* δ_{t^*}*, only challengers with cost* $c_C \leq \phi(\delta_{t^*})$ *attempt to advance to* δ_{t^*+1}*, and the defender always responds with war. If* $c_D \geq \frac{1}{2}$ *then in any equilibrium, the challenger occupies the entire landmass.*

Thus, as long as the cost of war c_D is below the defender's entire initial share of the landmass $\frac{1}{2} - \delta_1$, there are many equilibria sustained by the same logic as our deterrence equilibrium. There is a final threshold at which the defender is "intrinsically" willing to fight due to a commitment problem similar to Powell's game of salami tactics with complete information; he anticipates that if the challenger advances beyond it, she will be unable to commit *not* to exploit salami tactics to eventually possess the entire landmass.[34]

At *any* threshold prior to this point, there exists an equilibrium in which the defender is willing to fight due to the logic of our model, even if the size of the transgression $\delta_{t^*+1} - \delta_{t^*}$ is infinitesimal relative to the cost of war c_D. At an equilibrium "red line" δ_{t^*}, the challenger expects the defender to respond to further advancement (however small) with war, so the defender can infer in equilibrium that a challenger who attempts to advance beyond δ_{t^*} in fact desires war in period t^*. Because the challenger's probability of victory $p(\delta_t)$ is such that advancing makes war *relatively* more attractive, the probability of appeasing an already-belligerent challenger by allowing further advancement beyond the red line is 0. Hence, responding with war is optimal.

Only when the cost of war c_D exceeds the defender's *entire* initial holdings $\frac{1}{2} - \delta_1$ do these deterrence equilibria collapse. In other words, if the defender is to have the final decision to fight, then our mechanism naturally requires the minimal condition that he prefer fighting

[34] Powell 1996, 2006.

a war to giving away his entire homeland up-front, and ceasing to exist.

An Endogenous Transgression In our baseline model, the magnitude of the challenger's transgression is exogenous. However, in many settings this is the challenger's choice; we therefore consider whether our basic insight is robust to leaving it up to her. To do so, we revisit the special payoff environment above, but consider an alternative game form. In the first period the challenger can make an endogenous "demand" δ_2 of how far to advance. Otherwise, the game form is identical to the baseline model.

In the Supplemental Appendix, we show that in this extension there exists an equilibrium in which the defender responds to any strictly positive transgression, however small, with war. The rationale is again similar to the baseline model. Even when the challenger can moderate her transgression, the deterrence equilibrium remains because it is still true that conceding to most transgressions would increase an already-belligerent challenger's payoff from war more than her payoff from peace. Only the largest of transgressions can potentially sate a belligerent challenger's thirst for war, and against such transgressions the defender intrinsically prefers to fight.

A Challenger Who Can Back Down In our baseline model the challenger presents the defender with a fait accompli; to resist the transgression means war. However, in many crisis bargaining models the defender's resistance to initial aggression does not result in immediate war; instead, the challenger has an opportunity to first back down.[35] We therefore consider how the deterrence equilibrium is impacted if the game sequence is modified to incorporate this possibility.

In the Supplemental Appendix we show that this alteration actually makes the deter-

[35]For example, see Lewis and Schultz 2003.

rence equilibrium of our model easier to sustain. Formally, whenever deterrence works in the baseline model, it also works in this variant. The reason is that the defender can entertain the possibility that the challenger is actually bluffing when he observes an attempted transgression, which makes him only more willing to resist. Indeed, in the motivating case where the challenger is very unlikely to be belligerent, justifying a defender's willingness to resist is easy; he simply expects the challenger to back down.

Interdependent War Values Our baseline model assumes that the defender's payoffs are fixed and independent of the challenger's type, as would be the case if he was uncertain about the challenger's cost of war, or her private valuation for certain war outcomes.[36] However, in many models of crisis bargaining a challenger also has private information about factors that would affect both parties' payoffs from war, such as the probability of victory. We therefore consider how introducing such private information would affect our results.[37]

This possibility naturally complicates the defender's inference problem. Intuitively, if the challenger has private information about her probability of victory, then upon observing a transgression the defender can simultaneously infer that appeasement is less likely to work – making him more willing to fight – and that he would be weak in a war – making him less willing to fight. Nevertheless, Corollaries 1 and 2 continue to hold unaltered – the deterrence equilibrium always exists whenever appeasement is impossible. However, when this is not the case, the equilibrium correspondence is more complex than in the baseline mode, and can exhibit new and interesting patterns. Such a model could be an interesting avenue for

[36]For examples of models with uncertainty about the cost of war, see Fearon 1995 or Sechser 2010. Glaser 1997 addresses uncertainty about a challenger's value for possessing a defender's homeland in his discussion of "greedy states."

[37]Fey and Ramsay 2011 explore the distinction between private and interdependent war values in a mechanism design context.

future work.

A Defender with Private Information Finally, our baseline model assumes away any private information possessed by the defender about her "resolve," in order to shift attention away from reputation to fear. However, since such private information is a common feature of most international crises, we verify that our main insights are robust to introducing it. In the Supplemental Appendix we formally consider a variant of the model in which the challenger is also uncertain about the defender's intrinsic willingness to fight over the transgression.[38]

In this extension we again find that Corollaries 1 and 2 hold unaltered. Interestingly, we also find that introducing even a small possibility that the defender is "intrinsically" willing to fight can sometimes uniquely select this equilibrium. Intuitively, the reason is that "deterrence begets deterrence" – more deterrence increases the defender's interim assessment that a challenger who transgresses is unappeasable, which makes him more willing to respond with war, generates a higher probability that the transgression will provoke him, and thereby results in yet more deterrence. Examining such "deterrence spirals" would also be interesting to explore in future work.

7 The Turkish Straits Crisis of 1946

To demonstrate that the logic we have identified sheds light on puzzling historical episodes, we examine the crisis over control of the Turkish Straits that occurred between the United States and the Soviet Union in the early days of the Cold War. In 1945 and 1946, the Soviet Union repeatedly demanded that Turkey allow it to place bases on the Turkish Straits.[39]

[38]Introducing such private information also requires augmenting the first period with an option for the defender to initiate war even if the challenger chooses not to transgress.

[39]Kuniholm 1980.

These demands, coupled with extensive Soviet military preparations in the Balkans, led American officials to prepare for armed aggression against Turkey. In August 1946, President Truman decided that the United States would fight a full-scale war to defend Turkey in the event of Soviet invasion. Although this commitment was never announced publicly, it was communicated to Stalin through multiple channels. Upon learning about Truman's decision, Stalin reversed course.[40]

In this section, we argue that the model of deterrence in this paper helps to explain why the United States was willing to fight a major war to defend Turkey, and why the Soviet Union found this commitment credible. First, we demonstrate that the incentives facing the United States were similar to those facing the defender in the model. Then, we argue that the willingness of the United States to defend Turkey, and the Soviet belief in the credibility of this commitment, was due to a logic similar to that in the model's deterrence equilibrium. Finally, we argue that other deterrence mechanisms fail to explain key features of the crisis.

Incentives The Turkish Straits Crisis contained the elements that are essential to our model: the defender's unwillingness to fight for the immediate stakes, the fear and uncertainty about the challenger's intentions, and the defender's preference to fight sooner rather than later. The United States government had limited economic and political ties to Turkey and attached little intrinsic value to the Turkish Straits or Turkish independence.[41] The Soviet Union, on the other hand, did have an appreciable direct interest in controlling the Straits. It had long sought military control of the Turkish Straits for the defensive purposes

[40] Mark 2005, 123-124.

[41] The U.S. had no obvious direct economic or political interests other than a very small trade in tobacco, machinery and vehicles. See Kuniholm 1980, 65-66. Later, the Truman administration began to claim an interest Turkey's democratization, but this was mere rhetoric to justify the alliance to the U.S. public. See Kayaoglu 2009.

of protecting trade and denying the use of the Straits by hostile powers, objectives that American officials actually showed sympathy with.[42]

Furthermore, decision makers anticipated that any war with the Soviet Union would be enormously costly, despite the U.S. monopoly in atomic weapons. The military anticipated that the Red Army would quickly overrun Western Europe and would launch offensives in the Middle East and Asia. Military planners expected that bombing would to be slow to produce results due to a lack of atomic bombs in the U.S. arsenal, a lack of usable airbases, and the Soviet Union's great size. They believed that ground operations would be necessary and that the U.S. may eventually have to invade the Soviet Union through Southern Europe in order to achieve victory.[43]

This combination of factors makes the United States' decision to fight for Turkey puzzling. In fact, the U.S. had made no plans for Turkey's defense in 1945 despite the belief that a Soviet attack was likely.[44] This only changed in 1946, when the United States government began to fear a general war with the Soviet Union, and to recognize to Turkey as a military asset in the event of such a war.

Fear By 1946, American officials began to fear that the Soviet Union would initiate another world war. Officials had come to believe that, as a result of their ideology and sense of insecurity, the Soviet Union was inherently expansionist and desired to dominate the Eurasian continent and eventually the world.[45] These ambitions did not necessarily imply that war would occur: most intelligence and military assessments assumed that the Soviet

[42] DeLuca 1977, 511-514; *Foreign Relations of the United States (FRUS) 1946 VII*, 827-829; Leffler 1985, 809-810.

[43] Ross 1996, 12-19, 31. Also see Mark 1997, 393 and Ross and Rosenberg 1989.

[44] See Mark 1997, 389 and Trachtenberg 1999, 38.

[45] Ross 1996, 3, 7; Gaddis 2000, 284.

Union was practical enough to avoid a destructive war with the United States and would accept the prevailing status quo.[46] However, officials could not be perfectly confident in this assessment, and they entertained the possibility that the Soviet Union would initiate general war in pursuit of their objectives. For example, in a meeting on the subject on June 12, 1946, President Truman speculated that the Soviet Union might start a war to divert public unrest, Secretary of the Navy Forrestal argued they might start a war if external circumstances were favorable for completing the "world revolution," and Admiral Leahy responded that the Soviets were simply unpredictable.[47] Officials also took seriously the stated Soviet belief that war with the capitalist world was inevitable.[48]

If such a war were to occur, it is clear that the United States would have preferred that the fighting begin before losing Turkey. In the war plans, Turkey was to be the first line of defense against a Soviet advance toward strategically vital areas of the Middle East, the loss of which would weaken the U.S. and its allies. Unlike the other countries of the Middle East, Turkey had a government willing to fight against a Soviet invasion and a disciplined and well-trained army that could resist an offensive for an estimated 4-5 months. Turkish resistance would give the United States access to the Suez Canal, the Persian Gulf, and most importantly, British-controlled air bases in Egypt from which the United States planned to bomb central Russia.[49]

Finally, some officials believed that a successful transgression would not only strengthen the Soviet Union, but would increase the Soviet appetite for general war. For example, Ambassador Edwin Wilson argued that, if the Soviet Union were allowed to overrun Turkey, they would then be unable to resist the temptation to advance toward the Suez Canal and the

[46] Ross 1996, 10; Mark 1997, 397.

