When Capabilities Backfire: How Improved Hassling Capabilities Produce Worse Outcomes *

Peter Schram†

January 11, 2021

Abstract

I formalize interactions between an endogenously rising challenger state and a rival, defender state that can accept the challenger's rise, go to war before the rise comes to fruition, or degrade the challenger's growth through low-level conflict operations that I call "hassling." The novelty here is that the defender has private information about its hassling capabilities; this implies that the challenger does not know how much it can rise without provoking the defender to hassle or go to war. I find that when the defender's ability to hassle improves, it can provoke a strategic response in the challenger that undermines the defender's ability to use its private information productively and results in lower utility for the defender. This model provides insight into both Saddam Hussein's decision-making leading up to the 2003 U.S. invasion and the stability-instability paradox.

*This paper was formerly entitled "When Deterrence Fails." Thanks to Brad Smith, Andrew Coe, Brenton Kenkel, Kris Ramsay, William Spaniel, James Fearon, Andrew Little, Michael Joseph, Braden Sopher, CPASS, the Virtual Formal Theory Workshop, and the Vanderbilt Conflict Workshop and Faculty workshop for helpful comments. I am indebted to Haonan Dong, Ethan Sun, and the Vanderbilt ROCCA lab for outstanding research assistance.

†Assistant Professor, Vanderbilt Department of Political Science

Leading up to 2003, Saddam Hussein repeatedly denied entry to United Nations weapons inspectors and made false statements about Iraq's chemical and biological weapons. These behaviors led Iraq's adversaries to believe that Iraq was developing weapons of mass destruction (WMD) when, in reality, Iraq had no nuclear weapons, and its pursuit of WMD was limited. Saddam had good reason to keep inspectors out: he believed that the inspectors would detail the extent and location of his armaments, information that could be used maliciously by Iraq's adversaries (Koblentz, 2018; Coe and Vaynman, 2020). Still, the behavior is puzzling given that it was the U.S., a much more powerful state, that was demanding inspectors be allowed in. The 1990–91 Gulf War illustrated that the Iraqi army was no match for U.S. forces. Additionally, transcripts from the Saddam regime suggested that Iraq knew that its behavior around WMD issues put it at risk for a confrontation with the United States. The well-known logic of commitment problems identifies this scenario as one in which preventive war is likely (Fearon, 1995; Powell, 2006; Debs and Monteiro, 2014; Gurantz and Hirsch, 2017)—which is ultimately what happened. The puzzle, then, is why was Saddam not deterred by the might of the U.S. military, when it seems that logically he should have allowed the inspectors in to avoid provoking a war he had little chance of winning.

Saddam was willing to gamble on keeping weapons inspectors out because he did not expect he would face the full might of the U.S. military. In 1998, in response to Iraq turning away weapons inspectors, the U.S. conducted Operation Desert Fox, a four-day bombing campaign against Iraqi weapons facilities. In this operation, the U.S. illustrated that it had the ability and willingness to use targeted low-level conflict operations in response to rising powers. If Saddam believed the U.S. was more likely to respond as it had in 1998 than it had in 1991, then there exists a troubling possibility: by being effective at low-level operations such as limited bombing campaigns, the U.S. may have undermined its own threat from war and emboldened Saddam to behave as he did. While low-level conflict capabilities can be useful in political crises—Operation Desert Fox was, at the time, a success—they may provoke problematic strategic responses from opponents. Today, the availability of technologies such

as cyberattacks, drone strikes, and precision strikes, tools that can cheaply and precisely destroy a nuclear program outside of a preventive war, may actually lead to more low-level conflict, or, in the case of the 2003 invasion, more war.

I formalize the above intuition. I consider a challenger that chooses how much to "transgress" (following Gurantz and Hirsch (2017)), an act that will make the challenger more powerful in the future. In response, a rival defender can accept the transgression, initiate a preventive war to stop the transgression, or "hassle," which, following Schram (Forthcoming), is the use of limited conflict to degrade the challenger's transgression while allowing bargaining to continue. What distinguishes this work from other formal research on conflict short of war (Bas and Coe, 2016; McCormack and Pascoe, 2017; Coe, 2018; Spaniel, 2019; Schram, Forthcoming; Joseph, 2020) is that the defender's hassling capabilities have both public and private components. The challenger uses its observations of the defender's public hassling capabilities in making its transgression decision, but it does not observe the defender's exact private willingness to hassle. Thus, the challenger does not know how much it can transgress before it provokes hassling or war. In this setting, I show that when the defender possesses better public hassling capabilities, it can produce worse outcomes—more aggressive transgressions, more war, and more hassling—which result in overall lower utility for the defender.

