# International Reputation with Changing Resolve[*]

Anne E. Sartori

Sloan School of Management, MIT

February 2, 2016

**Abstract**

This paper studies the effect of states' chaging levels of resolve on the formation and effect of reputations. When states' willingness to fight is very likely to change or differ between disputes, defenders do not acquire reputations for resolve that impact the behavior of later challengers. Surprisingly, however, some probability that a state's resolve will change actually motivates the formation of reputations, as states that anticipate the possibility of decreased resolve invest in reputations may be useful later. The paper helps to explain empirical findings that reputations are more likely to form and affect subsequent interactions when interactions occur with the same adversary, in the same region, or in disputes that are otherwise similar. It also illuminates states' attempts at reputation building in the face of emerging windows of vulnerability, and suggests a new hypothesis about the relationship between the quality of states' information about their adversaries' resolve and the formation and effect of reputations.

Word Count: 9,987

[*] Please do not cite without permission of the author. I thank John Kennan, Muhammet Bas, Alexandre Debs, Olivier Henripin, Lisa Martin, Bob Powell, Dustin Tingley, George Yin, and seminar audiences at the University of Michigan, Duke University, Yale University, Harvard University, the University of Iowa, the University of Wisconsin-Madison, the Buffett Center at Northwestern University, the American Political Science Association Annual Meeting, the Midwest Political Science Association Annual Meeting, and the Political Economy in the Chicago Area 2011 meeting for helpful comments. This paper is my answer to a question that Paul Huth asked me many years ago.

...it was assumed that the British Government's actual interest in the Falkland Islands was slight and that were it not for the islanders' lobby they would have been abandoned some time ago *along with Britain's other former colonies.* –Freedman and Gamba-Stonehouse (1991)[78], italics added[1]

In what many commentators have described as a "last imperial hurrah,"
Britain re-invaded the Falkland Islands, and after bitter fighting expelled
the Argentine invaders.–Howkins (2010)[261]

States scan each other's past behavior for clues to their current intentions. We call these clues reputations.[2] However, as the Falklands example illustrates, the information contained in reputations is not always correct; a state with a reputation for lacking resolve may fight, or a state with a reputation for having it may fail to stand firm. States fail to live up to their reputations because no two disputes are identical; the adversary and the military balance may be different; the domestic political situation may have changed. But how and when do states acquire reputations for their resolve, if resolve is itself a dynamic quality?

This article studies the formation and effect of reputations for resolve in a world in which a state's resolve may differ or change between situations. I show that states are likely to obtain reputations for their resolve, or willingness to fight, only when disputes are sufficiently similar in a variety of ways, as were Britain's many disputes with former colonies. However, the fact that a state's resolve may differ in future situations does not always inhibit reputation formation. On the contrary, the possibility that a state's resolve may change between disputes provides a positive incentive for states to engage in the behaviors that lead to them to acquire reputations; as I explain later, as disputes become extremely similar, states become less, not more, likely to acquire such reputations.

By exploring the impact of changing resolve on the formation of reputations, this article helps to reconcile theoretical arguments that explain the formation of reputations with those

[1] Also see Weisiger and Yarhi-Milo (2015)[478].

[2] I provide a more careful definition later.

that argue that the past should have no effect; they arise as special cases of a model of international politics in which the degree of similarity between disputes varies with the disputes in question. It also explains findings in the empirical literature that reputations are more likely to form and affect subsequent interactions when interactions occur with the same adversary, in the same region, or in disputes that are otherwise similar, and points to other areas for empirical research.

On the theoretical side, some scholars, drawing on models developed in economics, argue that reputations influence the credibility of a state's threats because a state's decision to stand firm in one situation or at one time signals its overall level of resolve, or willingness to fight (Nalebuff (1991), Huth (1997)[88], Tingley and Walter (2011)).[3] When a state has a reputation for resolve, others believe it more likely to stand firm in the current dispute, so that its deterrent threats are more credible. Other scholars counter that credibility is "context dependent," mostly or entirely determined by the interests at stake and the military capabilities, which change between one conflict and the next (Maxwell 1968; Press 2005). Rather than basing their decisions upon states' past behavior, adversaries thus make a "current calculus" of power and interests in determining how to proceed, based on observable, material characteristics (Press 2005).

On the empirical side, many but not all works that look to see if states' past behavior influences their current interactions find evidence consistent with the rational theory of reputations (e.g. Crescenzi (2007), Dafoe and Caughey (2016)). Important works find little or no evidence of reputations (Hopf 1994; Press 2005).[4] However, as Huth (1997) argued, taken as a whole, the evidence suggests that reputations are most likely to influence interactions if they were formed based on actions within the same region or the same adversary.[5] More recent work elaborates on this pattern Huth identified, suggesting that reputations have an effect on some future disputes, but mostly ones that are linked in one of a number of ways

[3] Exactly what a state is signaling depends on the model; for example, Alt, Calvert, and Humes (1988) talk about a reputation for "toughness." For an alternative, psychological theory of reputation formation, see Mercer (1996).

[4] Also see Mercer (1996), who argues that his evidence is inconsistent with a rationalist theory of reputations.

[5] See also Snyder and Diesing (1977)[187].

to the past situation (Crescenzi 2007; Weisiger and Yarhi-Milo 2015).[6] Overall, the half-full glass makes it possible for some to claim that reputational theory is incorrect and others to claim that it is important.[7]

In one sense, the debate over context dependence misses the boat. Of course resolve is (to some extent) context dependent, but that does not mean that states easily can assess their adversaries' resolve by performing a calculation. A portion of states' resolve is unobservable, and states have incentives to misrepresent what is unknown about their resolve (Fearon 1995). The imperfect observability of resolve is why crises often involve probes, such as the confrontation between the U.S. and the Soviets at "Checkpoint Charlie" during the Berlin Wall Crisis of 1961 (Brecher and Wilkenfeld 1997, 350-51). Scholars have posited a number of mechanisms by which states learn about the unobservable part of each other's resolve; one of these, in the context of repeated interactions, is reputation.[8] As Weisiger and Yarhi-Milo (2015)[474] have pointed out, reputation involves estimates of the opponent's interests based on past behavior; it also can involve estimates of other components of resolve, including capabilities and public opinion.

Yet, disputes do differ from each other in both observable and unobservable ways, and the extent to which a state's resolve changes between disputes is likely to depend upon the circumstances.[9] The situations in which works often have found evidence that a state's past behavior impacts its current interactions – multiple disputes in the same region and disputes in which the adversaries are more similar in power or interests – seem to be situations in which there is likely to be at least some continuity in a state's resolve from one interaction

[6] Lieberman (2013) argues that states can get reputations for their resolve, but that these reputations are only helpful when the context is similar. See, for example, page 116. Walter (2006) also finds evidence of reputations when governments fight separatist groups, a situation in which multiple disputes seem to have similarities. The pattern is not entirely clearcut, however; Press (2005)'s negative findings on reputation come from repeated crises involving Nazi Germany.

[7] Also see Clare and Danilovic (2010) for an interesting recent argument that states act to rebuild their reputations, reminiscent of Jervis (1970).

[8] Most recent models of bargaining and war assume private information about factors related to a state's willingness to fight because this willingness is imperfectly observable.

[9] Dafoe, Renshon, and Huth (2014)[372] also argue for the importance of investigating "when, how, and why" reputations and the related concept of honor matter, rather than that of whether they matter.

to the next, suggesting that the degree of context dependence may play a role in determining whether or not states' behavior leads them to acquire reputations or whether or not reputations affect the current interaction.[10]

The present paper adds to our understanding of international politics by showing with a single model why context dependence can but does not always impede the development and influence of reputations for resolve (and, in fact, can motivate the formation of reputations), and also why reputations are more likely to form in the situations in which they have more often been found: when disputes occur in the same region or are more similar in other ways, and when repeated disputes involve the same dyad over time. The paper studies a two-period crisis bargaining model with two novel assumptions. First, the model assumes that the unobservable part of a state's resolve (its "private resolve") is partly context dependent; it is correlated, but not identical, between disputes. The defender may have a greater or lesser unobserved willingness to fight in the second dispute because of how it evaluates the differing stakes, for military reasons, or for other reasons outside the model. The model parameterizes the extent to which private resolve changes between disputes, so that comparative statics shed light on how assumptions about context dependence influence the conclusions we draw about reputations. Second, the model assumes that the challenger with whom the defender interacts can, but need not, change between disputes. Again, the probability of change is parameterized, so that comparative statics illuminate why reputations are more likely to be important when the same two states interact repeatedly.

Like much of the previous formal research in international relations, I draw on a model from economics (Kennan 1998). The contribution of my paper is substantive; Kennan's excellent paper does not mention politics, and, in fact, contains very little text other than that used in proofs (Kennan 1998).[11] To investigate the importance of repeated interactions in

[10] Disputes in the same region are situations in which, all else equal, the defender's private resolve is more likely to persist over time because they are more likely to involve similar interests and a similar need to project power.

[11] Kennan has a closely related, published paper on reputations that also does not discuss politics or international relations except to say that "repeated contracts also arise in international trading relationships" (Kennan 2001, 719).

The closest works in political science are Alt, Calvert, and Humes (1988) and Morrow (1989), both of the same dyad, I add the possibility that the uninformed actor, in my case the challenger, can change between disputes. To investigate different kinds of context dependence, I also analyze a version in which the overall magnitude of a dispute (both observable and unobservable parts of the defender's resolve) can change in expected or unexpected ways between disputes. I also provide a more accessible summary of the logic of the proofs and perform comparative statics and a welfare analysis.

The analyses provide a foundation for both the credibility-is-context-dependent and the reputations-affect-behavior schools. As the context-dependence position asserts, reputations form and affect future behavior only when private resolve is not too transient; the information that the challenger learns when the defender fights in one dispute must be sufficiently relevant to the new situation for the reputation to impact the challenger's decisions. This result helps to explain why many of the "pro-reputational" findings in the literature (findings consistent with the idea that states' behavior can lead to reputations that influence their adversaries' decisions in subsequent situations) occur in studies in which disputes occur in the same region, or involve reputations forming when adversaries are more similar in power or policy. When observers see the defender stand firm, the information they learn about its private resolve is sufficiently relevant to other disputes in the same region, or other similar disputes, that it affects the way they treat the defender in those disputes.

However, consistent with reputational theory, the analyses also show that the defender can acquire a reputation for resolve that influences challengers' subsequent behavior as long as the persistence of private resolve passes a threshold. Resolute defenders have an incentive to acquire reputations for their resolve precisely because they may *not* be so resolute in their next disputes; if so, they may be able to lean on their reputations, rather than their lackluster willingness to fight, to attain their goals.

Moreover, some positive probability that private resolve will change between disputes (some degree of context dependence) actually facilitates reputation formation because, for a

which model reputation formation with changing resolve or toughness. In those papers, today's resolve does not depend on yesterday's; they are related in that they are different draws from the same data generating process. I assume instead that a state that is more resolute than average today is more likely to be resolute tomorrow.

resolute defender to acquire a reputation for resolve, an irresolute defender must accept a low offer and forego such a reputation. Foregoing this reputation is costly for the defender, and in circumstances in which a resolute defender acquires a reputation, the challenger compensates the irresolute defender for its failure to acquire one by making a somewhat higher low offer.[12] The more an irresolute defender believes its low resolve will persist (and so it will not want to fight in the future), the greater the value of a reputation, and the more compensation it requires for failing to acquire one. In situations in which the defender's resolve is too persistent, a defender that is irresolute today is impossible to tempt with an offer that is low enough that the challenger is willing to make it. For this reason, the relationship between the degree of context dependence and the formation and effect of reputations is non-monotonic; reputations are more likely to matter when private resolve is moderately context dependent.

The model also helps to explain why many of the empirical findings that are consistent with reputational theory come from studies of situations in which the same dyad interacts repeatedly. Reputations are more likely to form when the same dyad interacts repeatedly because these are situations in which reputations are valuable to the challenger as well as the defender. Ordinarily, scholars characterize reputations for resolve as valuable to the states that possess them. I show that while a defender's reputation can benefit that state, it also can benefit the challenging state. When the challenger believes itself likely to interact with the same defender again, information about the defender's resolve that is carried in its reputation helps the challenger to know the best way to act toward the defender in the future. For this reason, a challenger is more likely to initiate costly probes that can lead to reputation formation when it expects to interact again with the same defender. When different challengers interact with a defender over time, the defender may nevertheless develop a reputation, but this is less likely because reputation formation benefits only the defender and not the challenger.

Finally, the analyses have implications about the relationship between uncertainty and reputation formation, suggesting a new avenue for empirical research. It is tautological that reputations about the unknown portion of a state's resolve require some uncertainty.

[12] See Sechser (2010)[629] for a similar argument about compensation for reputations.

However, as I explain later, states acquire reputations only when the uncertainty about their resolve is not too great.

A caveat about context dependence is in order: there are various ways in which the stakes may be higher in a future dispute than in today's, and these have different impacts on reputation formation. When states anticipate that tomorrow's issue will be of a substantially larger magnitude (both observable and unobservable parts of the stakes are substantially amplified), reputations are so valuable that reputation formation can be impeded. Thus, for example, insofar as US decision makers anticipated crises in Europe when they fought over Korea, their actions were unlikely to lead to a reputation for resolve. When the substantially greater magnitude of a future dispute is unanticipated, however (as perhaps was the 2014 conflict with Russia over the Ukraine), a reputation that already was formed in prior disputes is likely to impact that dispute, ceteris paribus. Of course, if a high-stakes dispute involves little uncertainty (the adversary is simply known to be resolute), states do not need to assess their adversaries' resolve, and reputations do not play a role. Similarly, when circumstances suggest that the unobservable part of the defender's private resolve in the unexpected, high-stakes dispute is probably very different from its resolve in the earlier crisis, the reputation the defender has acquired will have little impact.

The article proceeds as follows. The next section defines resolve and reputation. The third describes the model. The fourth analyzes behavior in the second of the two disputes when the disputes are similar in magnitude, in order to give a sense of the value and impact of a reputation. The fifth discusses the implications of the model about the circumstances under which reputations do and do not form and affect the challenger's subsequent behavior. It also discusses how the implications differ when the subsequent dispute is (either expectedly or unexpectedly) of substantially greater magnitude. The sixth section briefly discusses the relationship of the model to windows of vulnerability, and empirical evidence more broadly, and the final section concludes. The appendix contains the logic of the proofs, comparative statics, welfare analysis, discussion of the equilibrium refinement, and tables that summarize the results of Kennan's (1998) proofs but with the re-parameterization I use in this paper.

## 1. What is a reputation for having or lacking resolve?

Before turning to the question of how the changing nature of resolve affects the formation and effect of reputations, I explain what I mean by those terms. "Resolve" in this article is synonymous with "value for war," which strongly affects a country's willingness to fight. Scholars differ in what constitutes a state's value for war; in my view, it is affected by many factors, prominent among which are its military capabilities and skill at using them, its level of interest at stake, and both leaders' and publics' support for war as a means of achieving objectives. "Reputation" for resolve is another state's belief about that part of a state's resolve about which others are uncertain (its private resolve), when that belief has been influenced by the state's past behavior.