[47] See Mark 2005, 119, 129.

[48] Gaddis 2000, 300-301.

[49] See Ross 1996, 30-38;Leffler 1985, 814-815.

Persian Gulf. He wrote that "once this occurs, another world conflict becomes inevitable" because of the military advantages the Soviets would then have against the West.[50] This is similar to the unappeasability condition in the model: conquering Turkey would increase Soviet military strength and make general war more attractive, outweighing any pacifying effect from satisfying the Soviet demands over Turkey itself. Therefore, a concession could not possibly appease the Soviet government if it was already belligerent.

Inference and Deterrence The Turkish Straits Crisis thus contained the incentive structure and uncertainty about challenger intentions that are essential to our model. Historical accounts of the crisis also suggest that, following the U.S. decision to defend Turkey, the Truman Administration was prepared to infer far-reaching Soviet ambitions from their willingness to attack Turkey *in the face of a U.S. commitment*, which is the key inference that sustains deterrence in equilibrium.

As mentioned above, most officials believed that the Soviet Union would be deterred by a U.S. commitment because it was generally thought that they wanted to avoid a major war.[51] Conversely, officials appear to have believed that the Soviet Union would only invade Turkey if it did desire such a war. Undersecretary of State Dean Acheson said he believed that the Soviet Union would most likely be deterred, but he also argued that the United States would "learn whether the Soviet policy includes an *affirmative* provision to go to war *now*" if deterrence failed.[52] This is the key inference in the deterrence equilibrium; that the defender can learn of the challenger's preference for war from her willingness to transgress in the face of a deterrent threat. It is also clear that this inference sustained the U.S. willingness to initiate war following a Soviet attack. President Truman, when asked if he understood

[50] *FRUS 1946 VII*, 819, 822.

[51] Mark 1997, 399.

[52] Mark 1997, 400. Emphasis in original.

that the decision to defend Turkey may mean war, responded that "we might as well find out whether the Russians were bent on world conquest now as in five or ten years."[53]

It is less clear whether the Soviets were deterred because of their understanding that American decision makers would interpret invasion as evidence of an intent to initiate war. As the first postwar crisis in which the Soviet Union attempted to take control of an area where it did not already have a military presence at the end of WWII, it seems likely that Stalin would have realized that invading Turkey would appear to the Americans as a dangerous new direction in Soviet policy, and that both parties ultimately came to understand the act of invasion as focal for revealing Soviet intentions. In fact, although the invasion didn't occur, this episode dramatically reshaped American perceptions of Soviet intentions, and Soviet Foreign Minister V.M. Molotov later admitted that they had overreached in Turkey.[54]

Alternative Explanations While other deterrence mechanisms may have also been relevant, some of them fail to explain key features of the crisis. It is possible that reputational concerns were driving decision-making, and that the United States felt it had to demonstrate its resolve to its allies and the Soviet Union.[55] However, the primary fear in losing Turkey was never reputational, it was strategic. Government officials and military estimates repeatedly emphasized that the major concern in the crisis was that losing Turkey would disadvantage the United States in the event of war with the Soviet Union. Truman himself, when told that a commitment to Turkey may mean war, pulled out a map and lectured his

[53]Mills 1951, 192. This logic continued to inform thinking long after the crisis. For example, in 1949, President Inonu of Turkey told U.S. officials that the Soviet Union would not attack Turkey by itself, but worried that, if Soviet leaders held "the conviction that war with the capitalist West is inevitable," they may attack Turkey as part of a decision to initiate general war. See Inonu to Royall, *Truman Office Files*, 1979.

[54]Mark 1997, 414.

[55]See *FRUS 1946 VII*, 858, for example of this type of thinking. While the report of the Joint Chiefs of Staff mentions the possible reputational cost of losing Turkey, it heavily emphasizes the strategic cost.

advisors about the strategic importance of the Middle East.[56]

Other commonly cited mechanisms in the deterrence literature do not seem to explain the outcome of the crisis. Any explanation involving audience costs would require threats to have been made publicly. While the United States did dispatch a naval force to the Mediterranean, it never publicly announced its decision to defend Turkey, instead communicating with the Soviet Union privately and downplaying the crisis in public. The crisis ended before efforts to mobilize public opinion had begun.[57] In addition, there was no obvious commitment device that would have automatically engaged the United States in a conflict, such as military forces stationed in Turkey as a "trip-wire." Finally, there was nothing probabilistic about the Americans' threat that "left something to chance." On the contrary, Truman clearly asserted that, if the Soviet Union invaded, he would follow the recommendation to defend Turkey "to the end."[58]

8 Conclusion

This paper examines a model of deterrence in which a defender is uncertain about the intentions of a challenger, and his fear that the challenger is unappeasably belligerent can sustain credible deterrence in equilibrium. We demonstrate that this fear can generate credible deterrence even when the probability that a challenger is belligerent is arbitrarily small and the value of the transgression being deterred is small relative to the cost of fighting a war. Unlike most previous analyses of deterrence, our theory does not assume that the defender is sometimes intrinsically willing to fight, or that he has access to commitment devices that help him to do so. Instead, our mechanism relies on the inference that the

[56] Acheson 1969, 196; Jones 1964, 63-64.

[57] See Trachtenberg 2012, 24-25, for examination of this crisis with respect to audience costs.

[58] *FRUS 1946 VII*, 840.

defender can make from a transgressive act taken in the face of an expectation of war.

After illustrating this simple insight, we derive several empirical implications about when deterrence will be credible that are previously untested in the empirical literature. We show that transgressions that make effective "tests" are not ones that are objectively large, but ones that carry a high military value *relative* to their direct value; the reason is that allowing such transgressions cannot appease an already belligerent challenger. We also show that events introducing uncertainty about a challenger's intentions can increase the likelihood of deterrence, and that the defender's fear can sometimes benefit her by allowing her to credibly deter at a negligible risk of avoidable wars.

Finally, we argue that this insight helps illuminate specific historical episodes of successful deterrence, such as the Turkish Straits Crisis and the Berlin Crisis, and may shed light on other cases such as the Cuban Missile Crisis. The logic of our model can also help to explain more contemporary episodes, such as North Korea's successful deterrence of an American air strike against their nuclear plant. The government of North Korea has threatened war in response to both economic sanctions and an air strike, and the United States appears to have taken these threats seriously, despite the fact that such a war would almost certainly lead to the destruction of the North Korean regime.[59]

How could such a threat be taken as credible? The available evidence suggests that both sides understand that North Korea is using certain actions as a test of the United States' intention to invade. Pyongyang's 2003 warning that an air strike on their nuclear plant would

[59]The threat in response to economic sanctions was in 1994, and in response to an airstrike was in 2003. See Wit, Poneman and Gallucci 2004; KCNA News Agency 2003. Secretary of Defense William Perry and Assistant Secretary of State Robert Gallucci both wrote that they believed that it would be "irresponsible" to treat the 1994 threat as bluff, and that an airstrike against the nuclear plant could trigger a war. See Carter and Perry 1999; Wit, Poneman and Gallucci 2004.

lead to "total war" explicitly stated that such an attack would be viewed a precursor to an invasion.[60] Similarly, in recommending an airstrike against a North Korean missile testing site, William Perry and former Assistant Secretary of Defense Ashton Carter wrote that the United States must be careful to warn North Korea that the attack would only be against a specific target, not against their country or their military.[61] Special Envoy Jack Pritchard responded that, despite the warning, Pyongyang might very well interpret the air strike as the "start of an effort to bring down [their] regime."[62] The incentive by North Korea to claim uncertainty about the United States' ultimate intentions, as well as the incentive by the U.S. to claim sharply limited goals, both follow directly from our logic.

These incentives, however, also point to some limits in our analysis and interesting avenues for future work. Most of the previous deterrence literature focuses on things that a defender can *do* – issue cheap talk claims, engage in costly signalling, employ commitment devices, etc. – to improve the credibility of his deterrent threats. Our analysis is different. We examine structural features of the environment outside of the defender's control that can sustain his credible deterrence in equilibrium – the defender's fear, and the properties of the transgression itself.

The logic of our model and the North Korean case, however, clearly suggest actions that the defender would like to "do" to increase the credibility of his deterrent threat – to claim that he fears the challenger is unappeasably belligerent (even when he does not), and to claim that he is using the transgression as a test of that belligerence in order to select the deterrence equilibrium when it exists. Our model, however, is insufficiently rich for such actions to affect equilibrium outcomes. Cheap talk cannot select equilibria, and there is nothing for the defender to signal – either he fears the challenger or he does not.

[60] KCNA News Agency 2003.

[61] Carter and Perry 2006.

[62] Pritchard 2006.

The history of the deterrence literature, however, suggests a way forward. Classical deterrence theory conceives of the credibility of deterrence as rooted in an intrinsic willingness to fight. In order to understand how a defender can increase his credibility, subsequent theorizing therefore assumed that a challenger was *uncertain* of that willingness. Our theory, in contrast, conceives of the credibility of deterrence as also rooted in fear; thus, the way forward to understand the previously-described incentives is to assume that the challenger is *uncertain of that fear*.

This sort of "higher-order uncertainty" has been considered in the study of the "spiral model"; Kydd analyzes a game in which a state is uncertain about what his enemy believes about him, and this complicates his ability to draw inferences from the enemy's arming decisions – is the enemy aggressive, or just afraid?[63] In Kydd's analysis, states wish to signal about their own intentions to improve their ability to draw inferences about their enemy's intentions. Our model, however, suggests that it may also be fruitful to study when states wish to signal about their fear to improve their ability to deter.

Such a modeling approach could potentially eliminate the issue of multiple equilibria in the model. Moreover, classical mechanisms for improving the credibility of deterrence could be understood in a new light. Under what conditions could cheap talk about *fear* increase the effectiveness of deterrence? What restrains a defender's willingness to claim that he fears a defender's future intentions? What sorts of costly signals most credibly communicate fear? Can such signals backfire on the defender and result in avoidable wars? Relatedly, how can a challenger credibly communicate limited aims and thus exploit a defender with a known preference for appeasement? Could such claims backfire by making her appear sufficiently weak to be exploited herself? To the extent that fear is an important component of credible deterrence, understanding such incentives is an important avenue for future work.

[63]See Kydd 1997.