Improvements in hassling capabilities produce worse outcomes for the defender in one of two ways. First, improvements in the defender's hassling capabilities can *embolden the challenger*. When Iraq chose to turn away weapons inspectors and lie about its capabilities, it did so knowing that the United States may respond with force. And, while war was a possible response, because the U.S. was good at hassling (as evidenced by, for example, Operation Desert Fox), Iraq expected hassling was the most likely response, and therefore determined that the benefits of turning away weapons inspectors outweighed the risks of the expected military response. In a counterfactual setting where the U.S. was worse at hassling, Iraq might have expected that refusing entry to inspectors would be met with a

greater likelihood of war. In this counterfactual setting, the increased risk of war could have overshadowed whatever gains Iraq expected from turning away weapons inspectors and convinced Iraq to be more open to inspectors. In other words, because the U.S. was good at hassling, it incentivized risk-taking behavior and emboldened Iraq to gamble.

Second, improvements in the defender's hassling capabilities can produce worse outcomes for the defender by *making the defender more predictable.* In the model, possessing private information is useful because it allows for posturing to attain better outcomes. The threat of war is an effective deterrent; if a challenger thinks that building a nuclear bomb will provoke its rivals to declare war today, then the challenger will not do so. When the defender postures effectively, it convinces the challenger that the defender would go to war over a transgression when, in fact, the defender would not have gone to war. Under some conditions, improvements in the defender's hassling capabilities diminish its ability to posture effectively, thus allowing the challenger to calibrate its transgressions to extract more bargaining surplus from the defender. I examine this mechanism in the context of the proxy conflicts that occurred during the Cold War and the stability-instability paradox.

This paper identifies new ways improved low-level conflict capabilities can lead to worse outcomes for the state making the improvements. Existing research has demonstrated similar results under alternate theoretical frameworks (Schelling, 1966; Powell, 1989, 2015; Bas and Coe, 2016; Spaniel and Malone, 2019; Baliga *et al.*, 2020). For example, while Bas and Coe and Baliga *et al.* show that better low-level capabilities can undermine a deterrent threat though a "failure-to-burn-bridges" logic,[1] those results would not arise if not for key model assumptions, namely, a stochastically-observed transgression process or ambiguous attribution (respectively). I show that improved low-level conflict capabilities can lead to worse outcomes when neither of the aforementioned assumptions are present, but there is uncertainty over a state's willingness to conduct low-level conflict. This is not to say that

[1]The logic of burning bridges is quite different from the predictability and emboldening mechanisms described here; see the Appendix for a discussion.

existing results arise from empirically invalid assumptions—for example, the model in Baliga *et al.* is built to describe cyberattacks and terrorism with uncertain attribution. Rather, a strength of this paper is that its key assumption follows from a foundational principle of security studies: that states possess private information over their willingness to engage in conflict (Waltz, 1979; Blainey, 1988; Fearon, 1995; Slantchev and Tarar, 2011; Fey and Ramsay, 2011; Ramsay, 2017). This paper's primary contribution is that it further clarifies the relationship between conflict capabilities and outcomes when there is a new type of private willingness to conduct conflict, thus advancing our understanding of how crises evolve when actors have multiple means of conducting conflict. Given the continued advancement of precision strike capabilities, information technologies, unmanned aerial vehicles, artificial intelligence, robotics and other technologies—all of which can be used in low-level conflict and classified as hassling capabilities—this is a relationship worth thoroughly understanding.

1 Theoretical Background

1.1 Key Concepts and Terminology

This paper references several types of political behaviors or capabilities. Here I provide definitions of these concepts and illustrate them using Operation Outside the Box. Operation Outside the Box was an Israeli hassling operation in response to the discovery that Syria was building a nuclear reactor in Al Kibar in 2007. In it, Israel used an electronic warfare attack to disable Syrian air defenses and conducted an airstrike on the reactor (Katz, 2010). Because building a reactor moved Syria closer to being able to produce a nuclear weapon, using the term as it is also used in Gurantz and Hirsch (2017), this was a "transgression."

Definition: *Actors* ***transgress*** *by taking actions that strengthen their future military capabilities.*

If transgressions are allowed to come to fruition, then the transgressing state will have

improved war-fighting abilities in future conflicts, thereby attaining more leverage in future negotiations. Transgressions can include investing in military nuclear, space, electronic, or cyber technologies (Lindsay *et al.*, 2016; Buchanan, 2020), amassing conventional forces (Copeland, 2001), supporting militants in rival states (Schultz, 2010), or securing valuable territory (Fearon, 1996; Powell, 2006). Through transgressions, the challenger may create a commitment problem (Debs and Monteiro, 2014; Bas and Coe, 2016). In response, rival states may initiate a preventive war to stop the transgression. But rival states have another option short of war: they can "hassle" the challenger, which was previously defined in Schram (Forthcoming). When Israel degraded Syria's nuclear facilities through an electronic warfare attack and airstrike, it engaged in hassling.