In this paper, as I discuss later, I am concerned with the formation of "consequential reputations for resolve" by the defender, beliefs about a defender's level of private resolve that have been influenced by its past behavior and affect a challenger's future actions. The "consequence" of a reputation differs depending on what a challenger is expected to do in the future if the defender's actions today contained no information. In the work that follows, I am considering a consequential reputation that allows the defender to improve what it expects to be a bad situation in a later dispute by demonstrating its toughness in the present one. This differs from a situation in which irresolute defenders fight to avoid revealing their weakness (as in economic models such as the chain-store paradox; see, e.g., Morrow (1994)[284-90]).

## 2. The Model

I study a simple model of two consecutive international interactions, or disputes, following Kennan (1998). In each dispute, two countries, a "challenger" and "defender," bargain over a good; if they do not reach agreement, they engage in conflict, which I usually will call "war." While the representation of bargaining and war is standard, the model contains the two new features discussed earlier. The first is the relationship between the defender's resolve in the two disputes. As is common in bargaining models of war, the defender has private information about its value for war (resolve) which may be either high or low. However, in this model, the unobservable part of the defender's resolve (its type) is partially persistent between disputes, and the probability with which it stays the same or changes can vary (is parameterized). Thus, at the time of the first dispute, the defender knows its willingness to fight in that dispute, but this does not fully determine its willingness to fight in the second. Because it knows its own resolve in the present, it has more information than the challenger about what its resolve is likely to be in the future, but does not know its future resolve with certainty.

Why assume that private resolve is partially persistent? The persistent part of a state's private resolve represents characteristics, such as the unknown part of military capabilities, that often have a lot of continuity, and/or characteristics that apply to a class of cases.[13] The transient part of a state's resolve comes from components that are more rapidly changing, sometimes domestic political support for war, and from the fact that a subsequent dispute might involve substantially different interests and a different adversary. For example, one could argue that Britain's resolve to maintain former colonies was linked (persisted at least partly from one to the next). However, its resolve in one case would not automatically determine its resolve in another; public opinion could differ, or its willingness to fight might depend on the strength of the adversary.

The second new feature in the model is that the defender and the current challenger may or may not interact again in the second dispute; again, the probability with which they do so can vary. If the defender does not interact with the same challenger in the second, it interacts with a new one.

As I explain in detail soon, the basic model assumes that the disputes are of the same general magnitude, though the defender may be more or less willing to fight in the second situation for various reasons, including the possibility that it considers the stakes to be somewhat higher. I later adapt the model to explore the possibility of reputation formation when the stakes in the second dispute either are expected to be much higher, or unexpectedly turn out to be so.

[13] See Meirowitz and Sartori (2008) on states' incentives to hide military capabilities.

Modeling two consecutive interactions allows me to study the conditions under which the defender's behavior in the first dispute leads to a reputation that impacts the challenger's behavior in the second dispute. In particular, I study how the defender's decision to fight in the first dispute can lead the challenger to be more accommodating in the second, and the circumstances under which this is more or less likely to be the case. The new features allow me to explore how a) the stability of resolve (the probability that a state that is resolute today is resolute tomorrow), and b) the likelihood of ongoing interactions between the same states, affect the probability that the defender's behavior in the first dispute leads to a reputation that affects the challenger's behavior towards it in the second.

The remainder of this section describes the model and its assumptions in more detail.

### 2.1. Actors, Sequence of Moves, and Payoffs

Each of the two ($t = 1, 2$) disputes begins with Nature's move, which represents circumstances arising over which the states dispute. Nature draws a number that determines the unobservable part of the defender's value for war in that dispute ($\eta$), which is either low ($\eta = 0$) or high ($\eta > 0$). The observable part of the defender's value for war is $\theta$, so that its total value is low ($\theta$) or high ($\eta + \theta$), as a result of Nature's draw. The defender is resolute in the first dispute with probability $h_1$. If a defender is resolute/irresolute in the first dispute, the probability that it is resolute/irresolute again in the second is $\rho$ with $\frac{1}{2} < \rho < 1$. Thus, the parameter $\rho$ models the persistence of private resolve. Alternatively, $1 - \rho$, the probability that the defender's private resolve differs in the next dispute, represents one kind of context dependence. The assumption that $\rho > \frac{1}{2}$ means that an irresolute defender is more likely to be irresolute than resolute in its next dispute (and a resolute defender is more likely than not to remain resolute); this does not turn out to be sufficient for reputation formation.

After Nature moves, the states bargain over an issue or territory of value $\Pi$ using the standard ultimatum-game bargaining protocol: The challenger makes an offer $a_t$. Then, the defender chooses to accept or reject the offer; if it rejects the offer, the states go to war. If the defender accepts the offer, its per-period payoff is $a_t$ and the challenger's is $\Pi - a_t$. The challenger's payoff from war is normalized to zero.

After the first dispute ends with an accepted bargain or war, the defender engages in another dispute. With probability $\lambda$ ($0 < \lambda \leq 1$), the challenger in the second dispute is the same; otherwise, the defender interacts with a new challenger. If the same two states participate in both disputes, both states' payoffs are the discounted sums of the two stage-game payoffs, with common discount factor $0 < \delta < 1$. If the first challenger does not interact in the second dispute, its second-period payoff is normalized to zero; thus, the first-period challenger's expected continuation value is $\lambda$ times the expected payoff of a second dispute in which it participates.[14] If a new challenger is involved in the second dispute, its payoff is its payoff from that dispute.

### 2.2. Information, Beliefs, and Equilibrium Concept

Nature's move at the start of each dispute is observed only by the defender. Thus, the defender has private information about its value for war in each time period. Since the challenger does not know whether the defender has a payoff of $\theta$ (irresolute) or $\eta+\theta$ (resolute), the challenger's uncertainty is about a small part of the defender's payoff when $\eta$ is small relative to $\theta$ and about a large part when it is large.

All other aspects of the game are common knowledge, including the persistence of resolve ($\rho$). For example, a defender that is resolute in the current context knows that it will be resolute in its next dispute with probability $\rho$; if a challenger learns that a defender is resolute today it knows that the defender will be resolute in the next dispute with probability $\rho$. I let $\zeta_t$ be the challenger's belief at the beginning of time $t$ about the probability that the defender is resolute in that period. If a new challenger is involved in the second dispute, that challenger observes behavior in the first dispute before engaging in the second.

Because of the uncertainty in the model, the appropriate equilibrium concept is one analogous to Perfect Bayesian Equilibrium (PBE), in which threats and promises are credible, strategies are rational given beliefs, and beliefs are updated according to Bayes' Rule whenever possible. In this game, second-period beliefs are jointly determined by Bayesian

[14] The challenger does not choose whether or not to be involved in a future dispute in this model, so this normalization of its second-period payoff to zero if it does not interact in that period is inconsequential.

updating of beliefs following first-period play and by the transition probability of resolve. The appendix discusses the use of forward induction to rule out equilibria with unreasonable off-the-equilibrium-path beliefs.

### 2.3. Additional Assumption: War is Costly

Formally, I assume that $\Pi - (\eta + \theta) > 0$.[15] This assumption implies that, absent reputational considerations, a bargain exists that both players prefer to war in period $t$, even if the defender is resolute in that dispute, though the states may never agree upon such a bargain.

### 2.4. Varying the Stakes

Scholars have questioned whether reputations can affect adversaries' behavior when different disputes involve different stakes. The basic model captures one way in which the stakes can change: the defender may have a different value for war in the second dispute; the part of its value for war about which it has private information (which includes any private information about the stakes) is higher or lower than before. However, that model assumes that the stakes are constant/the context is constant in two ways: the maximum possible value of a defender's value for war (observed plus unobserved components) is the same in the two disputes, as is the observed component of the defender's resolve. That is, any difference in the defender's value for the stakes between the two disputes is simply captured by whether its private resolve is high or low.

To further investigate the effects of changing stakes, I vary the model, first by assuming that a higher-stakes dispute is expected, and then by assuming it occurs but is unexpected. This distinction is important because reputation formation is forward-looking. States invest in their reputations because they anticipate that the reputations may help them to attain goals in future situations. As I show later, they also initiate probes that allow adversaries to form reputations because this information may be valuable in the future.

The first variant assumes that the states in the first dispute expect the second to be of

[15] The assumption that war is costly is common in the formal literature on bargaining and war; e.g., see Fearon (1995).

larger magnitude ("bigger" in all ways); to do this, I multiply the payoffs $\Pi$, $\eta$, and $\theta$ in the second dispute, which represent the different components of the states' values for the disputed issue (observed and unobserved), by a constant greater than one. Mathematically, this is equivalent to assuming that the discount factor is greater than one, so I investigate what happens when the assumption of discount factor ($\delta$) less than one does not hold.

The second variant assumes that states in the first dispute expect their next dispute to be of similar magnitude to the current one; thus, the incentives for reputation formation are identical to those in the basic model. However, when the second dispute arises, it turns out to be greater magnitude, one in which the payoffs $\Pi$, $\eta$, and $\theta$ are multiplied by a constant greater than one. The question, then, is whether any reputation acquired in the first dispute has an effect in the second dispute, which unexpectedly concerns a more vital state interest.

## 3. The Value of a Reputation for Resolve; Who Benefits?

During the 1999 NATO bombing campaign in Yugoslavia, Richard Betts was quoted as saying, "What's changed is the reputation and the honor of the administration and NATO" are at stake; "We've gotten into a war, we've committed our resources and our reputation to trying to do something" (Harris 1999). Betts' quote suggests that he believed that the US reputation was valuable. Was he right? When a state obtains a reputation for resolve, is there a benefit, and to whom?[16]

Surprisingly, the answer is complicated. As policymakers suggest, a reputation for having resolve is valuable to the defender. However, the defender's reputation also can benefit its *challenger* by providing it with information about the defender's resolve that, while imperfect, helps it better to tailor its policies to the type of defender that it faces. This section explains why.

If the defender acquires a reputation, it does so in the first dispute, and any effect of the reputation is felt in the second. As background, I begin by discussing why the challenger's beliefs at the start of the second dispute determine the defender's payoffs in that dispute, which depend on whether it is more-resolute or less, and also why, under most circumstances,

[16] The discussion in this section is based on the "welfare analysis" section of the appendix.

the challenger's beliefs also determine its own expected payoff.[17] Behavior in the second dispute is typical of screening games – the challenger makes a low offer if it believes it is likely to be interacting with a defender that will accept a low offer (the challenger is optimistic), and it makes a high offer if it believes it is likely to be interacting with a defender that will reject a low offer (it is pessimistic). If its beliefs are such that it is exactly indifferent, then it may choose one offer all the time or randomly make a high or a low offer.

Specifically, since the second dispute is the last period, it is equivalent to a one-period game. The challenger begins this game believing that there is some probability $\zeta_2$ that the defender is resolute. Each type of defender rejects any offer less than its value for war, since rejecting the offer leads to war. The challenger can make a pooling offer, a high enough offer that both resolute and irresolute defenders will accept it, or a screening offer, a low offer that only an irresolute defender will accept, so that it "screens" out which type is which. Its optimal offer when screening will be $a_2 = \theta$ (the lowest an irresolute defender will accept) and when pooling will be $a_2 = \eta + \theta$ (the lowest a resolute defender will accept), because giving away more does not increase the probability that the offer is accepted. If the challenger offers $a_2 = \theta$, it expects this offer to be accepted with probability $1 - \zeta_2$, the probability that it faces an irresolute defender, but if it offers $a_2 = \eta + \theta$, it expects it always to be accepted. The challenger's expected payoff in the case of a screening offer is $\zeta_2 * 0 + (1 - \zeta_2)(\Pi - \theta)$, and its payoff in the case of a pooling offer is $\Pi - \eta - \theta$, so that it makes a low, screening offer when

$$\zeta_2 < \frac{\eta}{\Pi - \theta} \equiv \zeta_2^*.$$

Similarly, the challenger makes a pooling offer when $\zeta_2 > \zeta_2^*$ and is indifferent between the two offers when $\zeta_2 = \zeta_2^*$.

[17] This belief determines the challenger's expected payoff except for a knife-edge case when the belief is precisely equal to a threshold and the resolute defender has a number of mixed strategies that it can play in equilibria. While it sounds strange that the challenger's belief determines its payoff, note that this belief contains information about the defender's type, and also influences the challenger's present behavior.

### 3.1. The Defender Benefits, but Mostly When It Is Irresolute

Because the defender gets a higher offer when the challenger believes it more likely to be resolute (an offer of $\eta + \theta$ rather than $\theta$), these beliefs can be valuable to the defender. When the defender's actions in the first dispute lead the challenger's beliefs at the start of the second dispute to exceed the threshold (so that it makes the higher offer), when otherwise the beliefs would have been below the threshold (it would have made the lower offer), I consider the defender to have acquired a "consequential" reputation for resolve.

The defender benefits in the second dispute from a reputation for resolve if and only if it is irresolute in that dispute.[18] A defender that is irresolute in the second dispute is better off by $\eta$ with pooling than with screening, because it is offered and accepts $\eta + \theta$ with pooling and only $\theta$ with screening. A defender that is resolute in the second dispute receives the same payoff, regardless of the challenger's beliefs at the start of that dispute, because the high offer is equal to the resolute state's value for war. However, the defender in the first dispute has an incentive to acquire a reputation for resolve even if it is resolute because resolve is only partially persistent; it will reap the benefit in the less-likely event that it turns out to be irresolute in the second dispute.

All models are simplifications; in this one, a reputation for resolve does not benefit the defender at all when it turns out to be resolute in the second period because a defender in that situation is just as happy to fight as to accept the better bargain. This result, in this extreme version, is an artifact of the bargaining protocol in this model. However, a reputation for resolve is likely to be at least as valuable to a defender when it is irresolute as when it is resolute with many other protocols, because the resolute defender is (at least weakly) better off than the irresolute one without the reputation. Thus, the fact that resolute defenders may turn out to be irresolute in future disputes provides an incentive for them to invest in their reputations for resolve.

[18] This result is based on the assumption that the challenger will make a low offer in the second dispute if it learns nothing about the defender's type in the first dispute. If the defender's reputation were, instead, to convince a challenger that intended to be generous in the future to become more stingy, that reputation would make the defender worse off.

As I discuss later, scholars often observe states that are resolute acting to maintain or develop reputations for resolve. This work suggests that they do so because they know that their resolve may weaken – and, if so, the reputations they form today may well come in handy.

### 3.2. The Challenger Also Benefits

Not only the defender benefits from its reputation; when conditions are such that reputations form, the challenger benefits, too, by tailoring its behavior in the second dispute to the type the defender is more likely to be. Remember that, in the second dispute, the challenger would like to make a low offer ($\theta$) to an irresolute defender, since an irresolute defender is willing to accept that offer, and a high offer ($\theta + \eta$) to an resolute defender, which would reject the low offer in favor of war. As I explain later, the defender acquires a reputation for being resolute by standing firm in the first dispute, and does so only if it is resolute in that dispute. The defender's past actions provide information about its past type, which in turn provides information about its present type. This allows the challenger to make a low offer when the defender is more likely to be irresolute and a high offer when it is more likely to be resolute.