9 Appendix (Intended for Publication)

Proof of Lemma 1 Straightforward and omitted. ■

Proof of Proposition 1 The defender's strategy consists of a probability of responding to the transgression with war, which we denote α. The challenger's utility from not transgressing is n_C^1, and from transgressing is $\alpha \cdot w_C^1(\theta_C) + (1-\alpha) \cdot \max\{w_C^2(\theta_C), n_C^2\}$. The latter is strictly increasing in θ_C and greater than n_C^1 for all α when $\theta_C = \bar{\theta}_C^1$. Thus, the challenger's strategy must be to always transgress, or to transgress i.f.f her type is above a cutpoint $\bar{\theta}_C \leq \bar{\theta}_C^1$ at which she is indifferent between transgressing and not.

The necessary and sufficient conditions for existence of the two pure strategy equilibria ($\alpha^* = 0$ the no deterrence equilibrium, and $\alpha^* = 1$ the deterrence equilibrium) are described in the main text and straightforward to derive. There may also exist mixed strategy equilibria in which the defender responds with war with a strictly interior probability $\alpha^* \in (0,1)$. For such an equilibrium to hold, the defender must be indifferent between responding with war and allowing the transgression. This requires that,

$$P\left(\theta_C \geq \bar{\theta}_C^2 \,|\, \theta_C \geq \bar{\theta}_C\right) = \bar{\beta} \tag{2}$$

i.e. the defender's posterior belief that the challenger will initiate war if allowed to transgress is equal to his threshold belief $\bar{\beta}$. The left hand side approaches $P\left(\theta_C \geq \bar{\theta}_C^2\right)$ as $\bar{\theta}_C$ approaches the lower bound of the type space, is equal to 1 at $\bar{\theta}_C = \bar{\theta}_C^2$, and is strictly increasing in between. Thus, a cutpoint satisfying (2) exists i.f.f. the no deterrence equilibrium exists ($P\left(\theta_C \geq \bar{\theta}_C^2\right) < \bar{\beta}$). We denote this cutpoint $\bar{\theta}_C^*$, which must be $< \bar{\theta}_C^2$.

We now check conditions such that there exists some $\alpha^* \in (0,1)$ that induces the challenger to play the cutpoint strategy $\bar{\theta}_C^* < \bar{\theta}_C^2$. A necessary condition and sufficient condition is that this type be indifferent between transgressing and not, i.e. there exists an α^* s.t.

$$\alpha^* \cdot w_C^1\left(\bar{\theta}_C^*\right) + (1-\alpha^*) \cdot n_C^2 = n_C^1. \tag{3}$$

If $\bar{\theta}_C^* > \bar{\theta}_C^1 \iff P\left(\theta_C \geq \bar{\theta}_C^2 \,|\, \theta_C \geq \bar{\theta}_C^1\right) < \bar{\beta}$ (i.e. the deterrence equilibrium does not exist) then the condition cannot be satisfied since this would imply that both $w_C^1\left(\bar{\theta}_C^*\right)$ and n_C^2 are greater than n_C^1. Conversely, if $\bar{\theta}_C^* < \bar{\theta}_C^1$ then an α^* satisfying (3) exists and is unique.

Thus, a unique mixed strategy equilibrium exists i.f.f. both the no deterrence and deterrence equilibria exist, and the equilibrium strategies $\left(\alpha^*, \bar{\theta}_C^*\right)$ are uniquely characterized by (2) and (3). We now show that when there are multiple equilibria, i.e. $\bar{\beta} \in \left[P\left(\theta_C \geq \bar{\theta}_C^2\right), P\left(\theta_C \geq \bar{\theta}_C^2 \,|\, \theta_C \geq \bar{\theta}_C^1\right)\right]$, the defender is strictly better off in the deterrence equilibrium than in either the no deterrence or mixed strategy equilibrium. The defender's utility in the deterrence equilibrium is $U^{de} = P\left(\theta_C < \bar{\theta}_C^1\right) \cdot n_D^1 + P\left(\theta_C \geq \bar{\theta}_C^1\right) \cdot w_D^1$. His utility in the mixed strategy equilibrium is

$$\begin{aligned} U^{ms} &= P\left(\theta_C < \bar{\theta}_C^*\right) \cdot n_D^1 + P\left(\theta_C \geq \bar{\theta}_C^*\right) \cdot \left(P\left(\theta_C < \bar{\theta}_C^2 \,|\, \theta_C \geq \bar{\theta}_C^*\right) \cdot n_D^2 + P\left(\theta_C \geq \bar{\theta}_C^2 \,|\, \theta_C \geq \bar{\theta}_C^*\right) \cdot w_D^2\right) \\ &= P\left(\theta_C < \bar{\theta}_C^*\right) \cdot n_D^1 + P\left(\theta_C \geq \bar{\theta}_C^*\right) \cdot w_D^1 \qquad \text{by def'n of } \bar{\theta}_C^*. \end{aligned}$$

This is less than U^{de} since $\bar{\theta}_C^* < \bar{\theta}_C^1$ by construction $\rightarrow P\left(\theta_C < \bar{\theta}_C^*\right) < P\left(\theta_C < \bar{\theta}_C^1\right)$, and $n_D^1 > w_D^1$. Finally, his utility in the no deterrence equilibrium is

$$\begin{aligned} U^{nd} &= P\left(\theta_C < \bar{\theta}_C^2\right) \cdot n_D^2 + P\left(\theta_C \geq \bar{\theta}_C^2\right) \cdot w_D^2. \\ &= P\left(\theta_C < \bar{\theta}_C^1\right) \cdot \underbrace{\left(P\left(\theta_C < \bar{\theta}_C^2 \mid \theta_C < \bar{\theta}_C^1\right) \cdot n_D^2 + P\left(\theta_C \geq \bar{\theta}_C^2 \mid \theta_C < \bar{\theta}_C^1\right) \cdot w_D^2\right)}_{< n_D^1 \text{ since } n_D^1 > n_D^2 > w_D^1 > w_D^2} \\ &\quad + P\left(\theta_C \geq \bar{\theta}_C^1\right) \cdot \underbrace{\left(P\left(\theta_C < \bar{\theta}_C^2 \mid \theta_C \geq \bar{\theta}_C^1\right) \cdot n_D^2 + P\left(\theta_C \geq \bar{\theta}_C^2 \mid \theta_C \geq \bar{\theta}_C^1\right) \cdot w_D^2\right)}_{< w_D^1 \text{ since the deterrence equilibrium exists}} \\ &< P\left(\theta_C < \bar{\theta}_C^1\right) \cdot n_D^1 + P\left(\theta_C \geq \bar{\theta}_C^1\right) \cdot w_D^1 = U^{de} \end{aligned}$$

Proof of Corollary 2

$$\begin{aligned}\delta_C^m\left(\bar{\theta}_C^1\right) &\geq \delta_C^d \iff w_C^1\left(\bar{\theta}_C^1\right) \geq n_C^1 + \left(\delta_C^d - \delta_C^m\left(\bar{\theta}_C^1\right)\right) \iff w_C^2\left(\bar{\theta}_C^1\right) \geq n_C^2 \\ &\iff \bar{\theta}_C^2 \leq \bar{\theta}_C^1 \iff P\left(\theta_C \geq \bar{\theta}_C^2 \,|\, \theta_C \geq \bar{\theta}_C^1\right) = 1 \quad \blacksquare\end{aligned}$$

Proof of Proposition 2 Holding $\bar{\theta}_C^1$ (i.e. the challenger's first period payoffs) fixed, the ineffectiveness of appeasement $P\left(\theta_C \geq \bar{\theta}_C^2 \mid \theta_C \geq \bar{\theta}_C^1\right)$ is decreasing in $\bar{\theta}_C^2$. Thus for the first claim, it suffices to show $\bar{\theta}_C^2$ is decreasing in $\delta_C^m - \delta_C^d$. By assumption, $n_C^2 = n_C^1 + \delta_C^d$ and $w_C^2(\theta_C) = w_C^1(\theta_C) + \delta_C^m$ and $n_C^2 = w_C^2\left(\bar{\theta}_C^2\right)$, which together imply that $n_C^1 = w_C^1\left(\bar{\theta}_C^2\right) + \left(\delta_C^m - \delta_C^d\right)$. This implies the desired property since $w_C^1(\theta_C)$ is increasing in θ_C.

To show that the probability of deterrence is increasing in $\delta_C^m - \delta_C^d$, note that with the assumed equilibrium selection and by Proposition 1, the probability of deterrence is 0 if $P\left(\theta_C \geq \bar{\theta}_C^2 \mid \theta_C \geq \bar{\theta}_C^1\right) < \bar{\beta}$ and $P\left(\theta_C < \bar{\theta}_C^1\right)$ otherwise. Holding $\bar{\beta}$ (the defender's payoffs) fixed, the probability of deterrence is therefore (step-wise) increasing in $P\left(\theta_C \geq \bar{\theta}_C^2 \mid \theta_C \geq \bar{\theta}_C^1\right)$. Since this is increasing in $\delta_C^m - \delta_C^d$ the result is shown. $\blacksquare$

Proof of Proposition 3 With complete information the probability of deterrence is $P\left(\theta_1 < \bar{\theta}_C^1, \theta_1 \geq \bar{\theta}_C^2\right)$. With incomplete information, if the deterrence equilibrium exists, i.e., $P\left(\theta_C \geq \bar{\theta}_C^2 \,|\, \theta_C \geq \bar{\theta}_C^1\right) > \bar{\beta}$, then the probability of deterrence is $P\left(\theta_1 < \bar{\theta}_C^1\right)$ which is $> P\left(\theta_1 < \bar{\theta}_C^1, \theta_1 \geq \bar{\theta}_C^2\right)$. If the deterrence equilibrium does not exist then the probability of deterrence is 0. Since a necessary condition for this is $\bar{\theta}_C^2 > \bar{\theta}_C^1$, the probability of deterrence would also be 0 if the defender knew the challenger's type.

Now, the defender is always better off not knowing the challenger's type if appeasement is ineffective $\left(\bar{\theta}_C^2 \leq \bar{\theta}_C^1\right)$; types $< \bar{\theta}_C^2$ are deterred when they otherwise would not be, and for all other types the outcome is identical. If appeasement could be effective $\left(\bar{\theta}_C^1 < \bar{\theta}_C^2\right)$ then she gains $n_D^2 - n_D^1$ against peaceful types $< \bar{\theta}_C^1$ who would have otherwise transgressed, but loses $n_D^2 - w_D^1$ by fighting preventable wars against appeasable types $\theta_C \in \left[\bar{\theta}_C^1, \bar{\theta}_C^2\right)$. Against all other types the outcomes are identical. Thus, in expectation the defender is better off not knowing the challenger's type if and only if

$$P\left(\theta_C < \bar{\theta}_C^1\right) \cdot \left(n_D^1 - n_D^2\right) > P\left(\theta_C \in \left[\bar{\theta}_C^1, \bar{\theta}_C^2\right)\right) \cdot \left(n_D^2 - w_D^1\right),$$

i.e. the benefit of deterring types $< \bar{\theta}_C^1$ exceeds the cost of preventable wars $n_C^2 - w_C^1$ against appeasable types. It is straightforward to that show this condition reduces to the condition stated in the Proposition. ■

References

Acharya, Avidit and Edoardo Grillo. 2014. "War with Crazy Types." *Political Science Research and Methods,* forthcoming.