Definition: Hassling *is the limited use of costly and destructive military assets against another state with the intent of blunting power shifts to allow for bargaining to occur.*

Hassling operates like a steam valve: in an international system where a rising power may provoke rivals to declare a preventive war, hassling can diffuse the situation at a lower cost to both parties than war. While war halts a transgression and resolves the political crisis (as it is in the common "costly lottery" treatment of war (Fearon, 1995)), hassling degrades the transgression while still allowing for negotiations to occur. Operation Outside the Box, an example of hassling, was a blow to the Syrian nuclear program, but neither resolved the host of political disputes between Syria and Israel nor prevented future investments in a nuclear program. Operation Outside the Box is therefore different from the 2003 invasion of Iraq, a preventive war. The 2003 invasion destroyed—not blunted—Iraq's potential nuclear aspirations and eliminated the possibility of future peaceful negotiations with the Baathist government by overthrowing it. Many instances of limited strikes against nuclear facilities (Reiter, 2005; Fuhrmann and Kreps, 2010), hybrid conflict (Lanoszka, 2016), gray-zone conflict (Mazarr, 2015; Gannon *et al.*, 2020), and (limited) preventive war (Levy, 2011) could qualify as hassling. Hassling also operates similarly to how other works treat arming, sanc-

tions, and containment regimes (McCormack and Pascoe, 2017; Coe, 2018; Joseph, 2020). Of course, some states have a better ability to hassle than others.

Definition: Hassling capabilities *are the tools that an actor uses to hassle.*

I examine the equilibrium effects of improvements in hassling capabilities. While other work considers this topic, the novelty here is that there is a private component to a defender's hassling capabilities and, thus, its willingness to hassle. Here, the challenger is uncertain over how much it can transgress before it provokes hassling or war.

I model hassling capabilities as having both public and private components to mirror real life. The public component can be interpreted as what other states know for certain about the defender's hassling capabilities, which largely comes from military intelligence and observations of a defender's previous hassling operations. For example, in 2007, Syria knew that Israel had a robust hassling capability, as Syria had observed in Operation Opera (1981). The private component can be interpreted as what other states do not know for certain about the defender's hassling capabilities, which largely comes from uncertainty over the defender's willingness to hassle or the defender's secret technological capabilities. To continue the example, Syria likely did not know that Israel possessed a weapon that could disrupt its air defenses. Of course, that a state possesses private information about its ability to engage in conflict is not new, and the claim that states have private information about their hassling capabilities is analogous to claims that states have private information about their war capabilities (see Ramsay (2017) for a review). Following this logic, when I refer to "public improvements in hassling capabilities," I mean improvements to the defender's ability to hassle that can be observed by rivals. These improvements could occur through publicly announced upgrades in weapons capabilities, military training exercises for hassling-type operations, or even strategic leaking of classified data on capabilities.

In the model, I do not assume that improvements in hassling capabilities also affect war payoffs, though I discuss this modification in the Additional Results section. As some justi-

fication, there is an ongoing debate over the value of airstrikes in war (Pape, 1996; Horowitz and Reiter, 2001; Post, 2019), with some case analysis suggesting airstrikes appear more valuable in peacetime strikes against weapons facilities—i.e. in hassling—than in war (Kreps and Fuhrmann, 2011). Thus, the capabilities for effective hassling may look quite different than the capabilities for an effective war, an observation that is given further credence when noting how different the force postures of Operation Desert Fox (hassling) and the 2003 Iraq War (a protracted counterinsurgency) were. In fact, there may even exist a negative relationship between hassling capabilities and war capabilities, especially when militaries face finite budgets and must make strategic allocation decisions.

While it was a success, that Israel had to conduct Operation Outside the Box represents a kind of failure. For Israel, ideally, Syria would have never initially invested in the reactor.

Definition: *A defender experiences a* ***capability failure*** *when, following improvements in capabilities, the state experiences overall worse outcomes (i.e. lower utilities).*

This paper ties the concept of a "capability failure" to the defender attaining a lower utility. The concept is not tied to a specific set of outcomes like the concept of a "deterrence failure" typically is (see Schelling (1980, 3–10) or Huth (1999)) because the outcomes here are more multifaceted than is standard. Here the challenger can select from a continuum of transgressions, and the defender can accept, go to war, or select from a continuum of hassling options. While this definition obfuscates minor capability failures (small utility losses) from major capability failures (large utility losses), it still characterizes the theoretically interesting case where better capabilities produce worse utilities for the defender.

1.2 Theory

One state, a challenger, chooses how aggressively to transgress. Upon observing the transgression, a rival defender has three choices: it can accept transgression, go to war to decisively challenge the transgression, or hassle to degrade the transgression. Because the defender's

response depends on the defender's hassling capabilities which has a private component, the challenger does not know how the defender will respond to a selected level of transgression. Better public hassling capabilities can hurt the defender through two mechanisms.

First, improvements in hassling capabilities can make the defender more predictable. When private hassling capabilities (as opposed to public hassling capabilities) play a large role in determining the defender's overall hassling capabilities, the defender's response to a selected transgression is difficult to predict. This unpredictability can be valuable to the defender; because it is difficult to know how the defender will behave, the challenger may scale back the scope of its transgression to avoid war. This is akin to the defender posturing or bluffing. However, when the defender publicly becomes better at hassling, sometimes the challenger can better predict when their transgressions would be met with hassling and not war. For example, because the United States had an robust ability to conduct low-level covert operations and because a war would be costly, the Soviet Union likely knew that their invasion of Afghanistan in 1979 would not be met with war with the U.S., but could be met with hassling. When public improvements in hassling capabilities reduce the uncertainty about how the defender will respond, the defender cannot effectively posture, thus undermining the deterrent threat from war. Essentially, improvements in public hassling capabilities can diminish the defender's benefits from its private information.