### 3.3. Limit on Context Dependence

My goal is to understand the conditions under which the defender obtains a "consequential reputation," a reputation that affects the challenger's behavior by leading it to make a high offer, when it would make a low one if the defender had no reputation. If this is ever to be the case, it must be when the challenger that observes resolute behavior in the first dispute believes the defender to have been resolute with certainty in that first dispute. Thus, a necessary condition for the formation of consequential reputations is that the challenger make a high (pooling) second-period offer when first-period actions reveal the defender definitely to have been resolute in the first dispute, but a low offer when first-period actions reveal no information.[19] If the challenger learns from the first dispute that the defender was resolute,

[19] This is Kennan's Condition O.

its belief that the defender is resolute at the start of the second is $\rho$. (This is the definition of $\rho$.) We saw earlier that a challenger benefits from making a generous offer in the second dispute when its belief $\zeta$ is greater than the threshold $\zeta_2^*$. Thus, the persistence of resolve $\rho$ must be at least $\zeta_2^* = \frac{\eta}{\Pi - \theta}$ for consequential reputations to form; put differently, resolve cannot be "too" context dependent.

## 4. Results: When Past Actions Matter

> [R]eputation was clearly worth fighting for. Israel's resolve reputation after 1956 created a period of 11 years of stability in which Nasser was on many occasions under pressure to attack yet he refrained from a challenge to deterrence.
>
> Although the border between Israel and Egypt was peaceful for 11 years, this was not the case on Israel's other border with Jordan and especially Syria. –Lieberman (2013)[116]

States' past behavior influences current events, but not always. As I noted earlier, both reputational and context dependence arguments arise as cases of the model that I study. In one equilibrium, defenders acquire reputations that affect adversaries' future behavior; in another, the states' behavior does not involve the formation of reputations.[20] Comparative statics indicate when states are likely to be in one situation or the other. I briefly describe these equilibria and my method of performing comparative statics before turning to the implications about when past behavior affects the present.

In the equilibrium with reputations, defenders that are resolute in the first dispute obtain consequential reputations by standing firm. Specifically, the challenger makes a low, screening offer in the first dispute; when the defender is resolute, it rejects the offer, fights, and

[20] See the appendix for more information about the equilibria. In addition to the equilibria that I use for my analyses, the game has a continuum of pooling equilibria that do not survive forward induction, as I discuss below, and a partial screening equilibrium that involves the formation of reputations that do not affect the challenger's behavior.

obtains a reputation for resolve, but when it is irresolute, it accepts the offer and obtains a reputation for lacking resolve. The challenger makes a high offer in the second dispute to a defender that has a reputation for resolve, and a low offer to a defender without such a reputation; when the defender is irresolute in the second dispute, it accepts either offer, and when it is resolute, it accepts only the generous one. (Because the defender's resolve may change between disputes, a challenger may "mistakenly" make a generous offer to an irresolute defender in the second dispute, or too low of an offer to a resolute one, despite the information carried by the defender's reputation.)

In the equilibrium without reputations, the challenger makes a high, pooling offer that all defenders accept in the first dispute, so it learns nothing about the defender's resolve from the defender's acceptance of its offer. The challenger makes a low offer in the second dispute; if the defender is irresolute, it accepts this low offer, but if it is resolute, it rejects the offer and the states go to war.

Earlier, I discussed how the challenger's behavior in the second dispute depended on its level of optimism about the defender's resolve. The equilibria that this section describes begin with behavior in the first dispute, but behavior depends on an analogous threshold; when the challenger's belief about how likely the defender is to be resolute is below the threshold, the challenger risks a low offer, or probe; this is the circumstance in which we observe reputations for resolve. If the challenger is more pessimistic, it dares not risk a low offer, and we do not observe reputations for resolve. Which of these equilibria exists depends on the challenger's level of optimism at the start of the first dispute – that is, its belief $\zeta_1$ about the probability with which the defender is resolute at the start of the first dispute. As Figure 4.1 shows, when the challenger is more optimistic, it makes the screening offer (the screening equilibrium exists), and when it is quite pessimistic, it makes the high, pooling offer (the pooling equilibrium exists). As long as the distribution of initial beliefs is continuous in the real world, varying a threshold varies the probabilities with which the model predicts we will observe the behavior associated with the equilibria on either side of the threshold. When $\zeta_1^T$ is higher, the model predicts that screening occurs more often (for a higher range of initial beliefs) and pooling occurs less often (for a lower range of initial

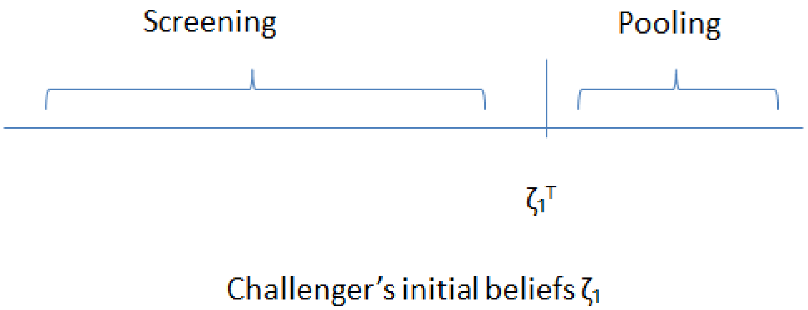

Figure 4.1: Equilibria as a Function of the Challenger's Belief at the Start of the First Dispute about the Probability that the Defender is Resolute

beliefs).

The position of the threshold depends on the exogenous variables in the model, in particular on the persistence of resolve, the probability with which the same two states will interact in the second dispute, and the amount of information that the challenger has about the defender's resolve.[21] To investigate the conditions under which reputations form and affect behavior, I consider how a change in a parameter affects the threshold $\zeta_1^T$; for example, when the persistence of resolve ($\rho$) is high enough that it satisfies the reputation condition, increasing $\rho$ decreases $\zeta_1^T$, making screening and the associated reputation formation less likely.

Modeling the possibility that resolve changes between disputes leads to insights into why reputations are more likely to form and affect adversaries' later behavior under some circumstances than others. The sections that follow discuss three topics: the complicated relationship between context dependence and reputations, the reasons why reputations are more likely to form in repeated interactions within the same dyad, and the effect of uncertainty on reputations.

[21] The threshold is a function of the parameters $\rho$, $\Pi$, $\delta$, $\theta, \eta$, and $\lambda$.

### 4.1. The Effect of Context Dependence

Earlier, I explained why we should only observe the effect of defenders' reputations for resolve when the defender's private resolve is sufficiently likely to persist to a future dispute; a minimum level of persistence is necessary for the information learned in one dispute to be sufficiently relevant to the next. This might be the case, for example, when a state faces a series of disputes over decolonialization or with former colonies. While not particularly surprising, this logic helps to explain two of the empirical regularities about reputations – that they are more likely to matter within multiple disputes in the same region, and that they are more likely to matter when adversaries are more similar, including similar in power and/or policy – which involve situations in which a state's resolve is more likely than average to persist between disputes.

More surprising, perhaps, is that context dependence, or the possibility that resolve will change, also facilitates the formation of reputations. Beyond the minimum persistence of resolve that is needed for reputation formation, as a state's resolve becomes yet more likely to persist to the next dispute, the defender becomes *less* likely to acquire a consequential reputation for resolve; that is, at some point, more context dependence makes it more likely that the defender acquires a reputation. This result is due to strategic interaction in the dispute in which the defender possibly acquires a reputation. In order for a defender to acquire a reputation for resolve by rejecting an offer, rejection must provide the challenger with information that the defender is more resolute than the challenger had previously thought. By Bayes' rule, for rejecting the offer to provide this information, the defender must be willing to accept the offer at least some of the time when it is irresolute; in doing so, it foregoes acquiring a reputation for resolve. Remember that a reputation will turn out to be valuable to a defender that is irresolute today if it is irresolute again in its next dispute. As resolve becomes less likely to change, the expected value of a reputation to a state that is irresolute goes up; for this reason, it requires greater compensation in the present dispute to be willing to forego the reputation. As the low, reputational offer becomes more expensive to the challenger, the challenger becomes less willing to make that offer. Ultimately, reputation formation is more likely when resolve is not too likely to persist because it is less costly to the challenger in this situation.[22] Since models are simplifications, I would not expect a complete absence of reputations below a threshold in reality; rather, I would expect a nonlinear relationship, with reputations more likely to form and affect subsequent disputes when resolve is moderately likely to persist across situations.

Combined, these logics lead to the first implication:

1) **Private resolve must pass a threshold level of persistence for reputations to form; beyond that threshold, reputation formation becomes less likely as private resolve becomes more persistent. Thus, states are most likely to acquire reputations for resolve that influence adversaries' later behavior when resolve is moderately likely to persist.**

While the logic of the base model points to the non-monotonic relationship between context dependence and reputations that I have discussed, as I noted earlier, there are many kinds of context dependence, and the consequences of the various kinds differ.

My analyses assumed that only the unobserved portion of states' resolve differs between disputes. Another possibility is that one or both states expect the stakes to be much higher in their next dispute; this implies that both observable and unobservable components of resolve will be much greater. Reputation formation occurs when the challenger makes an offer just high enough to compensate an irresolute defender for the future value of a lost reputation, one that is still low enough to be rejected by a defender that is resolute in the current dispute. When an irresolute defender expects its next dispute to be much larger in magnitude, it may expect its reputation to be so valuable that it will reject any low offer that the resolute defender will refuse. and so a resolute defender cannot acquire a reputation. That is, when states expect a future dispute of unusually high stakes, we should rarely observe the formation of reputations.

When states do not expect a future, high-stakes dispute, however, reputations can form

[22] The challenger's offer in the equilibrium with reputation formation is $\theta + \delta\eta\rho$, which is higher when the persistence of resolve, $\rho$, is higher. As $\rho$ goes up, making this offer becomes less attractive to the challenger relative to the offer in the pooling equilibrium. The situation is more complicated than this because the value of the pooling offer to the challenger also changes, but on net, the pooling offer becomes relatively more attractive.

in the current dispute exactly as previously described. If a high-stakes dispute then unexpectedly occurs, the defender's reputation still will influence its credibility. If the defender stands firm in one dispute, the challenger learns that it was resolute in that dispute. As long as the persistence of private resolve is expected to be high enough between a low-stakes and a high-stakes dispute, this information is sufficiently relevant that the reputation has an effect on the unexpected, high-stakes dispute. If, however, the difference in stakes means that the defender's private resolve is likely to change (the actual persistence of resolve, $\rho'$, turns out to be higher than the persistence the states anticipated when the reputation formed), the challenger's knowledge that the defender was previously resolute will not lead it to be more accommodating in the high-stakes dispute.

These logics lead to a second implication about context dependence and reputations:

**2 a) When disputing states expect that the defender's next dispute will involve substantially greater stakes, the defender is unlikely to obtain a reputation for resolve. b) When a defender acquires a reputation for resolve and then its next dispute is of unexpectedly high stakes, the reputation will nevertheless have an effect on the high-stakes dispute.**

My broader investigation of context dependence leads to some caveats about the formation and effect of reputations. In a classic statement about the importance of reputations, Thomas Schelling wrote, "When we talk about the loss of face that would occur if we backed out of Formosa under duress, or out of Berlin, the loss of face that matters most is the loss of Soviet belief that we will do, elsewhere and subsequently, what we insist we will do here and now" (Schelling 1966, 55-6). Insofar as the U.S., at the time of its conflicts in Asia, anticipated Soviet threats to significant interests in Europe, my analyses suggest that Schelling was wrong; we should not have seen the U.S. getting a reputation in Asia that would affect its later interactions. This conclusion is consistent with Ted Hopf's finding that, "following the biggest postwar defeat–Vietnam...Soviets saw the United States as both capable and resolute" (Hopf 1994, 30).

### 4.2. Why Reputations are Likely to Affect Repeated Interactions within the Same Dyad

When the challenger expects repeated interactions with the same defender, it expects to benefit from learning the defender's level of resolve in the present dispute because its resolve today is an indicator, though imperfect, of its resolve tomorrow. Knowing the defender's resolve today allows it to tailor its behavior to the defender's level of resolve tomorrow – not perfectly, because the defender's level of resolve may change – but better than if it did not have this information. When the challenger expects the present dispute to be an isolated event, it does not expect to benefit from learning the defender's resolve.

Thus, challengers that believe they face a greater probability of repeated interactions with the same defender are more willing to risk low offers (probes) that may get rejected (leading to reputation formation) than challengers that believe themselves unlikely to interact again with this defender. Put differently, challengers are willing to offer low settlements when repeated interactions are more likely even when they are more pessimistic that their offers will be accepted, because rejection of the offer provides information that is more likely to have future value. That is, the model implies:

**3. Reputation formation is more likely when the countries believe it more likely that the same states will interact again in another dispute.**

Whether or not interactions involve the same countries over time and the extent to which resolve is changing thus have different effects on reputation formation in the model. When a challenger expects it is more likely to interact again with the same defender, it has greater incentive to make a probe and acquire information about the defender. How much the challenger has to pay to acquire the information (what bargain it must offer), however, depends on the probability that the defender's resolve will persist to a subsequent dispute; when the defender's resolve is more likely to change, the challenger must offer the defender a better bargain today to get the information it wants. These variables are likely to be correlated in practice: for example, in an enduring rivalry such as the Egypt/Israel case, the states expect to interact repeatedly and resolve is likely to be persistent. Overall, reputation formation is most likely when the same states interact over time *and* resolve is expected to be moderately persistent.

### 4.3. The Effect of Uncertainty

Information about adversaries' willingness to fight varies by the situation. For example, there may be greater uncertainty about the defender's willingness to fight when the states are closer in military power, or more certainty in particular disputes due to public demonstrations of power, the quality of the press, or the quality of intelligence.

The reputation formation that is the subject of this paper is possible only with some uncertainty; if the defender's resolve were perfectly known to the challenger, then there would be no possibility for its actions to influence the challenger's beliefs about its resolve. Thus, it is unsurprising that reputation formation is more likely when the challenger has greater uncertainty about the defender's private resolve. (The private component of resolve, $\eta$, is a larger component of its total resolve.) However, the "limit on context dependence" condition that I described earlier also implies that too much uncertainty can preclude reputation formation. Earlier, I wrote the condition as $\rho > \frac{\eta}{\Pi - \theta}$, but it can be written as $\eta < \rho(\Pi - \theta)$. Holding constant the defender's mean level of resolve, this condition is less likely to be satisfied when $\eta$ is larger. Extreme uncertainty precludes reputation formation because it widens the gap between the screening and pooling offers in the second dispute. For a reputation to be consequential, the challenger must make a pooling offer when the defender has a reputation; this offer is $\theta + \eta$ in the second dispute. When the extent of uncertainty $\eta$ is large enough, the challenger always prefers to take the risk of the lower, screening offer, and reputation formation is impossible.

The relationship between the amount of uncertainty and reputation formation is the mirror image of that between the persistence of resolve and reputations: as uncertainty increases, so does the likelihood of reputations and the interdependence of commitments, up to a point. When uncertainty is too extensive, the challenger is unwilling to risk the much higher pooling offer it must make in the second dispute in order for the defender to accept the offer if it is resolute. The implication is therefore:

**4. Uncertainty about the defender's resolve cannot be too great for reputa-**

**tions to form, but, before that threshold, reputation formation is more likely when the challenger has greater uncertainty about the defender's resolve. Thus, reputation formation and the interdependence of commitments are most likely at moderate levels of uncertainty.**

## 5. Windows of Vulnerability and Other Evidence

At the start of the paper, I mentioned three categories of empirical studies in which reputations have been found most consistently: studies of multiple disputes in the same region, with adversaries that are similar in power and/or interests, and/or with the same adversary over time. Since I was familiar with these findings, they cannot be used to test the theory. However, the model that I study does help to explain these empirical findings, both because they are implications rather than assumptions, and because the analyses show why we might see the patterns we do. The model assumes that resolve is partially persistent, but not that low enough persistence precludes the formation of reputations; it assumes that the challenger with which the defender interacts can change between disputes, but not that the defender is more likely to obtain a consequential reputation for resolve as it becomes more likely to interact with the same challenger. Varying the parameters could have led to different relationships than those observed empirically, but they did not.