Acheson, Dean. 1969. *Present at the Creation: My Years in the State Department.* New York: Norton.

Alt, James, Randall Calvert and Brian Humes. 1988. "Reputation and Hegemonic Stability: A Game-Theoretic Analysis." *American Political Science Review* 82(2):445–466.

Baliga, Sandeep and Tomas Sjostrom. 2008. "Strategic Ambiguity and Arms Proliferation." *Journal of Political Economy* 116(6):1023–1057.

Baliga, Sandeep and Tomas Sjostrom. 2009. "Conflict Games with Payoff Uncertainty." Unpublished Draft.

Carter, Ashton B. and William J. Perry. 1999. *Preventive Defense: A New Security Strategy for America.* Washington, DC: Brookings Institution Press.

Carter, Ashton B. and William J. Perry. 2006. "If Necessary, Strike and Destroy." *Washington Post*, 22 June 2006.

Chassang, Sylvain and Gerard Padro i Miquel. 2010. "Conflict and Deterrence Under Strategic Risk." *Quarterly Journal of Economics* 125(4):1821–1858.

Chiozza, Giacomo and H.E. Goemans. 2004. "International Conflict and the Tenure of Leaders: Is War Still 'Ex Post' Inefficient?" *American Journal of Political Science* 48(3):604–619.

DeLuca, Anthony R. 1977. "Soviet-American Politics and the Turkish Straits." *Political Science Quarterly* 92(3):503–524.

Fearon, James. 1994. "Domestic Political Audiences and the Escalation of International Disputes." *American Political Science Review* 88(3):577–592.

Fearon, James. 1995. "Rationalist Explanations for War." *International Organization* 49(3):379–414.

Fearon, James. 1996. "Bargaining over Objects that Influence Future Bargaining Power." Unpublished Manuscript, Chicago, IL.

Fey, Mark and Kristopher W. Ramsay. 2011. "Uncertainty and Incentives in Crisis Bargaining: Game-Free Analysis of International Conflict." *American Journal of Political Science* 55(1):149–169.

Gaddis, John Lewis. 2000. *The United States and the Origins of the Cold War, 1941-1947*. New York: Columbia University Press.

Glaser, Charles. 1997. "The Security Dilemma Revisited." *World Politics* 50(1):171–201.

Huth, Paul K. 1999. "Deterrence and International Conflict: Empirical Findings and Theoretical Debates." *Annual Review of Political Science* 2:25–48.

Jackson, Matthew and Massimo Morelli. 2007. "Political Bias and War." *American Economic Review* 97(4):1353–1373.

Jones, Joseph Marion. 1964. *The Fifteen Weeks: An Inside Account of the Genesis of the Marshall Plan*. New York: Harcourt, Brace and World, Inc.

Kayaoglu, Barin. 2009. "Strategic Imperatives, Democratic Rhetoric: The United States and Turkey, 1945-1952." *Cold War History* 9(3):321–345.

KCNA News Agency. 2003. "North Korea warns 'total war' if USA attacks 'peaceful' plant." BBC Summary of World Broadcasts, February 6, 2003. Lexis-Nexis Academic: News.

Kuniholm, Bruce Robellet. 1980. *The Origins of the Cold War in the Near East*. Princeton, NJ: Princeton University Press.

Kydd, Andrew. 1997. "Game Theory and the Spiral Model." *World Politics* 49(3):371–400.

Kydd, Andrew H. and Roseanne W. McManus. 2014. "Threats and Assurances In Crisis Bargaining." *Journal of Conflict Resolution* forthcoming.

Leffler, Melvyn P. 1985. "Strategy, Diplomacy, and the Cold War: The United States, Turkey, and NATO, 1945-1952." *The Journal of American History* 71(4):807–825.

Lewis, Jeffrey B. and Kenneth A. Schultz. 2003. "Revealing Preferences: Empirical Estimation of a Crisis Bargaining Game with Incomplete Information." *Political Analysis* 11:345–367.

Mark, Eduard. 1997. "The War Scare of 1946 and its Consequences." *Diplomatic History* 21(3):383–415.

Mark, Eduard. 2005. The Turkish War Scare of 1946. In *Origins of the Cold War: An International History, Second Edition*, ed. Melvyn P. Leffler and David S. Painter. New York: Routledge.

May, Ernest R. and Philip D. Zelikow, eds. 1997. *The Kennedy Tapes: Inside the White House During the Cuban Missile Crisis*. Cambridge, MA: Belknap Press.

Milgrom, Paul and John Roberts. 1982. "Predation, Reputation and Entry Deterrence." *Journal of Economic Theory* 27(2):280–312.

Mills, Walter, ed. 1951. *The Forrestal Diaries.* New York: Viking Press.

Nalebuff, Barry. 1986. "Brinkmanship and Nuclear Deterrence: The Neutrality of Escalation." *Conflict Management and Peace Science* 9:19–30.

O'Neill, Barry. 1999. *Honor, Symbols and War.* Ann Arbor, MI: University of Michigan Press.

Powell, Robert. 1987. "Crisis Bargaining, Escalation, and MAD." *American Political Science Review* 81(3):717–735.

Powell, Robert. 1996. "Uncertainty, Shifting Power and Appeasement." *American Political Science Review* 90(4):749–764.

Powell, Robert. 2006. "War as a Commitment Problem." *International Organization* 60(1):169–203.

President Harry S. Truman Office Files, 1945-1953. 1979. "Memorandum of Interview with President Inonu of Turkey, Kenneth C. Royall Secretary of the Army." Reel 39: January 7, 1949. NLT (PSF-Subj) 403. Published by ProQuest, LLC.

Pritchard, Charles L. 2006. "No, Don't Blow it Up." *Washington Post*, 23 June 2006.

Ross, Steven T. 1996. *American War Plans, 1945-1950.* Portland, OR: Frank Cass.

Ross, Steven T. and David Alan Rosenberg, eds. 1989. *America's Plans for War Against the Soviet Union, 1945-1950, Volume 1.* New York: Garland Publishers.

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

Schwarz, Michael and Konstantin Sonin. 2008. "A Theory of Brinkmanship, Conflicts and Commitments." *Journal of Law, Economics and Organization* 24(1):163–183.

Sechser, Todd S. 2010. "Goliath's Curse: Coercive Threats and Asymmetric Power." *International Organization* 64(4):627–660.

Slantchev, Branislav L. 2011. *Military Threats: The Costs of Coercion and the Price of Peace*. New York, NY: Cambridge University Press.

Soman, Appu Kuttan. 2000. *Double-Edged Sword: Nuclear Diplomacy in Unequal Conflicts, The United States and China, 1950-1958*. Westport, CT: Praeger.

Trachtenberg, Marc. 1999. *A Constructed Peace: The Making of the European Settlement, 1945-1963*. Princeton, NJ: Princeton University Press.

Trachtenberg, Marc. 2007. "Preventive War and U.S. Foreign Policy." *Security Studies* 16(1):1–31.

Trachtenberg, Marc. 2012. "Audience Costs: An Historical Analysis." *Security Studies* 21(1):3–42.

U.S. Department of State. 1969. "Foreign Relations of the United States, 1946. Vol. VII: The Near East and Africa." United States Government Printing Office, Washington, DC.

U.S. Department of State. 1986. "Foreign Relations of the United States, 1952-54. Vol. VII, Part 2: Germany and Austria." United States Government Printing Office, Washington, DC.

Wit, Joel S., Daniel B. Poneman and Robert L. Gallucci. 2004. *Going Critical: The First North Korean Nuclear Crisis*. Washington, DC: Brookings Institution Press.

Zagare, Frank C. 2004. "Reconciling Rationality with Deterrence: A Re-examination of the Logical Foundations of Deterrence Theory." *Journal of Theoretical Politics* 16(2):107–141.

Figure 1: Model

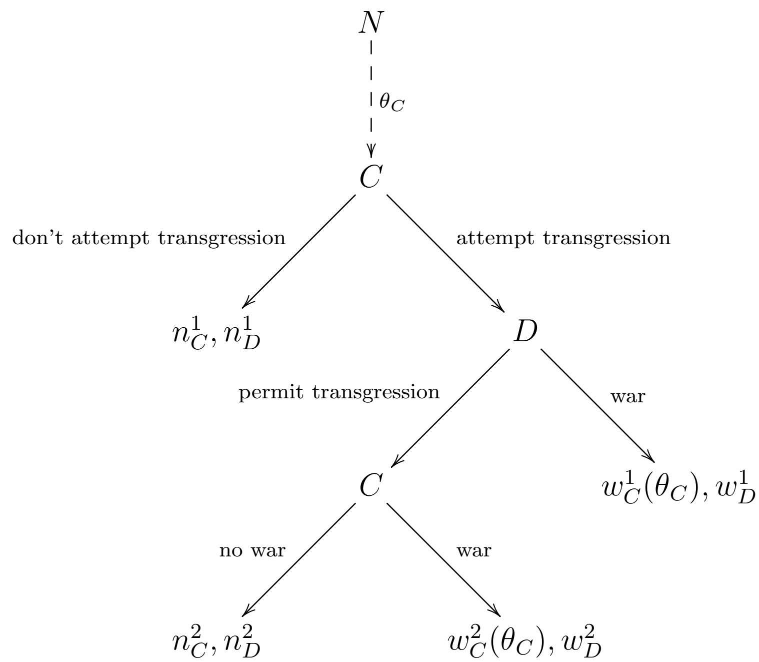
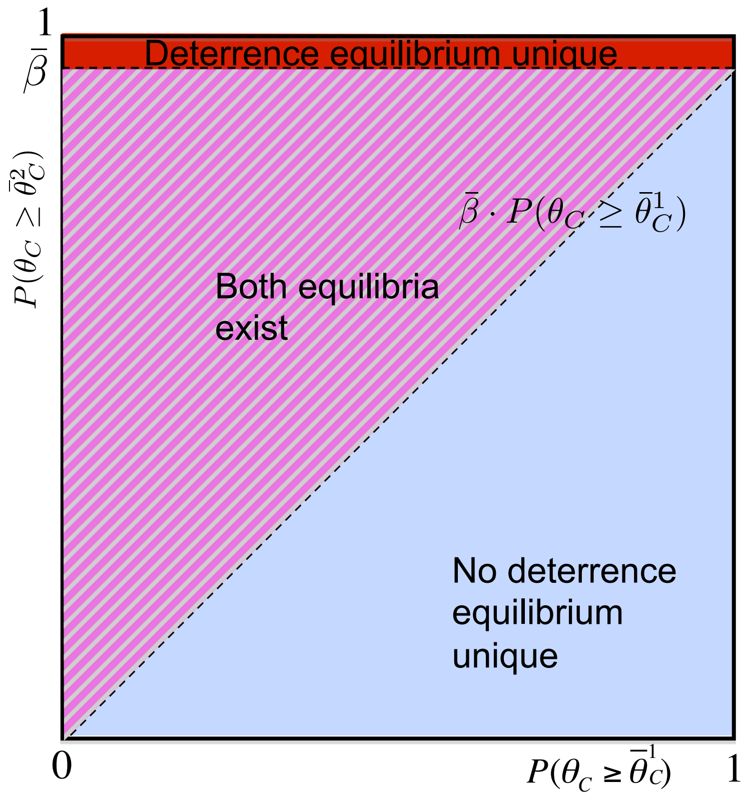