Second, improvements in hassling technologies can embolden a challenger. When choosing the extent to transgress, the challenger faces a trade-off: a larger transgression is advantageous if the defender does not declare war, but a larger transgression increases the likelihood that the defender will declare war. The terms of this trade-off—whether a small increase in transgression extent will produce a small or large increase in the likelihood of war—are dictated by the defender's hassling capabilities, that have both public and private components. Under some trade-off terms—say an 8% increase in transgression level produces a 20% increase in the likelihood of war—the challenger is unwilling to increase its level of

transgression. However, under different trade-off terms that could occur when the defender is better at hassling—a 27% increase in transgression level produces a 20% increase in the likelihood of war—the challenger would be willing to increase its level of transgression.[2] If the challenger is emboldened in this way, being better at hassling can actually cause worse outcomes for the defender to an extent that offsets its gains from being better at hassling.

If the defender is publicly better at hassling, it can result in a greater likelihood of war, more pervasive and intensive hassling campaigns, or both; in other words, improvements in hassling capabilities can produce a capability failure. But improved public hassling capabilities do not always produce capability failures. It could be that a defender improves its hassling capabilities by developing a wide range of complementary hassling technologies that could be used separately or together. If this is the case, then improvements in hassling capabilities can improve the defender's outcomes by making the defender less predictable. Alternatively, if a defender becomes so effective at hassling that it could easily degrade a rival's transgression, then the challenger may choose not to pay the costs of transgressing. Of course, achieving this degree of hassling efficacy may be difficult, as rival countries may be willing to absorb huge costs if it means, for example, developing nuclear weapons. Also, interestingly, improvements in public hassling capabilities are distinct from improvements in private hassling or wartime capabilities, which cannot produce a capability failure (as discussed below).

1.3 Related Theory

This model is perhaps closest to Gurantz and Hirsch (2017), but differs in two key respects: it allows the defender to select from a continuum of low-level conflict options (i.e. hassling), and it assumes the challenger is uncertain over the defender's true hassling capabilities. I focus on these issues below, but this model is rooted in a modeling tradition that examines endogenous transgressions (often arming) and deterrence, where a challenger undertakes some action that is detrimental to a defender, and a defender determines how best to respond.

[2]These numbers follow from the parameters used in Figure 2.

There are many variants of this interaction that consider uncertainty over the challenger's first move (Meirowitz and Sartori, 2008; Baliga and Sjöström, 2008; Schultz, 2010; Debs and Monteiro, 2014; Bas and Coe, 2016; Spaniel, 2019; Meirowitz *et al.*, 2019), domestic political considerations (Fearon, 1994; Smith, 1998; Di Lonardo and Tyson, 2017), deterrence when facing multiple threats (McMahon and Slantchev, 2015; Paine, 2020), signaling (Fearon, 1997; Kydd and McManus, 2017), uncertainty over a challenger's type (Trager, 2010; Acharya *et al.*, 2015; Gurantz and Hirsch, 2017), a first-mover's advantage (Chassang and Miquel, 2010; Baliga and Sjöström, 2020), ambiguous attribution (Baliga *et al.*, 2020), and others.[3]

In nearly every model cited above, when a state is dissatisfied it declares war. In practice, decision-makers select from many possible policies (Lanoszka, 2016; Gartzke and Lindsay, 2017; Mehta and Whitlark, 2017; Gannon *et al.*, 2020). An emerging branch of formal research considers a policymaker who faces a range of policy options, including reneging on bargained agreements (Schultz, 2010), limited war (Powell, 2015) or efforts that degrade a rising power (McCormack and Pascoe, 2017; Coe, 2018; Schram, Forthcoming; Joseph, 2020). The existing literature has shown that low-level conflict can be an alternative to preventive war. This paper builds on this finding, but demonstrates that an enhanced ability to conduct low-level conflict can have negative strategic effects. In other words, while existing research shows an enhanced ability to conduct low-level conflict is useful within a crisis, this paper shows it can also shape an opponent's behavior to produce riskier crises at higher frequencies.

The relationship between uncertainty and conflict is a central topic in contemporary international relations (Morrow, 1989; Fearon, 1995; Wagner, 2000; Fey and Ramsay, 2011; Slantchev and Tarar, 2011), with Ramsay (2017) providing an excellent review. This paper is the first to treat private conflict capabilities or willingness to fight as affecting different types of conflict differently, with this paper focusing exclusively on uncertainty over a state's ability and willingness to use hassling. The analysis conducted here is similar to related

[3]See Huth (1999) for a review of early work on deterrence.

research that considers how changes in wartime capabilities (or the features of peacetime) affects outcomes like utilities and the likelihood of war (Chassang and Padró i Miquel, 2009; Benson *et al.*, 2016; Spaniel and Malone, 2019; Spaniel, 2020).