Some case-study evidence about which I did not know when I did the study also is consistent with the findings. While U.S. leaders often have stated concern for the country's reputations, it is in some ways puzzling that "strong," or resolute states would fight for their reputations; after all, when a state is willing to defend particular interests (has low costs or high benefits from conflict in a particular situation), it has less need to rely on reputation.

My analyses suggest a motive behind strong states' fascination with reputations: One reason for leaders to stand firm is their concern that their country will not be as willing to fight in the future as it is now. Again as a simplification, resolve in the model is an absolute quantity, but in practice, leaders' decision to take their country to war may be based on their perception of their country's resolve relative to that of other countries involved in the situation. When leaders believe their country's resolve to be declining relative to others, they may choose to fight now to establish a reputation, knowing that this reputation will help them to obtain goals later, when they expect their own country to be less willing to obtain those goals through fighting.

If this is the case, we should expect to see states acting to obtain reputations for resolve when they perceive emerging windows of vulnerability or closing windows of opportunity, both situations in which states believe the balance of resolve is shifting in favor of their adversaries.[23] Studying several such cases involving the People's Republic of China (PRC) between 1949 and 1979, Christensen (2006) argues that sometimes the PRC's decision to fight in face of an opening window of vulnerability involved little intent to change the military situation by force; instead, the intent was to change the political situation, while at other times, the intent was to change both the political and the military situation.[24] The logic of the present article suggests a reputational component to the these decisions.

Christensen argues that, in the period covered by his study, the PRC used force most frequently when they perceived opening windows of vulnerability. In some of these cases, China sought to prevail militarily, but others were characterized by "the use of force primarily as a method of shaping longer-term security trends," (Christensen 2006, 52). For example, he argues that China's decision to use force in the Taiwan Strait crisis of 1954 was motivated by Mao's belief that "international forces were shifting against the PRC and would perhaps permanently prevent Beijing from attaining its goal of reunification with Taiwan and elimination of his Civil War enemy, Chiang Kai-Shek's Kuomintang (KMT)" (Christensen 2006, 58-9). Mao did not try to resolve the situation decisively through the use of military force; the PRC's shelling of the offshore islands was designed to "send a signal to both Taipei and the United States" that China would not accept a military alliance between the two (Christensen 2006, 59-60).

In sum, Mao appears to have used force in this case because he expected his country's vulnerability to increase, and to have done so to attain political goals by establishing his country as resolute. The present article explains why states would pursue this policy and

[23] I thank Steve Rosen for suggesting that the theory might help to explain state behavior in the face of emerging windows of vulnerability.

[24] I thank Steve Rosen for suggesting this example.

suggests that it often will be successful, though it was not successful in this case.

At least two of the hypotheses that I develop are testable using statistical methods. The model predicts that defenders are more likely to acquire reputations and then to obtain near-future success in deterrence, or else a favorable settlement, under particular circumstances. Since the likely persistence of private resolve has empirical referents, empirical work could test for the non-monotonic relationship implied by the theory.[25] It also could test more formally to determine if multiple interactions with the same adversary are more likely to lead to reputational effects; the existing literature suggests but generally does not attempt to show correlation.[26] The relationship between uncertainty and reputational effects is more difficult to investigate empirically, but a new literature shows promise in developing measures of uncertainty (Bas 2012; Kaplow and Gartzke 2015). Probably the most difficult hypotheses to test are about situations in which the stakes are changing substantially over time, since the relationship depends upon whether the change is expected or unexpected; case-study detective work would be needed to attempt to discern leaders' expectations at the start of the earlier crisis.

## 6. Conclusion

Scholars have debated whether or not states acquire reputations. The answer to the question, in turn, determines whether or not states' commitments are interdependent; if past

[25] A big-N test of this hypothesis would require an index representing the persistence of resolve. Creating such an index is challenging but possible. For example, private resolve may be more likely to persist when a country's regime or regime type is more stable (it is a stable democracy or is led by a stable autocrat), the latter if regime type affects countries' willingness to use force. It may be more likely to persist in stable autocracies than in stable democracies, which are subject to (perhaps-changing) pressures of public opinion. It also may be more likely to persist in the context of a rivalry and in other situations (e.g. decolonization, perhaps civil war situations) in which related issues are expected to arise repeatedly. Finally, a country's resolve may be less likely to persist in times of rapid changes of military technology.

[26] For example, the studies of same-dyad interactions over time generally do not compare the effect of behavior in past interactions with the same challenger to the effect of behavior in past interactions with other challengers; they consider same-dyad interactions only, and find reputation formation in those interactions.

actions influence present behavior, then failing to live up to one commitment can weaken the credibility of another. The empirical literature strongly suggests that "whether" is the wrong question: standing firm sometimes, but not always, leads to a reputation for resolve. However, while the empirical literature indicates situations in which reputations are most likely to form and affect future interactions, it does not sufficiently explain why. This article explains why reputation formation is most likely when a country is resolute and moderately likely to be resolute again in a future dispute, when the same countries interact repeatedly, and when the challenger has a moderate amount of information about the defender's resolve.

An important aspect of the debate concerns the effect of increasing stakes. While one day the issue might be Iraq and Syria, the next day it might be a passenger plane shot down over Ukraine. If the stakes are greater in a later situation, does this affect the formation and consequence of reputation? This article shows that states that stand firm with moderate probability of changing stakes are likely to obtain reputations that affect their adversaries' future decisions. However, if an anticipated change in stakes is dramatically large, the effect of standing firm depends on whether such changes are expected or unexpected; when states expect dramatically higher stakes in the future, reputation formation can be impossible.

The policy lesson from this research is therefore mixed. Commitments sometimes are interdependent, but are perhaps less likely to be so in two situations in which policymakers most want them to be: when their state is irresolute today and very likely to be irresolute again tomorrow, thus hoping for a reputation for resolve that will facilitate a bluff, and when policymakers expect future disputes of extreme importance.

When scholars discuss protracted conflicts, they sometimes see them as a puzzle for rationalist theories. If fighting reveals information, and states do not fight without uncertainty, then why do we see protracted or repeated wars? This article suggests a simple rationale for repeated or protracted international (or other) disputes between the same adversaries over what appear to be the same issues: learning always is imperfect in a world in which things change. As Morrow (1989, 227) notes, "repeated crises require that the parties cannot learn enough from the first crisis to deter the second." In the model I study, the challenger sometimes learns enough from the defender's past actions about its present resolve that the knowledge affects its behavior, but it never learns the defender's resolve with certainty, since resolve can differ from situation to situation. Plus ça change, moins c'est la même chose.

# References

Alt, J. E., R. L. Calvert, and B. D. Humes (1988, June). Reputation and hegemonic stability: A game-theoretic analysis. American Political Science Review 82(2), 445–466.

Bas, M. (2012). Measuring uncertainty in international relations; heteroskedastic strategic models. Conflict Management and Peace Sciences 29(5), 490–520.

Brecher, M. and J. Wilkenfeld (1997). A Study of Crisis. Ann Arbor: University of Michigan Press.

Christensen, T. J. (2006). Windows and war: Trend analysis and beijing's use of force. In A. I. Johnston and R. S. Ross (Eds.), New Directions in the Study of China's Foreign Policy, pp. 50–85. Stanford: Stanford University Press.

Clare, J. and V. Danilovic (2010). Multiple audiences and reputation building in international conflicts. Journal of Conflict Resolution 54(6), 860–82.

Crescenzi, M. J. C. (2007, April). Reputation and interstate conflict. American Journal of Political Science 51(2), 382–396.

Dafoe, A. and D. Caughey (2016). Honor and war: Southern U.S. presidents and the effects of concern for reputation. World Politics 68(2).

Dafoe, A., J. Renshon, and P. K. Huth (2014). Reputation and status as motives for war. Annual Review of Political Science (17), 371–393.

Fearon, J. D. (1995). Rationalist explanations for war. International Organization 49, 379–414.

Freedman, L. and V. Gamba-Stonehouse (1991). Signals of War; The Falklands Conflict of 1982. Princeton: Princeton University Press.

Harris, J. F. (1999, Friday, April 2). Stakes are growing for Clinton. The Washington Post, A1.

Hopf, T. (1994). Peripheral Visions: Deterrence Theory and American Foreign Policy in the Third World, 1965-1990. Ann Arbor: University of Michigan Press.

Howkins, A. (2010). A formal end to informal imperialism: Environmental nationalism, sovereignty disputes, and the decline of British interests in Argentina, 1933-1955. British Scholar 3(2), 235–262.

Huth, P. K. (1997, Autumn). Reputations and deterrence: A theoretical and empirical assessment. Security Studies 7(1), 72–99.

Jervis, R. (1970). The Logic of Images in International Relations. Princeton: Princeton University Press.

Kaplow, J. M. and E. Gartzke (2015). The determinants of uncertainty in international relations. Paper prepared for the Empirical Implications of Bargaining Theory Conference, May 14-15,2015, Princeton University.

Kennan, J. (1998). Informational rents in bargaining with serially correlated valuations. University of Wisconsin Madison, December 1, 1998 version.

Kennan, J. (2001). Repeated bargaining with persistent private information. Review of Economic Studies 68, 719–755.

Lieberman, E. (2013). Reconceptualizing Deterrence; Nudging Toward Rationality in Middle Eastern Rivalries. London and New York: Routlege Taylor  Francis Group.

Maxwell, S. (1968). Rationality in Deterrence. London: Institute for Strategic Studies.

Meirowitz, A. and A. E. Sartori (2008). Strategic uncertainty as a cause of war. Quarterly Journal of Political Science 3(4), 327–352.

Mercer, J. (1996). Reputation and International Politics. Ithaca: Cornell University Press.

Morrow, J. D. (1989). Capabilities, uncertainty, and resolve: A limited information model of crisis bargaining. American Journal of Political Science 33(4), 941–972.

Morrow, J. D. (1994). Game Theory for Political Scientists. Princeton: Princeton University Press.

Nalebuff, B. (1991). Rational deterrence in an imperfect world. World Politics 43(3), 313–335.

Press, D. (2005). Calculating Credibility; How Leaders Assess Military Threats. Ithaca: Cornell University Press.

Schelling, T. C. (1966). Arms and Influence. New Haven: Yale University Press.

Sechser, T. S. (2010). Goliath's curse; coercive threats and asymmetric power. International Organization 64(4), 627–60.

Snyder, G. H. and P. Diesing (1977). Conflict Among Nations. Princeton: Princeton University Press.

Tingley, D. H. and B. F. Walter (2011). The effect of repeated play on reputation building: An experimental approach. International Organization 65, 343–65.

Walter, B. (2006). Building reputation; why governments fight some separatists but not others. American Journal of Political Science 50(2), 313–30.

Weisiger, A. and K. Yarhi-Milo (2015). Revisiting reputation: How past actions matter in international politics. International Organization 69(2), 473–95.

# 7. Appendix

This appendix contains the full model and an overview of the proofs. The model that I use was developed by Kennan (1998) to explain labor economics. The work in this appendix aims to make the ideas behind the proofs more accessible. The appendix extends the work in Kennan (1998) as follows:

1. To understand the model's implications for international relations, I reparameterize the model and perform comparative statics, including representing the degree of uncertainty by $\eta$ so that it can be varied for comparative statics.[27]
2. I perform welfare analyses of the second dispute.
3. I allow the challenger (the uninformed actor) to differ between the two disputes, showing that the defender can obtain a reputation for resolve even when the states expect the second dispute to involve a new challenger; however, it is less likely to do so in this case.
4. I evaluate the model assuming that the second dispute will be of greater magnitude, either in expected ways or in unexpected ones.

I refer the reader to Kennan (1998) for full proofs. However, at the end of this appendix, I re-write the tables in which he summarizes his equilibrium analyses using the notation I use in this paper, in order to make using his work easier for the reader. A re-working of Kennan's proofs with the reparameterization used in this paper is available from the author. Any errors are, of course, my own.

## 7.1. The Model

To make the work easier to follow, I present the full model here, repeating the description that is in the main article.

[27] Parameterizing the degree of uncertainty does not change the model mathematically, but facilitates the interpretation.

### 7.1.1. Actors, Sequence of Moves, and Payoffs

Each time period (dispute) $t \in \{1,2\}$, begins with a move by Nature which determines the defender's value for war. Substantively, this corresponds to a dispute arising in which the defender has some value for war. Nature draws a number $n_t \in \{0,1\}$. Nature chooses $n_1 = 1$ with probability $h_1$ and $n_1 = 0$ with probability $1-$ $h_1$. The probability that $n_2 = n_1$ is $\rho$ with $0 < \rho < 1$.Thus, Nature chooses $n_2 = 1$ with probability $1 - \rho + \varphi n_1$ and $n_2 = 0$ with probability $\rho - \varphi n_1$ where $\varphi \equiv 2\rho - 1$.

The defender's value for war in period $t$ is $n_t\eta + \theta$, so it is either low ($\theta$) or high ($\eta + \theta$), with $\eta > 0$. The parameter $\rho$ thus determines how likely a defender's value for war is to be the same in the second period – that is, the degree of persistence of resolve. If a defender is resolute/irresolute in the first dispute ( $n_1 = 1$), the probability that it is resolute/irresolute again in the second is $\rho$.

After Nature moves, the states bargain using a simple ultimatum-game bargaining protocol: First, the challenger makes an offer $a_t$. Then, the defender chooses $o_t \in \{0,1\}$, where $o_t = 1$ means that it accepts the offer, and $o_t = 0$ means that it rejects the offer and the states go to war. If the defender accepts the offer, its per-period payoff is $a_t$ and the challenger's is $\Pi - a_t$. If the defender rejects the offer, the states go to war. The challenger's payoff from war is normalized to zero.

After the first dispute ($t = 1$) ends with an accepted bargain or war, the defender engages in another dispute. With probability $\lambda$ ($0 < \lambda \leq 1$), the challenger in the first dispute is the same in the second; with probability $(1 - \lambda)$ the defender interacts with a new challenger in the second period.[28] The parameter $\lambda$ thus represents how likely the states in the first dispute believe it to be that they will interact again. If the same two states participate in both disputes, both states' payoffs are the discounted sums of the two stage-game payoffs, with common discount factor $0 < \delta < 1$.[29] If the first-period challenger does not interact

[28] I assume $\lambda > 0$ because the challenger will always expect some positive probability, however small, of interacting with the defender again in real-world applications. In the case of $\lambda = 0$, the challenger only is affected by the current dispute, and does not care how the current dispute impacts the future dispute.

[29] Substantively, the discount factor can represent both discounting of payoffs that the states get in the future and the probability that a future dispute will arise.

in the second dispute, its second-period payoff is normalized to zero; thus, the first-period challenger's expected continuation value is $\lambda$ times the expected payoff of a second dispute in which it participates.[30] If a new challenger is involved in the second dispute, its total payoff is its payoff from that dispute.