Figure 2: Equilibria



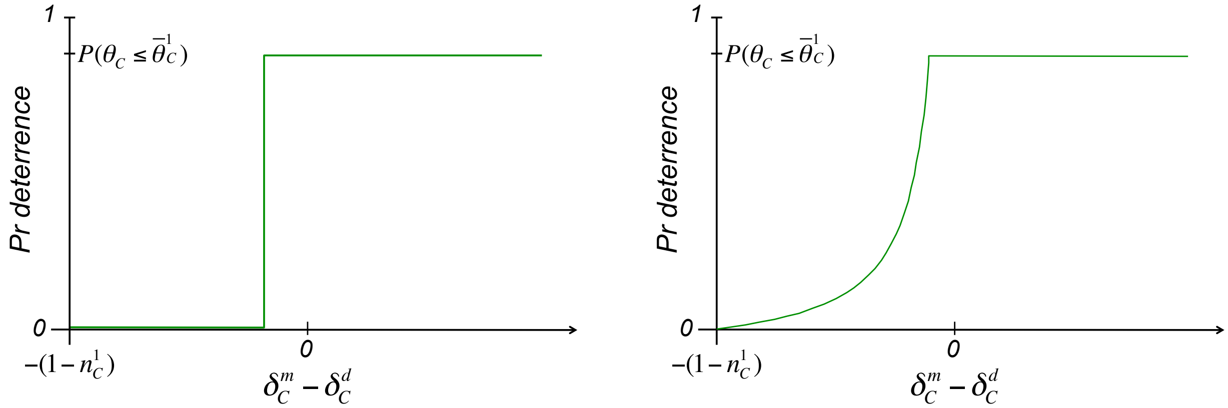


Figure 3: Probability of deterrence as function of $\delta_C^m - \delta_C^m$

Figure 4: Probability of Deterrence when Challenger's Gains = Defender's Costs

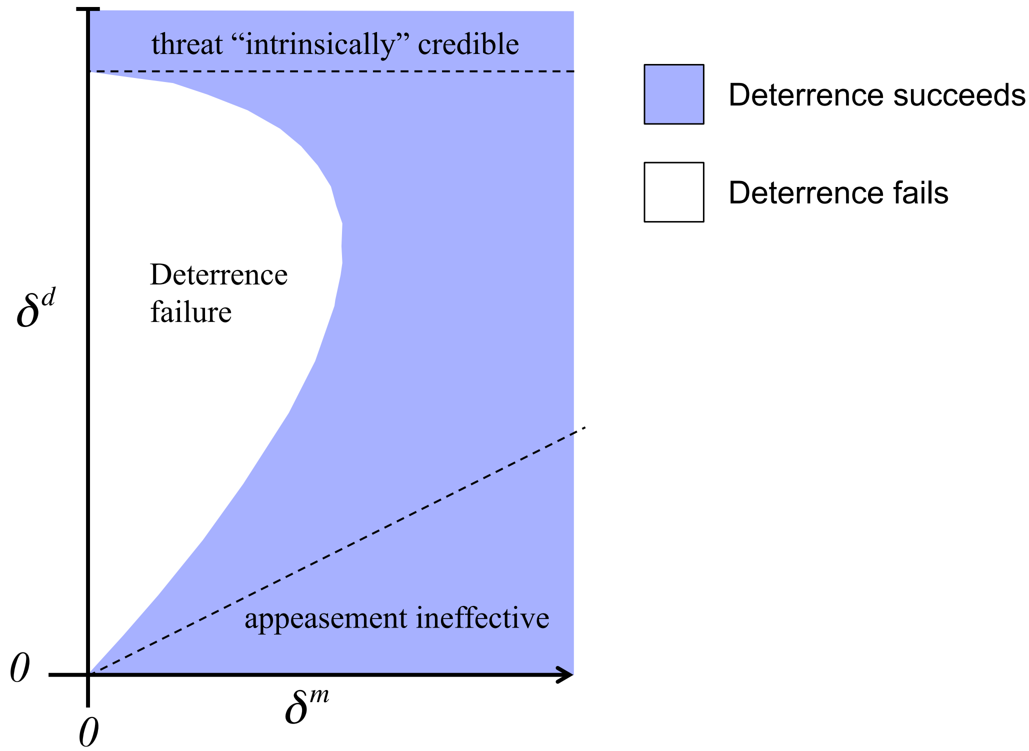
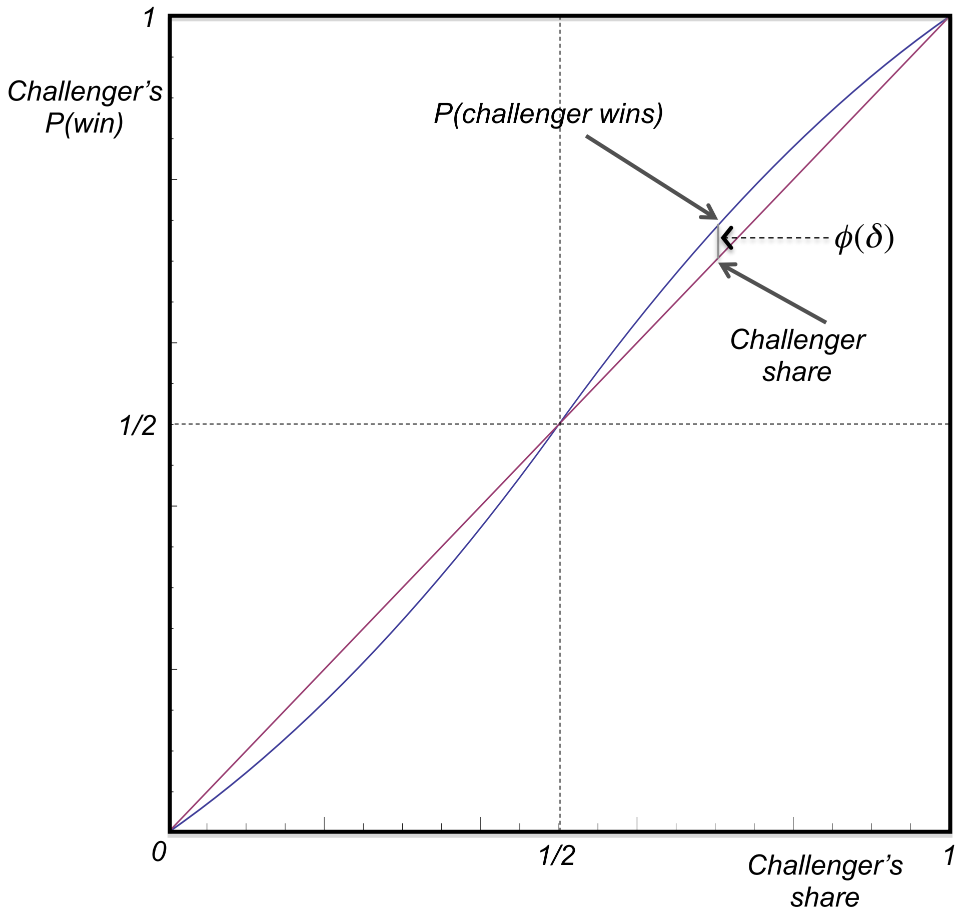


Figure 5: Probability that Challenger Wins in Generalized Example



A Supplemental Appendix – Not for Publication

"Fear, Appeasement, and the Effectiveness of Deterrence," April 1 2015

A.1 Robustness to Salami Tactics and Endogenous Demands

In this section we consider robustness to two alternative bargaining protocols – a) extending the sequence so that the game resembles a model of salami tactics, and b) endogenizing the demand made by the challenger. In the former extension the defender always has the final move in each period over whether to fight or concede.

Rather than fully solve out general versions of these games, we present two examples illustrating that our basic insight holds in these variants. Both examples are constructed from the following payoff environment for a finite period game of conflict over a landmass of size and value equal to 1. In both variants there is no discounting and no "flow" payoffs – payoffs are based on the holdings of the landmass in the period in which the game ends.

Payoff Environment Suppose a challenger and a defender jointly occupy a landmass of size and value equal to 1. Say the **advantaged** party at time t is that which holds a majority of the landmass, and let δ_t denote the **excess** holdings of the advantaged party in period t above $\frac{1}{2}$. If a war occurs in period t, the probability the advantaged party wins is:

$$p(\delta_t) = \left(\frac{1}{2} + \delta_t\right) + \phi(\delta_t)$$

where $\phi(\delta_t) = \frac{2\delta_t(1-2\delta_t)}{Z}$ and Z is very large.[64] Also suppose that the defender's cost of war is commonly known to be $c_D \geq \frac{1}{4}$. The challenger's type θ_C is unknown and uniformly distributed over $\theta_C \sim U\left[-\frac{1}{4}, 0\right]$, and her cost of war is $c_C = -\theta_C$. ■

[64] We require at least $Z > 6$ for $p(\delta_t)$ to be strictly increasing in δ_t.

The challenger's probability of victory in a war as a function of her position is depicted in Figure 5 in the main text. We now present the first extension.

Extension 1 (Salami Tactics)**.** *Consider a $T \geq 3$ period game, and a $T+1$-length series of increasing values $0 < \delta_1 ... < \delta_T = \frac{1}{2}$. Each δ_t represents the challenger's* ***excess*** *holdings above $\frac{1}{2}$ if the game advances to period t. Assume that the challenger is initially advantaged ($\delta_1 > 0$), and that the increments of advancement $\delta_t - \delta_{t-1}$ are less than the defender's cost of war c_D for all $t < T$. In each period t, the challenger decides whether or not to attempt to advance from δ_t to δ_{t+1}. If she doesn't attempt to advance the game ends. If she does attempt to advance, the defender chooses whether or not to respond with war, and the challenger's probability of victory is $p(\delta_t)$.*

In this extension there are a finite number of exogenously fixed positions to which the challenger can advance, each advancement represents a transgression of fixed size, and each increment $\delta_t - \delta_{t-1}$ of advancement short of possessing the entire landmass is less than the defender's cost $c_D \geq \frac{1}{4}$ of war. Thus, the defender is always vulnerable to "salami tactics." When $\delta_1 < \delta_2 < c_D < \delta_3$, the game essentially reduces to the baseline model; the reason is that the defender knows she will be unable to credibly resist any advancement beyond δ_2, anticipates that conceding at δ_2 will result in a concession of size $1 - \delta_2 > c_D$, and is therefore there willing to fight.[65]

The set of equilibria for this extension satisfies the following proposition.