Other research has similarly shown that improved low-level conflict capabilities can lead to worse outcomes, but under different assumptions and through different mechanisms. On assumptions, Powell (1989) and Powell (2015) assume low-level conflict is a tool for imposing costs in a war of attrition or for generating an elevated risk of conflict (respectively). Bas and Coe (2016) and Baliga *et al.* (2020) assume transgressions are stochastically observed or that they are imperfectly attributable (respectively). A particularly important (and uncommon) assumption made here is that the challenger can select from a flexible set of possible transgressions ($t \in \mathbb{R}_{\geq 0}$) rather than a dichotomous "transgress or not" choice ($t \in \{0, 1\}$). Within my model, this modeling choice effectively rules out the mechanisms behind "burning bridges" (Schelling, 1966, 1980), or "tying hands" (Fearon, 1997) as drivers for improved hassling capabilities producing capability failures. I describe this intuition further in the Appendix and in Corollary 3, but improvements in hassling capabilities cannot produce a capability failure unless private information also plays a role, and then only sometimes. As a result, the theoretical mechanisms (emboldening and predictability) are quite different from the seminal logic of burning bridges and more similar to research on information and crisis bargaining (Fearon, 1995; Fey and Ramsay, 2011; Meirowitz *et al.*, 2019; Spaniel, 2020).

2 Model

Two states, C and D, are in a crisis over a divisible asset with a normalized value of 1. C wants to transgress (denoted $t \in \mathcal{T} \subseteq \mathbb{R}_{\geq 0}$)[4] to improve its future wartime capabilities. In response to the selected t, D will accept it, engage in hassling to degrade the transgression (denoted $h \in \mathcal{H} \subseteq \mathbb{R}_{\geq 0}$), or go to war to prevent the transgression from coming to fruition.

[4] I define $\mathbb{R}_{\geq 0} = \{a : a \geq 0\}$.

When D does not declare war, C's transgression level t and D's hassling level h affects C's future likelihood of winning in war, which affects future bargaining between C and D. Critically, D's costs from hassling consist of public parameter $\alpha \in \mathcal{A} \subseteq \mathbb{R}_{\geq 0}$ and private type $\theta \in \Theta \subseteq \mathbb{R}_{\geq 0}$. The rationale for this is described in detail in Section 1.1, but to offer one interpretation, α can be thought of as D's observed hassling capacity, and θ could be thought of as D's private willingness to engage in hassling. In this setup, C does not know how much it can transgress before D will hassle or go to war. The game is as follows.

1. I define $\{\underline{\theta}, \bar{\theta}\} = \Theta$ with $\underline{\theta} < \bar{\theta}$. Nature sets $\theta = \underline{\theta}$ with probability $Pr(\underline{\theta})$ and sets $\theta = \bar{\theta}$ with probability $1 - Pr(\underline{\theta})$. D knows nature's selection θ, but C does not.

2. State C selects transgression level $t \in \mathcal{T} = \mathbb{R}_{\geq 0}$.

3. State D can either go to war by setting $w_D = 1$ or not go to war by setting $w_D = 0$ and selecting some level of hassling $h \in \mathcal{H} = \mathbb{R}_{\geq 0}$ (with $h = 0$ implying that D "accepts"). When D does not go to war, the game moves to the next stage. Going to war terminates the game and produces wartime payoffs $U_A = P(0,0) - \kappa_C$ and $U_D = 1 - P(0,0) - \kappa_D$ for states C and D (respectively; I characterize the P function below).

4. State D offers state C some value $x \in [0,1]$.

5. State C can declare "war" by setting $w_C = 1$ or can "accept" the offer by setting $w_C = 0$. When state C sets $w_C = 1$, state C receives their its updated wartime payoff $U_A = P(t,h) - \kappa_C$, and state D receives $U_D = 1 - P(t,h) - \kappa_D - \frac{h^2}{F(\alpha,\theta)}$, which is its updated wartime payoff $(1 - P(t,h) - \kappa_D)$ minus its costs from hassling $(\frac{h^2}{F(\alpha,\theta)})$. When state C sets $w_C = 0$, state C receives payoff $U_A = x$ and state D receives $U_D = 1 - x - \frac{h^2}{F(\alpha,\theta)}$.

Stages 1-3 describe C's transgression and D's initial response. Stages 4-5 describe future crisis bargaining after the transgression and hassling have been realized. As I discuss in the Appendix, similar results can arise through many different bargaining protocols outside of what is modeled in stages 4 and 5, or through similar modeling choices made in Gurantz

and Hirsch (2017), which makes this bargaining a black-box. As intuition for the game form, in Operation Outside the Box, Syria's investment in the reactor would happen in stage 2, Israel's hassling would happen in stage 3, and future crisis bargaining between Syria and Israel after the operation occurs would happen in stages 4 and 5. Israel's decision to hassle in 3 was thus driven by their concern over how crisis bargaining with a nuclear-armed Syria would play out in 4 and 5. Note that the treatment of war as a game-ending move and hassling as a tool that allows for continued negotiations is consistent with the common treatment of war as a costly lottery.