### 7.1.2. Information and Beliefs

Nature's move at the start of each time period is observed by the defender but not the challenger. Thus, the defender has private information about its value for war in each time period. All other aspects of the game are common knowledge, including the transition probability of resolve ($\rho$). For example, a defender that is resolute in the current context knows that it will be resolute in its next dispute with probability $\rho$; if a challenger learns that a defender is resolute today it knows that the defender will be resolute in the next dispute with probability $\rho$. I let $\zeta_t$ be the challenger's belief at the beginning of time $t$ about the probability that $n_t = 1$, where $\zeta_1 = h_1$ since the challenger knows $h_1$. If a new challenger is involved in the second dispute, the challenger in the second dispute observes behavior in the first dispute before engaging in the second.

### 7.1.3. Additional Assumptions

1. A defender's value for war is positively correlated over time periods; a state that is resolute today is more likely to be resolute tomorrow than one that is irresolute today, and a state that is irresolute today is more likely to be irresolute tomorrow than one that is resolute today. Formally, I assume that $\rho > \frac{1}{2}$.

2. War is costly, so that there is some distribution of the goods that both countries prefer to war. Formally, I assume that $\Pi - (\eta + \theta) > 0$. This assumption implies that, absent reputational considerations, there exists some bargain that both players prefer to war in period $t$, even if the defender is resolute.

[30] The challenger does not choose whether or not to be involved in a future dispute in this model, so normalizing its second-period payoff from not being so involved to zero is inconsequential.

### 7.1.4. Summary of Notation

| Notation | Meaning |
|---|---|
| $t$ | time period (dispute 1 or 2) |
| $\Pi$ | value of the disputed good |
| $\theta$ | irresolute defender's value for war |
| $\theta + \eta$ | resolute defender's value for war |
| $\rho$ | probability that defender's type in $t = 2$<br>is same as in $t = 1$ |
| $\varphi$ | $2\rho - 1$ |
| $\delta$ | discount factor |
| $\zeta_t$ | challenger's belief at start of $t$ that defender is resolute |
| $\lambda$ | probability that the same challenger interacts in $t = 1, 2$ |
| $a_t$ | challenger's offer in time $t$ |

Table 7.1: Notation

## 7.2. The Second Dispute

This section characterizes the equilibria of the second dispute, the second period of the game. (See Kennan, 1998, Section 3, pages 6-7.) This second dispute is equivalent to a one-period game in which the challenger's belief that the defender is resolute at the start of the game is $\zeta_2$.

As discussed earlier in the paper, equilibrium behavior in the second dispute is characteristic of screening games: if the challenger is optimistic enough, it makes a low offer, which only irresolute defenders accept; if it is more pessimistic, it makes a higher offer, which both irresolute and resolute defenders accept. Depending on whether $n_1 = 0$ (low type) or if $n_1 = 1$ (high type), $\sigma^{L \text{ or } H}(a_t)$ is the probability with which the offer is accepted in period $t$ ($o_t(a_t) = 1$).

**Proposition 1.** *The following describes second-period strategies in equilibrium.*

1. **Pooling**. If $\zeta_2 > \zeta_2^*$, then $\sigma_2$ is an equilibrium strategy profile iff

a) $\sigma_2^c = \{\eta + \theta\}$

b) $\sigma_2^L = 1$ if $a_2 > \theta; \sigma_2^L = 0$ if $a_2 < \theta; \sigma_2^L(\theta) = \alpha$ *where* $\alpha \in [0, 1]$;

c) $\sigma_2^H = 1$ if $a_2 \geq \eta + \theta; \sigma_2^H = 0$ if $a_2 < \eta + \theta;$ .

**2. Screening**. If $\zeta_2 < \zeta_2^*$, then $\sigma_2$ is an equilibrium strategy profile iff

a) $\sigma_2^c = \{\theta\}$

b) $\sigma_2^L = 1$ if $a_2 \geq \theta; \sigma_2^L = 0$ if $a < \theta$;

c) $\sigma_2^H = 1$ if $a_2 > \eta + \theta; \sigma_2^H = 0$ if $a_2 < \eta + \theta; \sigma_2^H(\eta + \theta) = \alpha$ *where* $\alpha \in [0, 1]$

**3. Randomizing.** If $\zeta_2 = \zeta_2^*$, then $\sigma_2$ is an equilibrium strategy profile iff

a) $\sigma_2^c(\theta) = \gamma, \sigma_2^c(\eta + \theta) = (1 - \gamma), where\ \gamma \in [0, 1]$

b) $\sigma_2^L = 1$ if $a_2 > \theta; \sigma_2^L = 0$ if $a < \theta; \sigma_2^L(\theta) = \alpha$ *where* $\alpha \in [0, 1], \gamma(1 - \alpha) = 0$;

c) $\sigma_2^H = 1$ if $a_2 > \eta+\theta; \sigma_2^H = 0$ if $a_2 < \eta+\theta; \sigma_2^H(\eta+\theta) = \beta$ *where* $\beta \in [0, 1], (1-\gamma)(1-\beta) = 0$;

Sketch of proof. To maximize its utility, any defender must reject $a < \theta$ and accept $a > \eta + \theta$.It must accept $a \in (\theta, \eta + \theta)$ when it is irresolute and reject it when it is resolute. Thus, the only offers that can be optimal for the challenger are $a \in \{\theta, \eta+\theta\}$. Randomization is an optimal strategy for the defender when the offer equals its value for war. However, in any pooling equilibrium in which the challenger offers $a = \eta + \theta$, the resolute defender must accept $a = \eta + \theta$ with certainty, or the challenger would do better by offering $\epsilon$ more. Similarly, in any equilibrium in which the challenger offers $a = \theta$, the irresolute defender must accept $a = \theta$ with certainty.

If $\zeta_2 \equiv \zeta_2^*$ then randomization over $\{\theta, \eta + \theta\}$ is optimal for the challenger, and unless the challenger's strategy chooses with certainty not to offer the defender's value for war, the defender's equilibrium strategy must accept with certainty when the offer equals its value for war.■

While the subgame that is the second dispute has many equilibria, each equilibrium has the same path of play and expected payoffs for a given value of $\zeta_2$, the challenger's belief at the beginning of the second period, except in the knife-edged case of randomization. More formally:

**Corollary 1.** *Second-period Equilibrium Payoffs. Challenger: In case 1 ($\zeta_2 > \zeta_2^*$), the* *challenger offers* $\eta + \theta$ *which is always accepted, so its payoff is* $\Pi - (\eta + \theta)$*. In case 2* $(\zeta_2 < \zeta_2^*)$*, the challenger offers* $\theta$ *which it believes will be accepted with probability* $(1- \zeta_2)$ *so its expected payoff is* $(1- \zeta_2)\ (\Pi - \theta)$*. Defender: In case 1, the defender's payoff is* $\eta + \theta$*. In case 2, the defender's payoff is* $\theta$ *if it is irresolute in period 2 and* $\eta + \theta$ *if it is resolute in that period.*

This corollary follows directly from the proposition.

### 7.2.1. Reputation Condition

If a reputation for resolve ever is to affect the challenger's behavior, it must do so when the challenger has learned all relevant information about the defender's first-dispute type from its behavior in that dispute. I study the case in which the challenger will make a low, screening offer in the second dispute if it learns nothing about the defender (that is, without any learning, $\zeta_2 < \zeta_2^*$, so I call this challenger "optimistic"); for consequential reputations for form in this situation, the challenger must make a high (pooling) second-period offer if and only if first-period actions reveal the defender definitely to have been resolute in the first dispute.

Remember that $\zeta_1$ is the challenger's belief at the start of the first dispute about the probability that the defender is resolute and $\rho$ is the probability with which a defender's level of resolve persists to the second dispute. If the challenger learns nothing in the first dispute, its belief at the start of the second is that the probability that the defender is resolute is thus $(2\rho - 1)\zeta_1 + 1 - \rho$. If it learns that the defender definitely is resolute in the first dispute, its belief that the defender is resolute at the start of the second is $\rho$, and if it learns that the defender definitely is irresolute in the first dispute, its belief that the defender is resolute at the start of the second is $1 - \rho$.

For the challenger to make a screening offer in the second dispute when its belief is $(2\rho - 1)\zeta_1 + 1 - \rho$ and a pooling offer when its belief is $\rho$, the first belief must be below the threshold belief between screening and pooling and the second above it. Thus, the following condition must be satisfied in order for the defender to obtain a consequential reputation when the challenger begins optimistic:[31]

*The Reputation Condition.* $(2\rho - 1)\zeta_1 + 1 - \rho < \zeta_2^* < \rho$.

Together with the equation characterizing the threshold $\zeta_2^*$, this condition implies that consequential reputations can form only when the resolute defender's type is persistent enough; $\rho > \frac{\eta}{\Pi - \theta}$ is necessary for the challenger to act differently in the second dispute if it learned in the first that the defender was resolute.

### 7.3. Equilibrium Concept and Refinement

Because the challenger has uncertainty about the defender's type, the appropriate equilibrium concept is analogous to Perfect Bayesian Equilibrium. Bayes' Rule does not restrict beliefs after zero-probability events, so in analyzing the game one must decide whether or not there are restrictions on beliefs following such events. There are two kinds of zero-probability events in the model – actions by the challenger that it never takes in equilibrium, and actions by the defender that it never takes in equilibrium. I treat these differently because there is no uncertainty in the model about the challenger's payoffs, so the challenger's choices should not affect anyone's beliefs directly.

There are offers that the challenger never chooses to make in equilibrium (for example, offers so high that all defenders would accept a lower offer). In analyzing the model, one must consider what would happen if the challenger were to deviate and make such an offer. After a deviation by the challenger, I assume that the defender does not update its beliefs about the challenger's type because the challenger has only one type. The challenger continues to hold its prior belief about the defender until the defender accepts or rejects the off-path offer, at which point the challenger updates its beliefs using Bayes' rule whenever possible.

There also are actions that the defender never chooses to take in equilibrium. For example, in a pooling equilibrium, the defender always accepts the offer that the challenger makes, so that rejection of this offer is a zero-probability event. Since the challenger is un-

[31] This condition is analogous to Kennan's Condition O.

certain about the defender's value for war, deviations by the defender might reasonably lead to changes in the challenger's beliefs about the defender's type. Following a deviation by the defender, I allow for any beliefs, solve for equilibria/actions, show most results, and then, following Kennan (1998), use introspective consistency – similar to the intuitive criterion, a standard equilibrium refinement – to simplify.

Specifically, there is a continuum of pooling equilibria that, following Kennan, I rule out because they rely on beliefs that do not satisfy introspective consistency. (I do not use these in my analyses or summarize the proofs in this appendix, but proofs are in Kennan (1998).) These involve offers $a_1 \in [\eta + \theta, \eta + \theta + \delta\eta(1 - \rho))]$. The idea of introspective consistency, which is an extension of the Intuitive Criterion, is that if rejection of an offer in this range is strictly worse than acceptance for the irresolute type, given that the challenger has some belief about the defender's type and responds optimally, and if this is not true for the resolute type, then the challenger should put zero probability on the irresolute type if the offer is rejected. (See Kennan, page 21). The only pooling offer that satisfies this condition is $a_1 = \eta + \theta + \delta\eta(1 - \rho)$.

### 7.4. Logic of the Proofs of Reputational and Non-Reputational Equilibria; Cut-Point Belief

When the challenger begins the first dispute optimistic enough that it will make a low offer in the second dispute if it learns nothing about the defender's resolve in the first dispute, the model has two equilibria involving pure strategies in the first dispute that satisfy introspective consistency. The first is a reputational equilibrium: The challenger makes a low offer in the first dispute, and the defender rejects the challenger's offer and fights in the first dispute if and only if it is resolute, obtaining a reputation for resolve; if and only if the defender has such a reputation, the challenger makes a high offer in the second dispute. The second is an equilibrium without reputation formation: the challenger makes a higher offer in the first dispute, which all defenders accept; the challenger makes a low offer in the second dispute, which the defender accepts if and only if it is irresolute. (The model also a mixed strategy equilibrium in which the defender acquires a reputation in the first dispute that has no effect on the challenger's later behavior; I do not use this equilibrium in my analyses.)

The key insight of the equilibria is that, if rejecting an offer leads to a reputation that gets the defender a better deal in the second period, the challenger must compensate the defender in the first period any time that it *accepts* an offer, in order to make up for the lost value of a reputation. In the reputational equilibrium, the defender rejects the offer and acquires a reputation for resolve if and only if it is resolute; thus, the challenger must compensate the irresolute defender for failing to acquire a reputation. In the equilibrium without reputation formation, the challenger makes a high offer in the first period that all defenders are willing to accept; this offer must be even higher than the resolute defender's value for war in order to compensate it for its failure to acquire a reputation.

As discussed in the body of the article, a reputation for resolve is valuable only to a defender that turns out to be irresolute in the second dispute, and its value is $\eta$ when it does turn out to be irresolute. Thus, the amount that a defender must be compensated for failing to acquire a reputation in the first dispute is the discounted value of $\eta$ multiplied by the probability with which it expects to be irresolute in the second dispute: $\delta\eta\rho$ for a defender that is irresolute in the first dispute, and $\delta\eta(1-\rho)$ for a defender that is resolute in the first dispute.

**Proposition 2.** ***Equilibrium with Reputation Formation (First-period Screening):*** *In the first period, the challenger offers $a_1$. An irresolute defender accepts an offer $a_1$ iff $a_1 \geq \theta+\delta\rho\eta$. A resolute defender accepts an offer $a_1$ iff $a_1 \geq \theta+\eta+\delta(1-\rho)\eta$, so it rejects the offer that the challenger makes on the equilibrium path. In the second period, the challenger believes that the defender is resolute with probability $(1-\rho)$ if it accepted the first-period offer and resolute with probability $\rho$ if it rejected the first period offer. It offers $\theta$ if the defender accepted the first-period offer, and $\theta+\eta$ if the defender rejected the first-period offer. In the second period, the defender's acceptance strategies are as in Proposition 1.*

**Sketch of Proof**:

The challenger offers $a_1$ in the first period (the model's notation).

1. A defender that is irresolute in the first period will be irresolute again in the second with probability $\rho$ by the model's set-up. In the posited equilibrium, in which complete separating occurs in the first period, the challenger must believe that the probability that the defender is resolute is $(1 - \rho)$ if the defender accepted the first-period offer and resolute with probability $\rho$ if it rejected the first period offer by Bayesian updating. When the Reputation Condition holds, the challenger will make a low offer of $\theta$ in the second period if the defender accepts the first-period offer by Proposition 1. Similarly, it will make a high offer of $\theta + \eta$ in the second period if the defender rejects the first-period offer.