Proposition 4. *If $c_D \in \left[\frac{1}{4}, \frac{1}{2}\right)$, then for any t^* such that $c_D < \frac{1}{2} - \delta_{t^*} \iff \delta_{t^*} < \frac{1}{2} - c_D$, there exists an equilibrium in which all types of challengers advance to δ_{t^*}, only challengers*

[65] When $\delta_1 < \delta_2 < c_D < \delta_3$, the game maps to the baseline model by letting the defender's payoffs be $n_D^1 = \frac{1}{2} - \delta_1$, $n_D^2 = \frac{1}{2} - \delta_2$, $w_D^1 = \left(\frac{1}{2} - \delta_1\right) - \phi(\delta_1) - c_D$, $w_D^2 = \left(\frac{1}{2} - \delta_2\right) - \phi(\delta_2) - c_D$, the challenger's payoffs be $n_C^1 = \frac{1}{2} + \delta_1$, $n_C^2 = \frac{1}{2} + \delta_2$, $w_C^1 = \left(\frac{1}{2} + \delta_1\right) + \phi(\delta_1) + \theta_C$, $w_C^2 = \left(\frac{1}{2} + \delta_2\right) + \phi(\delta_2) + \theta_C$, and $\theta_C \sim U\left[-\frac{1}{4}, 0\right]$.

with cost $c_C < \phi(\delta_{t^*})$ attempt to advance to δ_{t^*+1}, and the defender always responds with war. If $c_D > \frac{1}{2}$ then in any equilibrium the challenger occupies the entire landmass.

Proof: We first show the desired equilibrium when $c_D \in \left[\frac{1}{4}, \frac{1}{2}\right)$ and $\delta_t^* < \frac{1}{2} - c_D$. Define $\bar{t}$ as the period with the largest $\delta_{\bar{t}}$ strictly less than $\frac{1}{2} - c_D$, and observe that $\delta_{\bar{t}} < \frac{1}{4}$. Now consider the following strategy profile. After all histories, in periods $t \in [t^*, \bar{t}]$ the defender responds to advancement with war, and the challenger only attempts to advance if $c_C < \phi(\delta_t)$. In all other periods the defender never responds with war, and the challenger always advances. This profile produces the desired equilibrium outcomes and the challenger is best responding. So we must show that the defender doesn't wish to deviate.

Consider first a period $t < t^*$ in which the challenger attempts to advance. To get to this period the challenger must have advanced to t and the defender must have always permitted it. So this is on equilibrium path, if the defender plays his equilibrium strategy of again permitting advancement then the challenger will advance all the way to δ_t^* before advancing triggers war, and the defender's expected payoff is:

$$\left(\frac{1}{2} - \delta_t^*\right) - P\left(c_C < \phi\left(\delta_t^*\right)\right)\left(\phi\left(\delta_t^*\right) + c_D\right). \tag{A.1}$$

In words, the defender's equilibrium expected holdings are $\frac{1}{2} - \delta_t^*$, with probability $P\left(c_C < \phi\left(\delta_t^*\right)\right)$ war occurs in period t^*, and when this occurs the defender suffers the challenger's excess military advantage $\phi\left(\delta_t^*\right)$ and the cost of war c_D. If instead the defender responds with war in period t, his payoff is $\left(1 - p\left(\delta_t\right)\right) - c_D = \left(\frac{1}{2} - \delta_t - \phi\left(\delta_t\right)\right) - c_D$, which is $<$ eqn. $(A.1)$ i.f.f.

$$c_D > \frac{\left(\left(\delta_t^* + \phi\left(\delta_t^*\right)\right) - \left(\delta_t + \phi\left(\delta_t\right)\right)\right)}{\left(1 - P\left(c_C < \phi\left(\delta_t^*\right)\right)\right)} - \phi\left(\delta_t^*\right).$$

Since $\phi\left(\delta_t\right) \to 0 \; \forall \delta_t$ as $Z \to \infty$, the r.h.s. approaches $\delta_t^* - \delta_t < \frac{1}{4}$ (since $\delta_t > 0$ and $\delta_t^* \leq \delta_{\bar{t}} < \frac{1}{4}$) as $Z \to \infty$. So since $c_D \geq \frac{1}{4}$ there exists a Z sufficiently large such that the inequality is satisfied for all $t < t^*$. Intuitively, we can scale down the excess military advantage function

$\phi(\delta_t)$ by increasing Z sufficiently so that the calculation essentially reduces to whether the cost of war exceeds the foregone share of the landmass from allowing the challenger to advance from t all the way to t^*. This will always be true since (by assumption) the cost of war exceeds the challenger's excess holdings in the period where war occurs ($\delta_{t^*} < c_D$).

Now consider a period $t \geq \bar{t}$ in which the challenger attempts to advance. This is off path, but we do not need beliefs about the challenger's type since if she is allowed to advance the strategies are for her to continue to advance and the defender to permit it. So if the defender allows advancement in t the challenger will eventually possess the entire landmass and the defender's payoff will be 0. If instead he responds with war his payoff is $\left(\frac{1}{2} - \delta_t - \phi(\delta_t)\right) - c_D$. Since $\delta_{\bar{t}} < \frac{1}{2} - c_D$ and $\delta_t > \frac{1}{2} - c_D$ $\forall t > \bar{t}$, for Z sufficiently large it will be optimal for the defender to respond with war in $\bar{t}$ but not in $t > \bar{t}$. In words, at $\bar{t}$ the remaining landmass just exceeds the defender's cost of war, so he will respond with war knowing that should he allow advancement he will also allow it in all future periods. For $t > \bar{t}$, the challenger is already sufficiently advanced that letting her take the remaining landmass is optimal.

Finally, consider a period $\hat{t} \in [t^*, \bar{t})$ in which the challenger attempts to advance and the defender is supposed to respond with war. The challenger already advanced in period t^* expecting to trigger war. So the defender *infers* in equilibrium that her cost $c_C < \phi(\delta_{t^*})$, the threshold in the first period t^* in which she advanced expecting war. If she is allowed to again advance in period $\hat{t}$ to period $\hat{t}+1$, a further attempt to advance in $\hat{t}+1$ will provoke war. Anticipating this, the challenger will once again advance i.f.f. $c_C < \phi\left(\delta_{\hat{t}+1}\right)$. Recall that $\delta_{\hat{t}+1} \leq \delta_{\bar{t}} < \frac{1}{4}$ and $\phi(\delta_t)$ is increasing over $\left[0, \frac{1}{4}\right]$, so $\phi(\delta_{t^*}) < \phi\left(\delta_{\hat{t}+1}\right)$. In words, the region of the landmass is s.t. advancement makes war relatively more attractive to the challenger. So the defender can infer that a challenger who advanced to period $\hat{t}$ expecting war will again advance in period $\hat{t}+1$ even though it will trigger war for sure. So responding with war in $\hat{t}$ is optimal, since permitting advancement will only weaken the defender in the inevitable war.

We next argue that when $c_D \geq \frac{1}{2}$, equilibrium requires the challenger to occupy the entire landmass. Suppose the defender's strategy involves responding to further advancement with war with strictly positive probability in any period $t \geq 1$. This would yield utility $\left(\frac{1}{2} - \delta_t - \phi\left(\delta_t\right)\right) - c_D$ which is < 0 since $\delta_1 > 0$. If she were instead to deviate to always allowing advancement in every period, her utility would be ≥ 0; at worst the challenger's strategy will involve occupying the entire landmass (recall that the defender holds the final decision to fight). Thus, equilibrium requires the defender to always permit advancement. Equilibrium then must also requires the challenger to attempt advancement in every period regardless of her type since it will always be permitted, and the unique equilibrium outcome is that she will occupy the entire landmass. ■

We now consider the second extension, in which the challenger makes an endogenous "demand" δ_2 of how far to advance. As in the baseline model, in this example the defender can allow a positive demand or respond with war, and if she advances the challenger can exploit her gains afterward by unilaterally initiating war.

Extension 2 (Endogenous Transgression)**.** *Consider the following $T = 2$ period game. In period 1 the challenger's excess holdings are $\delta_1 > 0$, and she can attempt to advance to some $\delta_2 \in \left[\delta_1, \frac{1}{2}\right]$ of her choosing. The defender can permit the advancement or respond with war. If he permits it, then the game proceeds to the second period, and the challenger decides whether to unilaterally initiate war or enjoy her gains. After either choice the game ends.*

The set of equilibria in this extension satisfy the following proposition.

Proposition 5. *If $c_D \in \left[\frac{1}{4}, \frac{1}{2}\right)$ and the challenger's excess share δ_1 under the status quo is less than $\frac{1}{4} - \frac{c_D}{2}$, then there exists an equilibrium in which the defender responds to a strictly positive demand $\delta_2 \in \left(\delta_1, \frac{1}{2}\right]$, however small, with war. If $c_D \geq \frac{1}{2}$, then in any equilibrium the challenger demands the entire landmass, i.e. $\delta_2 = 1$, and it is accepted.*

Proof: We first show the desired equilibrium when $c_D \in \left[\frac{1}{4}, \frac{1}{2}\right)$; we construct an equilibrium where all demands are on-path. Challengers with cost $c_C > \phi(\delta_1)$ demand the status quo ($\delta_2^* (c_C) = \delta_1$), it is accepted, they do not initiate war in period 2, and the game ends. All challengers with cost $c_C \leq \phi(\delta_1)$ mix identically over all positive demands $\delta_2 \in \left(\delta_1, \frac{1}{2}\right]$ and the defender always responds with war. Should any such demand be accepted (off path), challengers with cost $c_C < \phi(\delta_2)$ unilaterally initiate war in the second period.

To see this is an equilibrium, consider first the defender's strategy. If he sees no demand ($\delta_2 = \delta_1$), he infers that the challenger will initiate war in the second period with probability 0 and so maintaining the status quo is optimal. Should he see a positive demand ($\delta_2 > \delta_1$), he can infer that the challenger's cost is below $\phi(\delta_1)$ but no more, since all such challengers mix identically over all positive demands. If the demand he receives satisfies $\delta_2 \in \left(\delta_1, \frac{1}{2} - \delta_1\right)$, then $\phi(\delta_1) < \phi(\delta_2)$, and since challengers with cost $c_C < \phi(\delta_2)$ will unilaterally initiate war in period 2, the probability of appeasing an already belligerent challenger by accepting such a demand is 0. Thus responding with war is optimal. If instead $\delta_2 \in \left[\frac{1}{2} - \delta_1, \frac{1}{2}\right]$, then even if allowing the demand would appease the challenger for sure the defender prefers to respond with war, since accepting such a demand will leave the defender with no more than δ_1, while responding with war leaves him with $\left(\frac{1}{2} - \delta_1 - \phi(\delta_1)\right) - c_D > \delta_1$ when $\delta_1 < \frac{1}{4} - \frac{c_D}{2}$ for sufficiently large Z.