Scenario	**C's utility**	**D's utility**
D initiates war at stage 3 (before h and t are realized)	$P(0,0) - \kappa_C$	$1 - P(0,0) - \kappa_D$
C initiates war at stage 5 (after h and t are realized)	$P(t,h) - \kappa_C$	$1 - P(t,h) - \kappa_D - \frac{h^2}{F(\alpha,\theta)}$
C accepts at stage 5 (after h and t are realized)	x	$1 - x - \frac{h^2}{F(\alpha,\theta)}$

Table 1: Summarized payoffs for actors.

The function $P : \mathcal{T} \times \mathcal{H} \rightarrow [0,1]$ is the likelihood that C wins in a war. I assume functional form $P(t,h) = max\left\{min\left\{1, \rho + t - h\right\}, \rho\right\}$ with $\rho \in [0,1]$, which implies P is weakly increasing in t and $-h$, and that P falls between ρ and 1 inclusive. The constant ρ is C's likelihood of winning a war before transgression or hassling comes to fruition (where $t = 0$ and $h = 0$). That $P(t,h)$ must be (weakly) greater than ρ implies that while hassling can degrade transgressions t, hassling can never make C a declining state, and, at most, hassling returns C to a baseline war victory likelihood of ρ.[5]

If D initiates war in stage 3, the actors fight over the asset. This stage closely resembles the Fearon (1995) costly lottery treatment of war. C's likelihood of winning in war is $P(0,0) = \rho$, and $\kappa_C > 0$ and $\kappa_D > 0$ are C's and D's costs from war, respectively.

[5]While low-level conflict could potentially turn a state into a declining power, this falls outside of the scope of the analysis.

If C initiates war in stage 5, the fighting happens after the transgression and hassling have both occurred. The P function, κ_C, and κ_D have been discussed above. The expression $\frac{h^2}{F(\alpha,\theta)}$ denotes the additional costs that D incurs from hassling, where the function $F : \{\mathcal{A} \times \Theta\} \rightarrow \{\mathbb{R}_{>0}\}$, with F is strictly increasing in α and θ. This functional form implies that D faces lower costs of hassling as α and θ increase, and that D pays no costs when $h = 0$ (i.e., from accepting C's transgression). For ease, I assume that C faces no costs to transgress; I relax this assumption in the Appendix, and the substantive results do not change.

If C accepts in stage 5, this represents a bargained outcome after the transgression and hassling have both occurred. The value x denotes the offer D makes to C.

2.1 Equilibrium Concepts and Assumptions

I limit attention to pure strategy perfect Bayesian Nash equilibria. As I discuss in the Appendix, considering mixed strategies do not change the substance of the results. The actions taken in the game depend on the public and private components of D's capabilities. A strategy for D is a mapping from the selected level of transgression t and its hassling capabilities α and θ to its action space w_D, h, and x. Because C does not know the value of θ, C's strategy is a mapping from the known parameter α to its action space t and w_C. I let σ denote a pair of strategies or $\sigma = (\sigma_A, \sigma_D)$. I let x^*, h^*, w_C^*, w_D^*, t^*, σ_A^*, and σ_D^* denote equilibrium actions and strategies, and $\sigma^*(\alpha, \theta)$ denote the equilibrium.

At this point, I can formally define a "capability failure."

***Definition**: Improvements in publicly observed hassling capabilities (i.e. moving from $\underline{\alpha}$ to $\bar{\alpha}$ with $\underline{\alpha} < \bar{\alpha}$) produce a **capability failure** when, $U_D\left(\sigma^*(\theta, \underline{\alpha})\right) \geq U_D\left(\sigma^*(\theta, \bar{\alpha})\right)$ for all $\theta \in \Theta$ and $U_D\left(\sigma^*(\theta, \underline{\alpha})\right) > U_D\left(\sigma^*(\theta, \bar{\alpha})\right)$ for some $\theta \in \Theta$.*

Improvements in public hassling capabilities produce a capability failure only when every possible type of D experiences weakly less expected utility. I discuss why this definition is

preferred to using D's ex-ante expected utility (a less-strict definition) in the Appendix.

As a final assumption, I will limit my analysis below to scenarios where the constraints on the P function and x do not bind. This implies that in equilibrium, $P \in (p, 1)$ and $x \in (0, 1)$. This assumption is useful because it eliminates the possibility that the kink in the P function or the requirement that $x \in [0, 1]$ drive any of the results. Furthermore, this allows the analysis to be straightforward, without needing to consider excessive casework. In the Appendix I include alternate functional forms and specifications that do not rely on this assumption and the findings are substantively identical.

3 Equilibrium

The intuition is as follows. In the fourth and fifth stages, D does strictly better avoiding war, which D accomplishes by making C an offer equal to its wartime utility $x^* = \rho + t^* - h^* - \kappa_C$. In the third stage, D reacts to C's selected transgression (t^*) by either going to war ($w_D^* = 1$) or not going to war ($w_D^* = 0$) and selecting an optimal hassling level (h^*). In the second stage, C selects a transgression t^* based on how C expects D will respond, which depends on D's known parameter (α) and expectations over D's private type (θ).