2. The irresolute defender is willing to accept $a_1$ iff $a_1 \geq \theta + \delta\rho\eta$ because it expects to be irresolute again with probability $\rho$, in which case it expects to accept a second-period offer of $\theta$ if it accepts the first-period offer, while if it rejects the first-period offer it will be offered and accept $\theta + \eta$; this difference in expected payoff is multiplied by the discount factor $\delta$. (If it turns out to be resolute in the second dispute, its payoff will be the same whether or not it accepted the first-period offer.) (Set-up of the model and Proposition 1)

3. The resolute defender is willing to accept $a_1$ iff $a_1 \geq \theta + \eta + \delta(1 - \rho)\eta$ because it expects to be irresolute in the second dispute with probability $(1 - \rho)$, in which case the logic of step 2 applies. (Set-up of the model and Proposition 1)

4. Unless the challenger makes an offer that the resolute defender will accept with positive probability, which does not occur here because we are looking for a screening equilibrium, its optimal offer is the minimum that the irresolute defender is willing to accept (since it always prefers an accepted offer to war by Assumption 2). Thus it offers $a_1 = \theta + \delta\rho\eta$.

5. A defender that is resolute in the first period strictly prefers to reject the offer $a_1 = \theta + \delta\rho\eta$ because $\theta + \delta\rho\eta < \theta + \eta + \delta(1 - \rho)\eta$. The reason is that $\delta\rho\eta < \eta + \delta(1 - \rho)\eta$ can be written as $\delta(2\rho - 1) < 1$. This inequality holds because $\rho$, $\delta < 1$ by assumption of the model.

6. If the challenger were to deviate to a higher offer, this deviation could only be profitable

if the resolute type of defender were to put positive probability on accepting. This condition would require $a_1 \geq \theta + \eta + \delta(1-\rho)\eta$, which would be accepted by all defenders, so that the challenger would learn nothing from acceptance and would make the low offer in the second dispute. Thus, the challenger's expected payoff from this deviation would be its payoff from the equilibrium without reputation formation, described next. When the challenger's prior beliefs about the defender's type ($\zeta_1$) are such that its expected payoff from playing its equilibrium strategy is at least as high as its expected payoff in the equilibrium without reputation formation, then it does not have an incentive to deviate from its reputational equilibrium strategy. Later, in the section on the "cut-point belief," I show that the challenger prefers to make the screening offer of $a_1 = \theta + \delta\rho\eta$ in the first dispute as long as $\zeta_1 \leq \frac{\eta(1-\delta\varphi)}{(1-\delta\rho)(\Pi-\theta)+\delta\eta(1-\rho)} \equiv \zeta_1^T$.

**Proposition 3.** ***Equilibrium without Reputation Formation (First-period Pooling):*** *In the first period, the challenger offers $a_1 = \theta + \eta + \delta(1-\rho)\eta$. An irresolute defender accepts an offer $a_1$ iff $a_1 \geq \theta + \delta\rho\eta$. A resolute defender accepts an offer $a_1$ iff $a_1 \geq \theta + \eta + \delta(1-\rho)\eta$, so both irresolute and resolute defenders accept the offer that the challenger makes on the equilibrium path. In the second period, if the defender accepted the first-period offer, the challenger believes that the defender is resolute with probability $\zeta_1\rho + (1-\zeta_1)(1-\rho)$ and offers $\theta$. If the defender rejected the first-period offer (off the equilibrium path), the challenger believes that the defender is resolute with probability $\rho$ and offers $\theta + \eta$ in the second dispute. The defender's second-period acceptance strategy is as in Proposition 1.*

**Sketch of Proof**:

1. In the posited equilibrium all defenders must take the same action in the first period on the equilibrium path so that the challenger does not learn from the action.[32] Since the challenger prefers any accepted offer to war by Assumption 2, if it will make an

[32] The game has another equilibrium with partial pooling in the first period, in which the challenger learns about the defender's type from its first-period action but not enough to change the challenger's behavior in the second period. This is an equilibrium with reputations that have no consequence, but also one with unreasonable off-the-equilibrium path beliefs.

offer that the resolute defender will accept with certainty, the challenger's payoff is maximized by making the minimum offer that the resolute defender is willing to accept, $a_1 = \theta + \eta + \delta(1-\rho)\eta$.

2. If the irresolute defender accepts the first-period offer, it expects $\theta + \eta + \delta(1-\rho)\eta + \delta[\rho(\theta) + (1-\rho)(\theta+\eta)]$ as its full game payoff because it expects to be irresolute again with probability $\rho$ and accept the challenger's second-period equilibrium offer of $\theta$, and if it is resolute, by Proposition 1 it will reject the offer and receive the resolute type's value for war, $\theta + \eta$. If it rejects the first-period offer in favor of war, it expects $\theta + \delta(\theta+\eta)$ between the first-dispute payoff and the discounted payoff from the second dispute, because the challenger's belief after a rejected offer, the Reputation Condition, and Proposition 1 imply that the challenger will make the higher offer in the second dispute. The irresolute defender strictly prefers to accept the first-period offer because $\theta + \eta + \delta(1-\rho)\eta + \delta[\rho(\theta) + (1-\rho)(\theta+\eta)] > \theta + \delta(\theta+\eta)$.(This condition simplifies to $\eta > 0$, which is true by assumption in the model.)

3. If the resolute defender accepts the first-period offer, it expects $\theta + \eta + \delta(1-\rho)\eta + \delta[(1-\rho)(\theta) + \rho(\theta+\eta)]$ because it expects to become irresolute with probability $1-\rho$, and the logic is the same as with the irresolute defender. If it rejects the first-period offer in favor of war, it expects $(1+\delta)(\theta+\eta)$. With some algebra, one can see that it is exactly indifferent between accepting and rejecting the offer. Thus, it can be an equilibrium strategy for it to accept.

4. If the challenger were to deviate from its first-period offer, its optimal deviation would be the minimum that the irresolute defender was willing to accept, $a_1 = \theta + \delta\rho\eta$, the first-period offer in the reputational equilibrium (Assumption 2). When its initial beliefs $\zeta_1$ are such that the expected payoff from its pooling equilibrium strategy is at least as high as its expected payoff from its reputational equilibrium, it has no incentive to deviate from its pooling equilibrium strategy.

5. By Bayes' Rule, any beliefs by the challenger following rejection of the first-period offer are possible, since rejection is a zero-probability event in equilibrium.■

**Definition 1.** *I define the threshold belief* $\zeta_1^T$ *as the belief at the start of the first dispute* *below which the challenger strictly prefers the reputational offer and above which (when the challenger is more pessimistic that the defender is resolute) the challenger strictly prefers the pooling offer, and at which the challenger is indifferent between the two offers.*

### 7.4.1. Cut-Point Belief

I now find the threshold between screening and pooling, $\zeta_1^T$.

We can compare the present-day payoffs, taking into account that the screening payoff continuation value is higher by $\zeta_1 \delta[\rho(\Pi - \theta) - \eta]$. Then, the challenger screens when

$(1 - \zeta_1)(\Pi - \theta - \delta\eta\rho) + \zeta_1\delta[\rho(\Pi - \theta) - \eta] \geq \Pi - (\eta + \theta) \ -\delta\eta(1 - \rho)$.

Rearranging terms and noting that $(1 - \delta\rho)(\Pi - \theta) + \delta\eta(1 - \rho) > 0$ because $\delta, \rho < 1$ and $\delta, \eta > 0$ by assumption in the set-up of the model, and $\Pi - \theta > 0$ by assumption 2, the challenger screens when

$$\zeta_1 \leq \frac{\eta(1 - \delta\varphi)}{(1 - \delta\rho)(\Pi - \theta) + \delta\eta(1 - \rho)} \equiv \zeta_1^T$$

and pools otherwise. (It also may pool when indifferent.) This threshold ($\zeta_1^T$) always is strictly positive because $\eta > 0$ and $\delta, \rho < 1$, so $\varphi \equiv 2\rho - 1 < 1$ and $\delta\varphi < 1$; $\Pi - \theta > 0$ by Assumption 2. Because the threshold is positive, the reputational equilibrium exists for some parameter values.

### 7.4.2. Comparative Statics

Here, I investigate how the thresholds vary with regard to two parameters of the model: the persistence of resolve ($\rho$), and the amount of information about the defender that is known by the challenger ($\eta$). I later consider the effect of changing the probability that the same challenger will be involved in the second dispute.

- I first investigate how the threshold between screening and pooling ($\zeta_1^T = \frac{\eta(1-\delta\varphi)}{(1-\delta\rho)(\Pi-\theta)+\delta\eta(1-\rho)}$) changes with the persistence of resolve ($\rho$).

$$\frac{\partial \zeta_1^T}{\partial \rho} = \frac{-2\delta\eta\left[(1 - \delta\rho)(\Pi - \theta) + \delta\eta(1 - \rho)\right] - \eta\left[1 - \delta(2\rho - 1)\right]\left[-\delta(\Pi - \theta + \eta)\right]}{\left[(1 - \delta\rho)(\Pi - \theta) + \delta\eta(1 - \rho)\right]^2}$$

The sign of the partial derivative is the sign of the numerator. Since all terms are multiplied by $\delta n > 0$, I divide the numerator by $\delta n$ and determine the sign of the remaining numerator, which simplifies to

$$(\delta - 1)(\Pi - \theta - \eta).$$

This quantity is negative because $\Pi - \theta > \eta$ by Assumption 2 and $\delta < 1$, so *the threshold goes down/screening becomes less attractive versus pooling at the offer as the persistence of resolve ($\rho$) increases.* In substantive terms, this means that reputation formation becomes less likely as the persistence of resolve increases beyond the threshold set by the Reputation Condition ($\rho = \frac{\eta}{\Pi - \theta}$).

- I next investigate how the threshold between screening and pooling changes with the degree of the challenger's uncertainty about the defender's resolve. Note that this question cannot be answered by simply varying either $\theta$ or $\eta$, because doing so varies the defender's mean value for war. Instead, I hold constant the mean value for war of an average defender (assuming the two types are equally likely, this is $\frac{2\theta+\eta}{2}$) and vary the amount of the payoff about which the challenger is uncertain, $\eta$. I do this by performing a change of variables. Let $\phi \equiv \frac{2\theta+\eta}{2}$ (so that $\theta = \phi - \frac{\eta}{2}$).

$$\begin{aligned} \zeta_1^T &\equiv \frac{\eta(1-\delta\varphi)}{(1-\delta\rho)(\Pi-\theta)+\delta\eta(1-\rho)} \\ &= \frac{\eta(1-\delta\varphi)}{(1-\delta\rho)(\Pi-\phi+\frac{\eta}{2})+\delta\eta(1-\rho)} \end{aligned}$$

$$\frac{\partial \zeta_1^T}{\partial \eta} = \frac{\left[(1-\delta\rho)(\Pi-\phi+\frac{\eta}{2})+\delta\eta(1-\rho)\right](1-\delta\varphi) - \eta(1-\delta\varphi)\left[\frac{1}{2}(1-\delta\rho)+\delta(1-\rho))\right]}{\left[(1-\delta\rho)(\Pi-\phi+\frac{\eta}{2})+\delta\eta(1-\rho)\right]^2}$$

Since the denominator is squared, the sign of the partial derivative is the sign of the numerator. Substituting back in for $\phi$ and simplifying, the numerator is

$$(1-\delta\varphi)\left[(1-\delta\rho)(\Pi-\theta-\frac{\eta}{2})\right].$$

Since $1 - \delta\varphi > 0$ because by $\varphi \equiv 2\rho - 1 < 1$, the sign of the partial is the sign of the term in square brackets. The term $\Pi - \theta - \frac{1}{2}\eta > 0$ by Assumption 2, and $(1 - \delta\rho) > 0$ by

$\delta, \rho < 1$. Thus, *the threshold between screening and pooling rises as the amount of uncertainty the challenger has about the defender's payoff rises*, holding constant the defender's mean payoff from war. Assuming continuity in the distribution of initial beliefs, this means that reputation formation is more likely in situations of greater uncertainty.

### 7.5. Welfare Analysis

This welfare analysis performs the following thought experiments under the condition that the challenger will make a low offer in the second dispute if it learns nothing about the defender's type from its first-dispute behavior. Consider the challenger's payoffs in the second dispute in the equilibrium with reputation formation. Is the challenger better off when the defender has a reputation for resolve (or for lacking resolve) than it would be if the defender had no reputation but the same probability of being the type that has a reputation for resolve (or for lacking resolve)? How well off is each type of defender, compared to its expected well-being if it were the same type but without a reputation?

The intuition behind the welfare analysis for the challenger is simple: If the defender reveals information about its type through its first-dispute behavior (separates at least partly in the first dispute), it is always (weakly) worthwhile to the challenger to take this information into account in the second dispute. Given this information, the challenger has an optimal second-dispute action, which is the action it takes in the equilibrium with reputations. If the challenger would have taken a different action without this information, the information makes the challenger better off. If the challenger coincidentally would have taken the same action, the challenger's payoff is the same with and without reputations. As long as the reputational equilibrium exists, it must be the case that the challenger is at least as well off playing its equilibrium action in the second period as it would be if it failed to take into account the information carried in the defender's reputation, but in fact the defender is the same defender, so the distribution of the defender's type is identical to what it would have been in the reputational equilibrium. This is similarly true if the defender has a reputation for lacking resolve, which means that its true probability of being resolute is now $(1 - \rho)$.

The defender also is (weakly) better off with the exchange of information that occurs with reputations. If the defender revealed through its behavior that it was resolute in the first dispute and turns out to be irresolute in the second dispute, it is better off with a reputation for resolve in the second dispute because it is able to bluff. If it revealed that it was irresolute and turns out to be irresolute again in the second dispute, its welfare will be the same in the second dispute, because I examine a situation in which the challenger will make a low offer if it learns nothing from the first dispute. If the defender turns out to be resolute in the second dispute, its welfare is unaffected by whether or not it revealed information in the first dispute; a resolute defender's payoff in the second dispute (where there are no considerations of future in the model) is always equal to its value for war, which it either obtains by accepting a bargain or by rejecting a lower offer in favor of fighting.

**Proposition 4.** *The set of outcomes in the second dispute if states play their equilibrium strategies in the reputational equilibrium is Pareto superior to the set of outcomes if states play the equilibrium of the one-period game without reputations: A) The challenger's payoff in the second dispute always is at least as high and sometimes is higher in the equilibrium with consequential reputations than it would be without reputations in the same situation. B) The defender's payoff in the second dispute always is at least as high and sometimes is higher in the equilibrium with consequential reputations than it would be without reputations in the same situation.*

Background: By Proposition 1, without reputations, the challenger's payoff in the second dispute depends on its beliefs at the start of that dispute, which determine whether it makes a high offer $(\theta + \eta)$ or a low offer $(\theta)$. I examine a situation in which the challenger will make a low offer in the second dispute if it learns nothing about the defender's resolve from its behavior in the first dispute, but as in the reputational equilibrium, the challenger will make a high offer if it learns that the defender was definitely resolute in the first dispute (the Reputation Condition is satisfied).

**Sketch of Proof.**

**Case A. The defender has a reputation for resolve.**

1. In the equilibrium with consequential reputations, the challenger makes a high offer

when the defender has this reputation and the defender always accepts; the challenger's payoff is $(\Pi - \theta - \eta)$ and the defender's is $\theta + \eta$.