To see that the challenger wishes to play her equilibrium strategy, first note that period 2 strategies are straightforwardly optimal since the challenger is the last mover. In period 1, any positive demand will provoke war, and all challengers with cost $c_C \leq \phi(\delta_1)$ prefer war to the status quo. So such challengers are indifferent between all positive demands and are willing to mix according to the equilibrium strategy. Finally, challengers with cost $c_C > \phi(\delta_1)$ prefer the status quo to war and so making a 0 demand $\delta_2 = \delta_1$ is optimal.

We last argue that when $c_D \geq \frac{1}{2}$, equilibrium requires the challenger to occupy the entire landmass. If she demands $\delta_2 = 1$ and it is accepted, then in the second period it is optimal to end the game with peace regardless of her type. In the first period, the defender will thus accept such a demand, since it will yield utility 0, while war yields utility $\left(\frac{1}{2} - \delta_1 - \phi(\delta_1)\right) - c_D < 0$ for $c_D \geq \frac{1}{2}$. Finally, in any equilibrium the challenger must demand $\delta_2 = 1$ in the first period; all other demands will yield strictly lower utility regardless of the defender's response, or her own anticipated strategy in the second period. ■

Proposition 5 further demonstrates that our result is not an artifact of having a fixed size of the transgression. When the status quo division is sufficiently close to an even division, there exists equilibria in which the defender responds to *any* positive demand, however small, with war. The logic is again identical to the two-period model. At the status quo, the challenger expects the defender to respond to any positive demand with war. Hence, the defender can infer in equilibrium that a challenger who makes such a demand desires war under the status quo. Because the challenger's probability of victory $p(\delta_t)$ is such that advancements $\delta_2 \in \left(\delta_1, \frac{1}{2} - \delta_1\right)$ make war relatively more attractive, the probability of appeasing an already-belligerent challenger by permitting such an advancement is 0. Alternatively, while advancements $\delta_2 \in \left[\frac{1}{2} - \delta_1, \frac{1}{2}\right]$ have some hope of successful appeasement, they are so large that the defender prefers to suffer the cost of war.

A.2 Robustness to challenger backing down

Proposition 6. *Consider an alternative game Γ' form in which the challenger can back down in the first stage if the defender resists. Whenever the deterrence equilibrium exists in the original game Γ it also exists in Γ'.*

Proof: In Γ' the deterrence equilibrium takes the following form; the defender always resists, the challenger is deterred unless she prefers immediate war, and when she transgresses she also fights upon resistance. Now if the defender always resists, challenger types $\theta_C \geq \bar{\theta}_C^1$ still prefer to transgress because the defender will resist and they will then proceed with war. Challenger types $\theta_C < \bar{\theta}_C^1$ cannot get away with the transgression because the defender always resists, can back down upon encountering resistance, and are therefore indifferent between transgressing and not; they are thus willing to play the required strategy of not transgressing. Upon observing a transgression the defender therefore continues to infer that the challenger is of type $\theta_C \geq \bar{\theta}_C^1$, and in this case resisting is equivalent to unilaterally initiating war himself; his incentives and inferences are unchanged and he is therefore willing to carry out his equilibrium strategy. ■

A.3 Game with interdependent war values

Both players' payoffs in the event of war depend on the challenger's type $\theta_C \in \Theta \subset \mathbb{R}$ that is unknown to the defender but known to the challenger, where Θ is an interval and θ_C has a prior distribution $f(\theta_C)$ with full support over Θ. The challenger's type is therefore to be interpreted as a state of the world that affects both players' payoffs over which the challenger has private information. Our notation and assumptions for the challenger's payoffs are unchanged. For the defender, we now express the dependence of his war payoff on the challenger's type using $w_D^t(\theta_C)$, and make the following slightly-modified assumptions.

1. For all challenger types, allowing the transgression makes the defender strictly worse off in both peace ($n_D^2 < n_D^1$) and war ($w_D^2(\theta_C) < w_D^1(\theta_C)\ \forall \theta_C$).

2. For all challenger types, allowing the transgression is strictly better than responding with war if the challenger will subsequently choose peace ($n_D^2 > w_D^1(\theta_C)\ \forall \theta_C$).

Note that our defender assumptions jointly imply that the defender strictly prefers peace to war in each t for every type of challenger. Moreover, conditional on defender assumptions $(1)-(2)$, any arbitrary dependence of the defender's war payoff $w_D^t(\theta_C)$ on the challenger's type can be accommodated. However, it is natural to assume that $w_D^t(\theta_C)$ is weakly decreasing in θ_C, i.e., a more belligerent challenger means a weaker defender. Our setup is not completely without loss of generality because it cannot capture when the challenger is privately informed about factors affecting the defender's war payoffs but not her own; however, it is sufficiently general to capture private information about the probability of victory.

Challenger Incentives In the second period, the challenger transgresses i.f.f. $\theta_C \geq \bar{\theta}_C^2$. In the first period, challengers of type $\theta_C \geq \bar{\theta}_C^1$ always transgress. Challengers of type $\theta_C < \bar{\theta}_C^1$ transgress i.f.f.,

$$\alpha \cdot w_C^1(\theta_C) + (1-\alpha) \cdot \max\left\{n_C^2, w_C^2(\theta_C)\right\} \geq n_C^1.$$

For each such type, there exists a unique interior probability $\hat{\alpha}(\theta_C)$ that would make them indifferent between transgressing and not, and given that probability the challenger would play a cutpoint strategy at θ_C. It is simple to verify that for $\theta_C \leq \bar{\theta}_C^1$, $\hat{\alpha}(\theta_C)$ is always well defined, strictly increasing in θ_C, strictly interior to $(0,1)$, and $\hat{\alpha}\left(\bar{\theta}_C^1\right) = 1$.

Defender's Incentives Suppose that the challenger uses a threshold for transgressing equal to $\hat{\theta}_C$. Then upon observing a transgression, the defender's payoff from war is

$$\int_{\hat{\theta}_C}^{\infty} w_D^1(\theta_C) \frac{f(\theta_C)}{1-F\left(\hat{\theta}_C\right)} d\theta_C$$

and from appeasement is,

$$\int_{\hat{\theta}_C}^{\max\left\{\hat{\theta}_C,\bar{\theta}_C^2\right\}} n_D^2 \frac{f(\theta_C)}{1-F\left(\hat{\theta}_C\right)} d\theta_C + \int_{\max\left\{\hat{\theta}_C,\bar{\theta}_C^2\right\}}^{\infty} w_D^2(\theta_C) \cdot \frac{f(\theta_C)}{1-F\left(\hat{\theta}_C\right)} d\theta_C.$$

Hence she will prefer to respond to the transgression with war i.f.f.

$$\int_{\max\left\{\hat{\theta}_C,\bar{\theta}_C^2\right\}}^{\infty} \left(w_D^1(\theta_C) - w_D^2(\theta_C)\right) \cdot \frac{f(\theta_C)}{1-F\left(\hat{\theta}_C\right)} \geq \int_{\hat{\theta}_C}^{\max\left\{\hat{\theta}_C,\bar{\theta}_C^2\right\}} \left(n_D^2 - w_D^1(\theta_C)\right) \frac{f(\theta_C)}{1-F\left(\hat{\theta}_C\right)} d\theta_C$$

Now it is straightforward to show that the condition above is satisfied i.f.f.

$$\bar{\beta}\left(\hat{\theta}_C\right) \leq P\left(\theta \geq \bar{\theta}_C^2 \,|\, \theta \geq \hat{\theta}_C\right), \tag{A.2}$$

where

$$\bar{\beta}\left(\hat{\theta}_C\right) = \frac{n_D^2 - E\left[w_D^1(\theta_C) \mid \theta_C \in \left[\hat{\theta}_C, \bar{\theta}_C^2\right]\right]}{\left(n_D^2 - E\left[w_D^1(\theta_C) \mid \theta_C \in \left[\hat{\theta}_C, \bar{\theta}_C^2\right]\right]\right) + E\left[w_D^1(\theta_C) - w_D^2(\theta_C) \mid \theta_C \geq \bar{\theta}_C^2\right]} \tag{A.3}$$

Intuitively, $n_D^2 - E\left[w_D^1(\theta_C) \mid \theta_C \in \left[\hat{\theta}_C, \bar{\theta}_C^2\right]\right]$ is the benefit from appeasement conditional on the challenger being appeasable. Similarly, $E\left[w_D^1(\theta_C) - w_D^2(\theta_C) \mid \theta_C \geq \bar{\theta}_C^2\right]$ is the benefit from preemptive war conditional on the challenger being unappeasable. Finally, as before $P\left(\theta_C \geq \bar{\theta}_C^2 \,|\, \theta_C \geq \hat{\theta}_C\right)$ is the interim probability that the challenger is unappeasable.

Now note the following. First, $\bar{\beta}\left(\hat{\theta}_C\right)$ is strictly interior to $[0,1]$ for any value of $\hat{\theta}_C$ by our payoff assumptions, since appeasement is beneficial when it is possible $\left(\theta_C \in \left[\hat{\theta}_C, \bar{\theta}_C^2\right]\right)$ and war early is better than war later when it is not $\left(\theta_C > \bar{\theta}_C^2\right)$. Second, $\bar{\beta}\left(\hat{\theta}_C\right)$ is weakly increasing in $\hat{\theta}_C$ in the natural case where a belligerent challenger is "bad news" for the defender (i.e. $w_D^t(\theta_C)$ is decreasing in θ_C) since then $E\left[w_D^1(\theta_C) \mid \theta_C \in \left[\hat{\theta}_C, \bar{\theta}_C^2\right]\right]$ is decreasing in $\hat{\theta}_C$. Third and as in the baseline model, $P\left(\theta \geq \bar{\theta}_C^2 \,|\, \theta \geq \hat{\theta}_C\right)$ is increasing in $\hat{\theta}_C$ – that is, the defender's interim assessment that appeasement will be ineffective is higher when the challenger uses a higher threshold for transgressing.

Equilibrium Characterization Applying the analysis above, we now have the following complete equilibrium characterization.