Because C faces no costs to transgress (see the Appendix for a model with costs), C's utility from its transgression level is increasing, unless it provokes D to war. Thus, C will attempt to avoid war with at least some types of D. C will select a t that would either make a $D(\alpha, \underline{\theta})$ or a $D(\alpha, \bar{\theta})$ indifferent between going to war in stage 3 while C is weaker, or hassling and letting bargaining play out with a stronger-but-hassled C in stages 4 and 5. I let $t(\alpha, \theta)$ denote the level of transgression that would make a $D(\alpha, \theta)$ indifferent between war (implicitly in stage 3) and hassling (implicitly with the bargained outcome). If C selects $t(\alpha, \underline{\theta})$, C avoids war altogether. If C selects $t(\alpha, \bar{\theta})$, $D(\alpha, \underline{\theta})$ (but not $D(\alpha, \bar{\theta})$) will respond with war.

The equilibrium takes one of two cases: a case where C avoids war, and a case where

C sometimes risks war. Proposition 1 lists some features of the equilibrium, excluding some of the straightforward strategic play described above. The Appendix contains the full equilibrium and all proofs.

Proposition 1: *Under the assumptions above, for a fixed* $\alpha \in \{\underline{\alpha}, \bar{\alpha}\}$*, the following actions are part of the Perfect Bayesian Nash Equilibrium. Letting* $Q(\alpha) = Pr(\underline{\theta})(\kappa_C + \kappa_D) + \left(Pr(\bar{\theta}) - Pr(\underline{\theta})\right)\frac{F(\alpha,\underline{\theta})}{4} - Pr(\bar{\theta})\frac{F(\alpha,\bar{\theta})}{4}$*:*

- ***Case 1.*** *When* $Q(\alpha) \geq 0$ *holds, C selects transgression level* $t^* = t(\alpha, \underline{\theta})$*, which results in both types of D hassling, setting* $h^* = \frac{F(\alpha,\theta)}{2}$ *for all* $\theta \in \{\underline{\theta}, \bar{\theta}\}$*.* $D(\alpha, \underline{\theta})$ *attains utility* $U_D(\sigma^*(\underline{\theta}, \alpha)) = 1 - \rho - \kappa_D$*, and* $D(\alpha, \bar{\theta})$ *attains* $U_D(\sigma^*(\bar{\theta}, \alpha)) = 1 - \rho - \kappa_D + \frac{F(\alpha,\bar{\theta})}{4} - \frac{F(\alpha,\underline{\theta})}{4}$*.*
- ***Case 2.*** *When* $Q(\alpha) < 0$ *holds, C selects transgression level* $t^* = t(\alpha, \bar{\theta})$*, which results in* $D(\alpha, \underline{\theta})$ *declaring war and* $D(\alpha, \bar{\theta})$ *hassling, setting* $h^* = \frac{F(\alpha,\bar{\theta})}{2}$*. Both* $D(\alpha, \underline{\theta})$ *and* $D(\alpha, \bar{\theta})$ *attain their wartime utility,* $U_D(\sigma^*(\theta, \alpha)) = 1 - \rho - \kappa_D$*.*

In Case 1, C avoids war altogether by selecting a transgression level that makes the type of D that is least willing to hassle, type $\underline{\theta}$, indifferent between hassling and war, giving $D(\alpha, \underline{\theta})$ its wartime utility $1 - \rho - \kappa_D$. When C selects $t(\alpha, \underline{\theta})$, the type of D that is more willing to hassle, type $\bar{\theta}$, will also hassle and will attain a utility above its wartime utility because hassling is cheaper for it. In Case 2, C sometimes risks war by selecting a transgression level that makes the type of D that is most willing to hassle, type $\bar{\theta}$, indifferent between hassling and war, giving $D(\alpha, \bar{\theta})$ its wartime utility $1 - \rho - \kappa_D$. When C selects $t(\alpha, \bar{\theta})$, the type of D that is less willing to hassle, type $\underline{\theta}$, will want to go to war and will attain its wartime utility. Note that Case 1 and Case 2 present different utilities for A; $Q(\alpha)$ is simply the cutpoint where, for values below the cutpoint, C does better by selecting the higher t^*—selecting $t(\alpha, \bar{\theta})$—and sometimes risking war relative to selecting $t(\alpha, \underline{\theta})$.