2. The defender acquired this reputation only if it was resolute in the first dispute, so the probability that it is resolute now is $\rho$ (by the set-up of the model). By Proposition 1, the challenger makes the high offer if $\zeta_2 > \frac{\eta}{\Pi - \theta}$ and is indifferent between making it and not if $\zeta_2 = \frac{\eta}{\Pi - \theta}$; since it is making the high offer, we know $\rho \geq \frac{\eta}{\Pi - \theta}$.
3. Without a consequential reputation, the challenger's payoff would be $(\Pi - \theta)(1 - \rho)$ ; the defender's would be $\theta$ if it is irresolute in the current dispute and $\theta + \eta$ if it is resolute. (Proposition 1)
4. Difference in payoffs:
    1. The difference between the challenger's payoff when the defender has a reputation for resolve and its payoff in the single-shot game if the defender has the same probability of being resolute (in Case A) is: $(\Pi - \theta - \eta) -$ $(\Pi - \theta)(1 - \rho) = -\eta -$ $(\Pi - \theta)(-\rho) = -\eta + \rho(\Pi - \theta)$.
    2. The difference between the payoff of a defender that is irresolute in the present dispute and has a reputation for resolve and one that is irresolute without a reputation is $\eta$. There is no difference between the payoff of a defender that is resolute in the present dispute and has a reputation for resolve and one that is resolute without a reputation. (By 1 and 3for both claims).
5. Since $\rho \geq \frac{\eta}{\Pi - \theta}$ by the Reputation Condition and $\rho, (\Pi - \theta) > 0$, $\rho(\Pi - \theta) \geq \eta$. Thus, the difference between the challenger's payoff when the defender has a reputation and when it does not is weakly positive. As long as the reputation condition holds as a strict inequality, ($\rho > \frac{\eta}{\Pi - \theta}$), the challenger's payoff when the defender has a reputation is strictly greater than when it does not.

**Case B. The defender does not have a reputation for resolve (has a reputation for lacking resolve).**

1. In the reputational equilibrium, the challenger makes the low offer and its payoff is $(\Pi - \theta)$ if the offer is accepted and 0 otherwise. By proposition 1, an irresolute

defender accepts the offer and its payoff is $\theta$; a resolute defender rejects the offer and its payoff is $\theta + \eta$.

2. By Proposition 1 the challenger also makes the low offer in the second-period equilibrium when the defender does not have a reputation. Since the defender's acceptance strategy also is the same in these situations, both states' payoffs must be identical.■

### 7.6. Multiple Challengers

#### 7.6.1. Equilibrium and Comparative Statics

In the first version of the model, the same two states are assumed to interact in the two disputes. I now modify the model to assume the existence of two challengers, one of which interacts with the defender in period 1 and the other of which interacts with the defender in period 2. Each challenger's payoff is simply its payoff for the period in which it interacts. However, the second-period challenger observes the defender's first dispute with the other challenger and potentially can learn about the defender's type from its interactions in that dispute. Later, I consider the more-elegant model in which there is some probability $\lambda$ that the same states will interact again; this yields the same implication.[33]

Since the second dispute is identical to that in the other model, the analyses of the second period are identical and Proposition 1 holds for this model. Moreover, since the defender expects second-period interactions as in the basic model, its acceptance strategy remains the same. Thus, the only way in which analysis of this model differs from that of the other is that the first-period challenger's payoffs from its possible offers only depend on the value of the offers and the defender's reaction to the offers today. The modified table shows the expected payoffs for the first-period challenger ($C_1$).

The set of equilibria is the same as in the basic model, but the thresholds differ. (The table below gives the challenger's payoffs for this version; the equilibria are as in the table for the version with a single challenger.) The challenger prefers to screen as opposed to pool at the high offer in the first period when:

[33] I thank Bob Powell for this suggestion.

$$\begin{aligned}(1-\zeta_1)(\Pi-\theta-\delta\eta\rho) &\geq \Pi-(\eta+\theta+\delta\eta(1-\rho))\\ \zeta_1 &\leq \frac{\eta(1-\delta\varphi)}{(\Pi-\theta-\delta\eta\rho)} \equiv \zeta_1^{TMC}\end{aligned}$$

where $\zeta_1^{TMC}$ is $\zeta_1^T$ for the model with multiple challengers.

In the model with a single challenger, the analogous threshold was

$$\zeta_1^{TSC} = \frac{\eta(1-\delta\varphi)}{(1-\delta\rho)(\Pi-\theta)+\delta\eta(1-\rho)} = \frac{\eta(1-\delta\varphi)}{(\Pi-\theta-\delta\eta\rho)+\delta[\eta-\rho(\Pi-\theta)]}.$$

The two thresholds are identical except for the second term in the denominator. The numerator is positive because $0 < \varphi < 1$ by Assumption 1; and $\eta > 0$, $0 < \delta < 1$. The common term in the denominator also is positive by Assumption 2 and $\delta < 1$. Thus, the difference in thresholds is determined by the sign of the second term in the denominator. This term, $\delta[\eta-\rho(\Pi-\theta)]$ is negative when the reputation condition is satisfied, because one part of the condition is that $\rho > \zeta_2^*$, and $\zeta_2^* = \frac{\eta}{\Pi-\theta}$ by Proposition 1. Thus, $\zeta_1^{TSC} < \zeta_1^{TMC}$. Intuitively, when the challenger expects to interact with the defender again, it is more willing to screen (willing to screen even if it is somewhat more pessimistic) because learning about the defender's type provides it with valuable information it can use in the next dispute; even though the defender's type may change between disputes, it is more likely to be the same in the next dispute than it is to be different.

- The first-period threshold belief $\zeta_1^{TMC}$ decreases as the persistence of resolve increases.

$$\frac{\partial\zeta_1^{TMC}}{\partial\rho} = \frac{-2\delta\eta(\Pi-\theta-\delta\eta\rho)-\eta(1-\delta\varphi)(-\delta\eta)}{(\Pi-\theta-\delta\eta\rho)^2}$$

The sign of the partial derivative is again the sign of the numerator, which can be simplified to

$$-\delta\eta[2(\Pi-\theta)-\eta-\delta\eta].$$

By Assumption 2, $\Pi-\theta > \eta$. Since $0 < \delta < 1$, the quantity in square brackets is positive, and since $\delta, \eta > 0$, the numerator is negative. Thus, $\frac{\partial\zeta_1^{TMC}}{\partial\rho} < 0$, and the first-period threshold belief decreases as the persistence of resolve increases. When the first-period threshold belief decreases, screening exists for a smaller range of prior beliefs and pooling for a larger range, and the formation of consequential reputations on the equilibrium path becomes less likely.

- As in the game with a single challenger, the threshold between screening and pooling at the high offer rises as the challenger becomes more uncertain of the defender's value for war. Put loosely, reputation formation is thus more likely in situations of greater uncertainty. I use the same change of variables as before ($\phi \equiv \frac{2\theta+\eta}{2}$, or $\theta = \phi - \frac{\eta}{2}$).

$$\zeta_1^{TMC} = \frac{\eta(1-\delta\varphi)}{(\Pi - \theta - \delta\eta\rho)} = \frac{\eta(1-\delta\varphi)}{(\Pi - (\phi - \frac{\eta}{2}) - \delta\eta\rho)}$$

$$\frac{\partial\zeta_1^{TMC}}{\partial\eta} = \frac{(\Pi - (\phi - \frac{\eta}{2}) - \delta\eta\rho)(1-\delta\varphi) - \eta(1-\delta\varphi)(\frac{1}{2} - \delta\rho)}{(\Pi - (\phi - \frac{\eta}{2}) - \delta\eta\rho)^2}$$

After taking the partial, I change the variables back to the original in order to investigate the sign. Then

$$\frac{\partial\zeta_1^{TMC}}{\partial\eta} = \frac{(\Pi - \theta - \frac{\eta}{2})(1-\delta\varphi)}{(\Pi - \theta - \delta\eta\rho)^2}.$$

The sign of the partial is the sign of the numerator, since the denominator is squared. We know that $1 - \delta\varphi > 0$ because $\varphi = 2\rho - 1 < 1$ and $\delta < 1$. Thus, the sign is that of

$$\Pi - \theta - \frac{\eta}{2}$$

This quantity is positive since $\Pi - \theta - \eta > 0$ by Assumption 2. Thus, *increasing the degree of uncertainty raises the threshold between screening and pooling*, making screening, and reputation formation, more likely.

### 7.6.2. Single and Multiple Challengers as Subgames of a Single Game

In many disputes, both challenger and defender may have some uncertainty over whether or not they will interact again any time soon. In this model, one way to conceptualize this uncertainty is as a game in which the challenger does not know whether today will be its only interaction with this defender or whether it will interact with this defender again in the future. (The defender's incentives are identical in the analyses of the two versions presented above.)

Consider the game as described earlier, but let the probability that the two states involved in the first dispute will interact again in the second be $\lambda$. If they do not do so, then the defender will interact with another challenger in the second dispute.

In this case

$$\begin{aligned} U^C[screen] &= \lambda\left[(1-\zeta_1)[(\Pi-\theta-\delta\eta\rho)+\delta(\Pi-\theta)\rho]+\zeta_1\delta[\Pi-(\eta+\theta)]\right. \\ &\quad +(1-\lambda)(1-\zeta_1)(\Pi-\theta-\delta\eta\rho) \end{aligned}$$

$$\begin{aligned} U^C[pool] &= \lambda\left[\Pi-(\eta+\theta+\delta\eta(1-\rho))+\delta(\Pi-\theta)((\rho-\zeta_1\varphi)\right] \\ &\quad +(1-\lambda)\left[\Pi-(\eta+\theta+\delta\eta(1-\rho))\right] \end{aligned}$$

The challenger prefers to screen when $U^C[screen] \geq U^C[pool]$. Since the challenger's payoff from screening (pooling) has a probability$\lambda$ of being the single-challenger payoff and a probability $(1-\lambda)$ of being the multiple-challenger payoff, the threshold between screening and pooling is

$$\lambda\zeta_1^{TSC}+(1-\lambda)\zeta_1^{TMC}.$$

or

$$\zeta_1^T=\frac{\lambda\eta(1-\delta\varphi)}{(1-\delta\rho)(\Pi-\theta)+\delta\eta(1-\rho)}+\frac{(1-\lambda)\eta(1-\delta\varphi)}{(\Pi-\theta-\delta\eta\rho)}$$

Since $\zeta_1^{TSC}<\zeta_1^{TMC}$, the threshold is rising in $\lambda$; that is, as the challenger becomes more certain it will interact again with the defender, it is more likely to make the screening offer in the first period that can lead to reputation formation.

Similarly, the threshold between partial screening and screening is a weighted function of the thresholds in the single-challenger and multiple-challenger versions:

$$\lambda\zeta_1^{1SC}+(1-\lambda)\zeta_1^{1MC}.$$

### 7.7. When the Second Dispute is of Greater Magnitude

Thusfar, the model allows the magnitude of the two disputes to differ only in that the defender may be resolute in one (consider it more important) and irresolute in the other (consider it less important). I now consider two situations in which the second dispute may be altogether of greater magnitude, assuming that the challenger will make a low offer in the second dispute if it learns nothing from the defender's first-period actions.

As I discuss in the text, the first variant assumes that the states in the first dispute expect the second to be of larger magnitude ("bigger" in all ways); to do this, I multiply the payoffs $\Pi$, $\eta$, and $\theta$ in the second dispute, which represent the different components of the states' values for the disputed issue (observed and unobserved), by a constant greater than one. Mathematically, this is equivalent to assuming that the discount factor is greater than one, so I investigate what happens when the assumption of discount factor ($\delta$) less than one does not hold.

In this case, a reputation becomes extremely valuable, and the reputational equilibrium may fail to exist, because the irresolute defender would need to be compensated more in the first dispute for its failure to acquire a reputation than the maximum the challenger can grant, $\Pi$. That is, the offer in the reputational equilibrium is $\theta+\eta+\delta\rho\eta$; while $\Pi > \theta+\eta$ by assumption 2, there is no reason to expect that $\Pi > \theta+\eta+\delta\rho\eta$ for large $\delta > 1$. Interestingly, the pooling equilibrium that I use for my comparative statics also may fail to exist, as it compensates a resolute defender for its failure to acquire a reputation.

The second variant of the model assumes that states in the first dispute expect their next dispute to be of similar magnitude to the current one; thus, the incentives for reputation formation are identical to those in the basic model. However, when the second dispute arises, it turns out to be greater magnitude, one in which the payoffs $\Pi$, $\eta$, and $\theta$ are multiplied by a constant greater than one, which I will call $c$. The question, then, is whether any reputation acquired in the first dispute has an effect in the second dispute, which unexpectedly concerns a more vital state interest.

As long as the persistence of private resolve is governed by the same process, reputation does have an effect in this situation. Looking at the second dispute, the challenger's belief at the start of the second dispute remains $\zeta_2 = \rho$ if the defender rejected the offer in the first dispute (has a reputation for resolve) and $\zeta_2 = 1-\rho$ if it accepted the offer. With these new payoffs, the challenger's expected payoff in the case of a screening offer is

$\zeta_2 * 0 + (1-\zeta_2)c(\Pi - \theta)$, and its payoff in the case of a pooling offer is $c(\Pi - \eta - \theta)$, so that it makes a low, screening offer when

$$\zeta_2 < \frac{c\eta}{c(\Pi - \theta)} \equiv \zeta_2^*.$$

That is, behavior in the second dispute is exactly as before, with the challenger having the same threshold belief between screening and pooling; the only difference is that all payoffs are of greater magnitude.

### 7.8. Equilibrium Tables

The tables below, versions of those in Kennan (1998) but with the notation and precise form of the model used in this paper, show the possible first-period offers, the defender's strategy ($\sigma_1^D$) if they are made, and the challenger's expected payoff for the whole game if it makes each offer. (See Kennan (1998, page 26).The defender's strategy column summarizes the acceptance strategies that are possible in an equilibrium of the subgame starting with the offer in the first column. In some cases the subgame has more than one equilibrium, so that the $\sigma_1^D$ column lists more than one strategy. The tables summarize all the equilibria, not just the ones that I use in my analyses.

Once the equilibria of the subgames starting with the defender's first-period decision are established, what remains is to determine the challenger's optimal offer strategy in the first period. The second table shows the possible offers and shows which are part of equilibria, and which, on the other hand, are dominated. For any dominated offer, the last column shows at least one offer that dominates the offer, to demonstrate that the dominated offer cannot be part of an equilibrium. This table identifies five offers or sets of offers that are possible in some equilibrium. There is one offer that is made in a semi-screening equilibrium, one offer that is made in a screening equilibrium, and a range of offers that can be made in pooling equilibria. Note that an offer may be dominated for the challenger for one equilibrium response by the defender in the subgame but not for another. For example, a first-period offer of $\theta$ is dominated for the challenger when the irresolute defender will accept with probability less than $Q$, but not when it will accept with probability $Q$, where

$$Q \equiv \frac{1}{(1-\zeta_1)} - \frac{\zeta_1 \varphi}{[\zeta_2^* - (1-\rho)](1-\zeta_1)}. \tag{7.1}$$

The existence of the semi-screening equilibrium is perhaps not obvious. In this equilibrium, the challenger offers exactly the irresolute defender's value for war, $\theta$. The resolute defender always rejects the offer, and the irresolute one rejects it just often enough that the challenger is just optimistic/pessimistic enough at the start of the second dispute to be indifferent between making a low (screening) or a high (pooling) offer in that dispute. If rejection of the offer were to lead the challenger to make the high (pooling) offer in the second dispute, the irresolute defender never would accept the offer, since the offer's period-1 value equals its value for war. However, since the challenger is indifferent between screening and pooling in the second dispute (given the defender's first-period randomization), it screens in the second dispute regardless of whether the first-period offer is accepted or rejected. Thus, the irresolute defender's continuation value is identical if it accepts or rejects the offer, and it is willing to randomize.