Proposition 7. *Equilibria of the model with interdependent values are as follows.*

- *The deterrence equilibrium exists i.f.f.*

$$\bar{\beta}\left(\bar{\theta}_C^1\right) \leq P\left(\theta \geq \bar{\theta}_C^2 \,|\, \theta \geq \bar{\theta}_C^1\right)$$

- *The no deterrence equilibrium exists i.f.f.*

$$P\left(\theta_C \geq \bar{\theta}_C^2\right) \leq \bar{\beta}\left(-\infty\right)$$

- *A mixed strategy equilibrium in which the challenger uses threshold $\hat{\theta}_C^* < \min\left\{\bar{\theta}_C^1, \bar{\theta}_C^2\right\}$ exists i.f.f*

$$\bar{\beta}\left(\hat{\theta}_C^*\right) = P\left(\theta \geq \bar{\theta}_C^2 \,|\, \theta \geq \hat{\theta}_C^*\right)$$

  In the equilibrium, the defender responds to the transgression with war with probability $\alpha^* = \hat{\alpha}\left(\hat{\theta}_C^*\right)$.

The most important observation from the above characterization is the following: because $\bar{\beta}\left(\hat{\theta}_C\right)$ is interior for all $\hat{\theta}_C$ (meaning that war sooner is better than war later), our basic insight holds unaltered. When appeasement is ineffective $\left(\bar{\theta}_C^2 \leq \bar{\theta}_C^1\right)$, the deterrence equilibrium exists for all distributions over the challenger's type θ_C and functions $w_D^t\left(\theta_C\right)$ mapping the challenger's type into the defender's payoff from war that satisfy the initial assumptions. Thus, Corollaries 1 and 2 continue to hold unaltered with interdependent values.

Other more subtle patterns of equilibria can occur with interdependent values. Because $\bar{\beta}\left(\hat{\theta}_C\right)$ can be steeply increasing in $\hat{\theta}_C$ rather than constant, it is no longer the case that the mixed strategy equilibrium can only exist when both pure strategy equilibria exist. Many different scenarios can occur, including an odd number of mixed strategy equilibria combined with an even number of pure strategy equilibria (including none), and a single pure strategy equilibrium combined with an even number of mixed strategy equilibria.

Intuitively, the reason for this multiplicity of equilibria is that a higher threshold for transgressing by the challenger has two countervailing effects. First, it makes the defender *less* willing to appease because her interim assessment of the probability that the challenger is unappeasable is higher. Second, it makes the challenger *more* willing to appease because inferring the challenger is a higher type also means that war is worse, making appeasement more attractive if it can be effective. These countervailing effects can then generate multiple equilibria: with higher thresholds, the defender can find appeasement less likely to be effective, but simultaneously more desirable if it would be effective.

A.4 Game with two-sided uncertainty

The defender is now assumed to have a type θ_D upon which his war payoffs in each period depend, so we write $w_D^t(\theta_D)$ to express this dependence. We maintain the assumption that payoffs in peace for both players are fixed and common knowledge, and make new assumptions on the defender's type that mirror those of the challenger. Specifically, θ_D also belongs to an interval, has some prior distribution $g(\theta_D)$ with full support, and is distributed independently of θ_C. Thus, war values are private and each side's uncertainty may be interpreted as about the opponent's cost of war. We modify the assumptions the defender's payoffs as follows:

1. For all defender types, allowing the transgression makes the defender strictly worse off in both peace ($n_D^2 < n_D^1$) and war ($w_D^2(\theta_D) < w_D^1(\theta_D)\ \forall \theta_D$).

2. In each period t the defender's war payoff $w_D^t(\theta_D)$ is continuous and strictly increasing in θ_D. In addition, there exists a unique defender type $\bar{\theta}_D^t$ that is indifferent between peace and war in period t.

3. The benefit $w_D^1(\theta_D) - w_D^2(\theta_D) > 0$ of war sooner vs. war is weakly increasing in the defender's type.

The first assumption extends the properties of the transgression to a setting where the defender's payoffs can vary, and the second mirrors the assumptions made on the challenger's type. Importantly, it implies that with strictly positive probability the defender's threat is "inherently" credible in that he is willing to go to war solely to prevent the transgression. Formally, for both players let $\bar{\theta}_i^{s,t}$ denote a player indifferent between peace in period s and war in period t – since $n_D^2 < n_D^1$ we have $\bar{\theta}_D^{2,1} < \bar{\theta}_D^{1,1}$ and types in between are willing to fight a war over the transgression.

The third assumption ensures that types who are overall more belligerent are also weakly more willing to go to war for preemptive reasons, and is necessary for the existence of cutpoint strategies. Finally, since the defender may unilaterally wish to initiate war in both periods, we augment the first period with a final stage in which the defender can start a war even if the challenger chooses not to transgress. It is unnecessary to augment the second period with a similar stage because any defender type who would unilaterally initiate war in the second stage would also initiate war in the first stage and end the game.

Challenger Incentives Challenger incentives are identical to the game with interdependent war values except for the following distinction – because the defender may now be of a type $\theta_D \geq \bar{\theta}_D^1$ who would start a war whether or not the challenger attempts to transgresses, α now denotes the probability that transgressing would *provoke* an otherwise peaceful challenger to start a war. If the defender uses a cutpoint strategy $\hat{\theta}_D \leq \bar{\theta}_D^1$ for responding to the transgression, then in equilibrium $\alpha = \frac{G(\bar{\theta}_D^1) - G(\hat{\theta}_D)}{G(\bar{\theta}_D^1)}$.

Defender's Incentives The defender's war payoffs now depend on her type θ_D; moreover, because types are independent the threshold $\hat{\theta}_C$ that the challenger uses for transgressing only affects her payoffs through the interim assessment β that the challenger would initiate war after being allowed to transgress. He therefore prefers to respond to the transgression with war when $\beta \geq \bar{\beta}(\theta_D)$, where

$$\bar{\beta}(\theta_D) = \frac{n_D^2 - w_D^1(\theta_D)}{(n_D^2 - w_D^1(\theta_D)) + (w_D^1(\theta_D) - w_D^2(\theta_D))}. \tag{A.4}$$

It is simple to verify that for $\theta_D \in [0, \bar{\theta}_D^{2,1})$ (where $\bar{\theta}_D^{2,1}$ is the defender type indifferent between immediate war and successful appeasement) the function $\bar{\beta}(\theta_D)$ is strictly interior to $[0, 1]$ and decreasing (by assumption 3). The latter property ensures that the defender always plays a cutpoint strategy, and we can therefore also work with the inverse function $\bar{\theta}_D(\beta) = \bar{\beta}^{-1}(\beta)$ denoting the defender type indifferent between appeasement and war when his interim assessment is β.

Equilibrium Characterization

Proposition 8. *Equilibria of the model with two-sided uncertainty are as follows.*

- *The deterrence equilibrium exists i.f.f.*

$$\bar{\beta}(-\infty) \leq P\left(\theta_C \geq \bar{\theta}_C^2 \,|\, \theta \geq \bar{\theta}_C^1\right)$$

- *The no deterrence equilibrium exists i.f.f.*

$$P\left(\theta_D \in \left[\bar{\theta}_D\left(P\left(\theta_C \geq \bar{\theta}_C^2\right)\right), \bar{\theta}_D^1\right] \mid \theta_D \leq \bar{\theta}_D^1\right) \leq \hat{\alpha}\left(-\infty\right)$$

- *An interior equilibrium with challenger threshold* $\hat{\theta}_C^* < \min\left\{\bar{\theta}_C^1, \bar{\theta}_C^2\right\}$ *exists i.f.f.*

$$P\left(\theta_D \in \left[\bar{\theta}_D\left(P\left(\theta_C \geq \bar{\theta}_C^2 \mid \theta_C \geq \hat{\theta}_C^*\right)\right), \bar{\theta}_D^1\right] \mid \theta_D \leq \bar{\theta}_D^1\right) = \hat{\alpha}\left(\hat{\theta}_C^*\right)$$

or equivalently

$$\frac{G\left(\bar{\theta}_D\left(\frac{1-F\left(\bar{\theta}_C^2\right)}{1-F\left(\hat{\theta}_C^*\right)}\right)\right) - G\left(\bar{\theta}_D^1\right)}{G\left(\bar{\theta}_D^1\right)} = \hat{\alpha}\left(\hat{\theta}_C^*\right)$$

In the equilibrium, the challenger transgresses when $\theta_C \geq \hat{\theta}_C^*$ *and the defender responds with war i.f.f.* $\theta_D \geq \bar{\theta}_D\left(P\left(\theta_C \geq \bar{\theta}_C^2 \mid \theta_C \geq \hat{\theta}_C^*\right)\right) = \bar{\theta}_D\left(\frac{1-F\left(\bar{\theta}_C^2\right)}{1-F\left(\hat{\theta}_C^*\right)}\right)$.

Again, the most important observation from the above characterization is that because $\bar{\beta}\left(\hat{\theta}_C\right)$ is interior for all $\hat{\theta}_C$ (meaning that war sooner is better than war later), our basic insight again holds unaltered. When appeasement is ineffective $\left(\bar{\theta}_C^2 \leq \bar{\theta}_C^1\right)$, the deterrence equilibrium exists for all distributions over the challenger's type θ_C and defender's type θ_D that satisfy the initial assumptions, and Corollaries 1 and 2 hold unaltered.

As with interdependent war values other more subtle patterns of equilibria can also occur. Intuitively, the reason is that deterrence begets deterrence – a higher threshold for transgressing (greater $\hat{\theta}_C$) generates a higher interim assessment $\frac{1-F\left(\bar{\theta}_C^2\right)}{1-F\left(\hat{\theta}_C^*\right)}$ that the challenger is unappeasable, generating a lower threshold $\bar{\theta}_D\left(\frac{1-F\left(\bar{\theta}_C^2\right)}{1-F\left(\hat{\theta}_C^*\right)}\right)$ for the defender to respond with war, a higher probability $\frac{G\left(\bar{\theta}_D\left(\frac{1-F\left(\bar{\theta}_C^2\right)}{1-F\left(\hat{\theta}_C^*\right)}\right)\right)-G\left(\bar{\theta}_D^1\right)}{G\left(\bar{\theta}_D^1\right)}$ that the defender will be provoked by an attempted transgression, and thus more deterrence. Under some conditions this dynamic can set off a "deterrence spiral" where the challenger is very unlikely to be unappeasable ex-ante yet the deterrence equilibrium is unique – a sufficient condition for this occurring is the standard condition that the "no deterrence" equilibrium be unstable and the slope of the challenger's best response function

$$\hat{\alpha}^{-1}\left(\frac{G\left(\bar{\theta}_D\left(\frac{1-F\left(\bar{\theta}_C^2\right)}{1-F\left(\hat{\theta}_C^*\right)}\right)\right)-G\left(\bar{\theta}_D^1\right)}{G\left(\bar{\theta}_D^1\right)}\right)$$

be greater than 1 (where $\hat{\alpha}^{-1}(\alpha)$ denotes the well-defined inverse of $\hat{\alpha}(\theta_C)$).