The subtlety lies in why the challenger does not deviate to an offer slightly higher than the irresolute defender's value for war, $\theta + \epsilon$, to coax all irresolute defenders into accepting; the reason is that in the equilibrium of the subgame following offers $a_1 \in (\theta, \theta + \delta\eta\rho)$, the defender also randomizes, accepting with the same probability $Q$ as in the subgame in which the challenger offers $\theta$. (In a one-period version of the model, an irresolute defender would accept any offer greater than its value for war with probability one, but it does not do so in the first period of the two-period model for reputational reasons.) For this reason, if the challenger deviates and offers $\theta + \epsilon$ instead of $\theta$, its period-1 payoff is slightly lower, $Q(\Pi - (\theta + \epsilon))$ instead of $Q(\Pi - \theta)$, and its expected continuation value is identical; in either case, if the defender in the first period rejects the offer, the challenger enters the second period with beliefs that make it indifferent between screening and pooling, so that which offer it makes is payoff irrelevant (though not payoff irrelevant to an irresolute defender). Thus, the challenger has no incentive to deviate from an offer of $\theta$ to $\theta + \epsilon$.

Which equilibria exist depend upon the challenger's initial beliefs, $\zeta_1$. For given initial beliefs, the challenger will choose a strategy that gives the highest expected payoff; each expected payoff takes into account that the states will play equilibrium strategies for the remainder of the game. As I explain earlier, I use only two equilibria for the comparative statics, the one with screening and the one with pooling at the highest possible pooling equilibrium offer.

| $\mathbf{a}_1$ | $\boldsymbol{\sigma}_1^D$ | **Challenger's expected payoff** |
|---|---|---|
| $< \theta$ | $\sigma_1^L(a_1) = \sigma_1^H(a_1) = 0$ | $0 + \delta(\Pi - \theta)(\rho - \zeta_1\varphi)$ |
| $\theta$ | $\sigma_1^L(a_1) = Q, \sigma_1^H(a_1) = 0$ | $(1 - \zeta_1)Q(\Pi - \theta) + \delta(\Pi - \theta)(\rho - \zeta_1\varphi)$ |
| $\theta$ | $\sigma_1^L(a_1) \in [0, Q); \sigma_1^H(a_1) = 0$ | $(1 - \zeta_1)\sigma_1^L(\Pi - \theta) + \delta(\Pi - \theta)(\rho - \zeta_1\varphi)$ |
| $(\theta, \theta + \delta\eta\rho)$ | $\sigma^L(a_1) = Q; \sigma^H(a_1) = 0$ | $(1 - \zeta_1)[Q(\Pi - a_1 + \delta(\Pi - \theta)\rho)]$ $+[(1 - \zeta_1)(1 - Q) + \zeta_1]\delta(\Pi - (\eta + \theta))$ |
| $\theta + \delta\eta\rho$ | $1 > \sigma^L(a_1) \geq Q; \sigma^H(a_1) = 0$ | $\sigma^L(a_1)(1 - \zeta_1)(\Pi - \theta - \delta\eta\rho)$ $+ \delta[\sigma^L(a_1)(1 - \zeta_1)(\Pi - \theta)\rho$ $+(\zeta_1 + (1 - \zeta_1)(1 - \sigma^L(a_1))\delta(\Pi - (\eta + \theta))]$ |
|  | $\sigma^L(a_L) = 1, \sigma^H(a_L) = 0$ | $(1 - \zeta_1)[(\Pi - \theta - \delta\eta\rho) + \delta(\Pi - \theta)\rho]$ $+\zeta_1\delta[\Pi - (\eta + \theta)]$ |
| $(\theta + \delta\eta\rho, \eta + \theta)$ | $\sigma^L(a_1) = 1,\ \sigma^H(a_1) = 0$ | $(1 - \zeta_1)[(\Pi - a_1) + \delta(\Pi - \theta)\rho]$ $+\zeta_1\delta[\Pi - (\eta + \theta)]$ |
| $\eta + \theta$ | $\sigma^L(a_1) = 1,\ \sigma^H(a_1) = 1$ | $\Pi - (\eta + \theta) + \delta(\Pi - \theta)(\rho - \zeta_1\varphi)$ |
|  | $\sigma^L(a_1) = 1,\ \sigma^H(a_1) = 0$ | $(1 - \zeta_1)[\Pi - (\eta + \theta) + \delta(\Pi - \theta)\rho]$ $+\zeta_1\delta(\Pi - (\eta + \theta))$ |
| $(\eta + \theta,$ $\eta + \theta + \delta\eta(1 - \rho))$ | $\sigma^L(a_1) = 1, \sigma^H(a_1) = 0$ | $(1 - \zeta_1)[\Pi - a_1 + \delta(\Pi - \theta)\rho]$ $+\zeta_1\delta(\Pi - (\eta + \theta))$ |
|  | $\sigma^L(a_1) = 1, \sigma^H(a_1) = 1$ | $\Pi - a_1 + \delta(\rho - \varphi\zeta_1)(\Pi - \theta)$ |
| $\eta + \theta + \delta\eta(1 - \rho)$ | $\sigma^L(a_1) = 1,\ \sigma^H(a_1) \in (0, 1)$ | $\zeta_1\sigma^H(a_1)[(\Pi - (\eta + \theta + \delta\eta(1 - \rho))$ $+\delta(1 - \rho)(\Pi - \theta)]$ $+(1 - \zeta_1)[\Pi - (\eta + \theta + \delta\eta(1 - \rho))$ $+\delta\rho(\Pi - \theta)] + \zeta_1(1 - \sigma^H(a_1))\delta(\Pi - (\eta + \theta))$ |
|  | $\sigma^L(a_1) = 1,\ \sigma^H(a_1) = 1$ | $\Pi - (\eta + \theta + \delta\eta(1 - \rho)) + \delta(\Pi - \theta)((\rho - \zeta_1\varphi)$ |
|  | $\sigma^L(a_1) = 1,\ \sigma^H(a_1) = 0$ | $(1 - \zeta_1)[\Pi - (\eta + \theta + \delta\eta(1 - \rho)) + \delta(\Pi - \theta)\rho]$ $+\zeta_1\delta(\Pi - (\eta + \theta))$ |
| $> \eta + \theta + \delta\eta(1 - \rho)$ | $\sigma^L(a_1) = \sigma^H(a_1) = 1$ | $\Pi - a_1 + \delta(\rho - \varphi\zeta_1)(\Pi - \theta)$ |

$a_1$: 1st period offer; $\sigma^i(a_1)$: probability that defender of type $i$ accepts this $a_1$

| $a_1$ | $\sigma_1^D$ | Dominated by | **Eqm. type** |
|---|---|---|---|
| $< \theta$ | $\sigma^L(a_1) = \sigma_1^H(a_1) = 0$ | $\theta$ | |
| $\boldsymbol{\theta}$ | $\sigma^L(a_1) = Q, \sigma^H(a_1) = 0$ | none | **part. screen** |
| $\theta$ | $\sigma_1^L(a_1) \in [0, Q); \sigma^H(a_1) = 0$ | $\theta + \epsilon$ | |
| $(\theta, \theta + \delta\eta\rho)$ | $\sigma^L(a_1) = Q; \sigma^H(a_1) = 0$ | $a_1 - \epsilon$ | |
| $\boldsymbol{\theta + \delta\eta\rho}$ | $1 > \sigma^L(a_1) \geq Q; \sigma^H(a_1) = 0$ | $\theta + \delta\eta\rho + \epsilon$ | |
| | $\sigma^L(a_1) = 1, \sigma^H(a_1) = 0$ | none | **screen** |
| $(\theta + \delta\eta\rho, \eta + \theta)$ | $\sigma^L(a_1) = 1,\ \sigma^H(a_1) = 0$ | $a_1 - \epsilon$ | |
| $\boldsymbol{\eta + \theta}$ | $\sigma^L(a_1) = 1,\ \sigma^H(a_1) = 1$ | none | **low pool** |
| | $\sigma^L(a_1) = 1,\ \sigma^H(a_1) = 0$ | $(\theta + \delta\eta\rho, \eta + \theta)$ | |
| $\boldsymbol{(\eta + \theta, \eta + \theta + \delta\eta(1 - \rho))}$ | $\sigma^L(a_1) = 1, \sigma^H(a_1) = 0$ | $(\theta + \delta\eta\rho, \eta + \theta)$ | |
| | $\sigma^L(a_1) = 1, \sigma^H(a_1) = 1$ | none | **med. pool** |
| $\boldsymbol{\eta + \theta + \delta\eta(1 - \rho)}$ | $\sigma^L(a_1) = 1,\ \sigma^H(a_1) \in (0, 1)$ | $a_1 + \epsilon$ | |
| | $\sigma^L(a_1) = 1,\ \sigma^H(a_1) = 1$ | none | **high pool** |
| | $\sigma^L(a_1) = 1,\ \sigma^H(a_1) = 0$ | $(\theta + \delta\eta\rho, \eta + \theta)$ | |
| $> \eta + \theta + \delta\eta(1 - \rho)$ | $\sigma^L(a_1) = \sigma^H, (a_1) = 1$ | $a_1 - \epsilon$ | |

$a_1$: 1st period offer;

$\sigma^i(a_1)$: probability that defender of type $i$ accepts this $a_1$ in subgame equilibrium

| $a_1$ | $\sigma_1^D$ | Dominated by | **Eqm. type** |
|---|---|---|---|
| $< \theta$ | $\sigma^L(a_1) = \sigma_1^H(a_1) = 0$ | $\theta$ | |
| $\boldsymbol{\theta}$ | $\sigma^L(a_1) = Q, \sigma^H(a_1) = 0$ | none | **part. screen** |
| $\theta$ | $\sigma_1^L(a_1) \in [0, Q); \sigma^H(a_1) = 0$ | $\theta + \epsilon$ | |
| $(\theta, \theta + \delta\eta\rho)$ | $\sigma^L(a_1) = Q; \sigma^H(a_1) = 0$ | $a_1 - \epsilon$ | |
| $\boldsymbol{\theta + \delta\eta\rho}$ | $1 > \sigma^L(a_1) \geq Q; \sigma^H(a_1) = 0$ | $\theta + \delta\eta\rho + \epsilon$ | |
| | $\sigma^L(a_1) = 1, \sigma^H(a_1) = 0$ | none | **screen** |
| $(\theta + \delta\eta\rho, \eta + \theta)$ | $\sigma^L(a_1) = 1,\ \sigma^H(a_1) = 0$ | $a_1 - \epsilon$ | |
| $\boldsymbol{\eta + \theta}$ | $\sigma^L(a_1) = 1,\ \sigma^H(a_1) = 1$ | none | **low pool** |
| | $\sigma^L(a_1) = 1,\ \sigma^H(a_1) = 0$ | $(\theta + \delta\eta\rho, \eta + \theta)$ | |
| $\boldsymbol{(\eta + \theta, \eta + \theta + \delta\eta(1 - \rho))}$ | $\sigma^L(a_1) = 1, \sigma^H(a_1) = 0$ | $(\theta + \delta\eta\rho, \eta + \theta)$ | |
| | $\sigma^L(a_1) = 1, \sigma^H(a_1) = 1$ | none | **med. pool** |
| $\boldsymbol{\eta + \theta + \delta\eta(1 - \rho)}$ | $\sigma^L(a_1) = 1,\ \sigma^H(a_1) \in (0, 1)$ | $a_1 + \epsilon$ | |
| | $\sigma^L(a_1) = 1,\ \sigma^H(a_1) = 1$ | none | **high pool** |
| | $\sigma^L(a_1) = 1,\ \sigma^H(a_1) = 0$ | $(\theta + \delta\eta\rho, \eta + \theta)$ | |
| $> \eta + \theta + \delta\eta(1 - \rho)$ | $\sigma^L(a_1) = \sigma^H, (a_1) = 1$ | $a_1 - \epsilon$ | |

$a_1$: 1st period offer;

$\sigma^i(a_1)$: probability that defender of type $i$ accepts this $a_1$ in subgame equilibrium

| $\mathbf{a}_1$ | $\boldsymbol{\sigma}_1^D$ | $C_1$**'s Expected Payoff if Multiple Challengers** |
|---|---|---|
| $< \theta$ | $\sigma_1^L = \sigma_1^H = 0$ | $0$ |
| $\theta$ | $\sigma_1^L = Q, \sigma_1^H = 0$ | $(1-\zeta_1)Q(\Pi - \theta)$ |
| $\theta$ | $\sigma_1^L \in [0, Q); \sigma_1^H = 0$ | $(1-\zeta_1)\sigma_1^L(\Pi - \theta)$ |
| $(\theta, \theta + \delta\eta\rho)$ | $\sigma^L(a_1) = Q; \sigma^H(a_1) = 0$ | $(1-\zeta_1)Q(\Pi - a_1)$ |
| $\theta + \delta\eta\rho$ | $1 > \sigma^L(a_1) \geq Q; \sigma^H(a_1) = 0$ | $\sigma^L(a)(1-\zeta_1)(\Pi - \theta - \delta\eta\rho)$ |
| | $\sigma^L(a_1) = 1, \sigma^H(a_1) = 0$ | $(1-\zeta_1)[(\Pi - \theta - \delta\eta\rho)$ |
| $(\theta + \delta\eta\rho, \eta + \theta)$ | $\sigma^L(a_1) = 1,\ \sigma^H(a_1) = 0$ | $(1-\zeta_1)(\Pi - a_1)$ |
| $\eta + \theta$ | $\sigma^L(a_1) = 1,\ \sigma^H(a_1) = 1$ | $\Pi - (\eta + \theta)$ |
| | $\sigma^L(a_1) = 1,\ \sigma^H(a_1) = 0$ | $(1-\zeta_1)[\Pi - (\eta + \theta)]$ |
| $(\eta + \theta,$ $\eta + \theta + \delta\eta(1-\rho))$ | $\sigma^L(a_1) = 1, \sigma^H(a_1) = 0$ | $(1-\zeta_1)(\Pi - a_1)$ |
| | $\sigma^L(a_1) = 1, \sigma^H(a_1) = 1$ | $\Pi - a_1$ |
| $\eta + \theta + \delta\eta(1-\rho)$ | $\sigma^L(a_1) = 1,\ \sigma^H(a_1) \in (0, 1)$ | $[(1-\zeta_1) + \zeta_1\sigma^H(a_1)]*$ $[\Pi - (\eta + \theta + \delta\eta(1-\rho))]$ |
| | $\sigma^L(a_1) = 1,\ \sigma^H(a_1) = 1$ | $\Pi - (\eta + \theta + \delta\eta(1-\rho))$ |
| | $\sigma^L(a_1) = 1,\ \sigma^H(a_1) = 0$ | $(1-\zeta_1)[\Pi - (\eta + \theta + \delta\eta(1-\rho))$ |
| $> \eta + \theta + \delta\eta(1-\rho)$ | $\sigma^L(a_1) = \sigma^H(a_1) = 1$ | $\Pi - a_1$ |

$a_1$: 1st period offer

$\sigma^i(a_1)$: probability that defender of type $i$ accepts this $a_1$ in subgame equilibrium