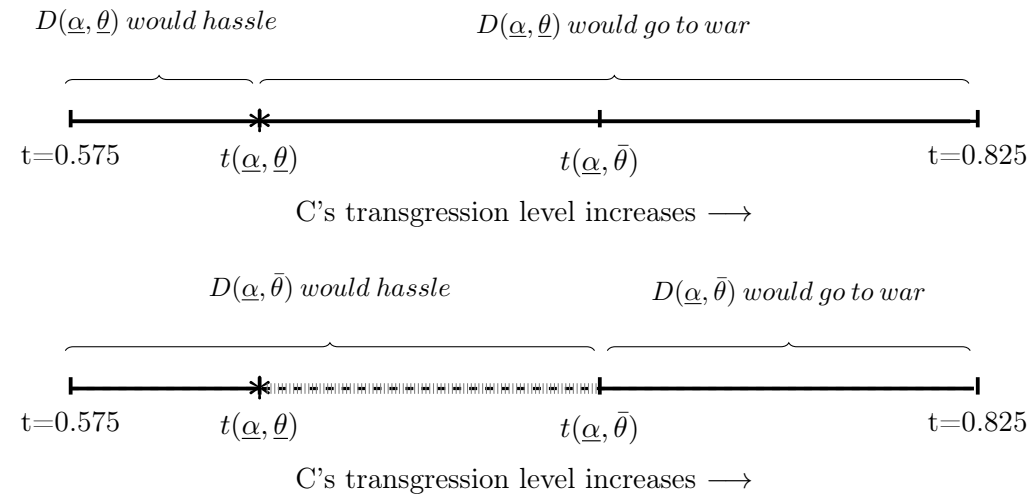
4 Results

In this section, I will first discuss how improvements in public hassling capabilities can lead to D becoming more predictable or C becoming emboldened, thus producing a capability failure. For clarity, I include more details on these results in the Appendix. I then discuss how these two mechanisms are the *only* way for a capability failure to occur in the model. I conclude this section by discussing what these results mean and extensions.

4.1 Predictability and Capability Failure

In the game, state D (weakly) benefits from its private information. An improvement in public hassling capabilities can result in different private types of D playing the game more similarly—for example, if it is common knowledge that D is outstanding at cyberwarfare, then C may be better able to predict how D will respond to a wide range of selected transgressions (e.g. with cyberattacks). As a result, when public capabilities are higher, state C can select a transgression that better extracts the benefit that D attains from its private information. This can occur despite the likelihood of war remaining unchanged. Figure 1 visualizes this intuition. For the selected parameters, under both $\underline{\alpha}$ and $\bar{\alpha}$, the equilibrium is described in Case 1 in Proposition 1, in which C always avoids war. Because C observes D's parameter α but not D's type θ, C selects $t(\alpha, \underline{\theta})$ to avoid war. In the top two panels of Figure 1 (where $\alpha = \underline{\alpha}$), C's optimal transgression level $t^* = t(\underline{\alpha}, \underline{\theta})$ is indicated by the asterisks. In the top panel, the selected t^* makes $D(\underline{\alpha}, \underline{\theta})$ indifferent between war and hassling, implying that $D(\underline{\alpha}, \underline{\theta})$ will attain its wartime utility. In the second panel, the space between the selected transgression and the transgression level that would have made a $D(\underline{\alpha}, \bar{\theta})$ indifferent between hassling and going to war (marked with a dashed line) represents D attaining some surplus. Instead of always receiving its wartime payoff, type $\bar{\theta}$ is better off *due to its private information.* In other words, so long as $D(\underline{\alpha}, \bar{\theta})$ is able to keep its type private in the lead-up to C's selection of t, $D(\underline{\alpha}, \bar{\theta})$ can attain some bargaining surplus.

For $\underline{\alpha}$



For $\bar{\alpha}$

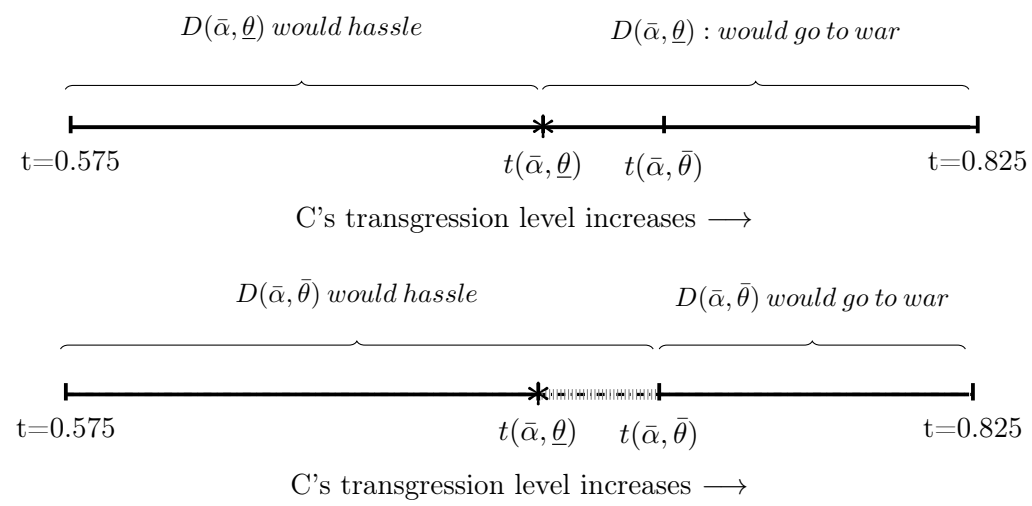


Figure 1: Optimal transgression and D's response (predictability).

C's selected level of transgression under parameters $\underline{\alpha}$ and $\bar{\alpha}$ are denoted by the asterisks. D's response to the selected t's are bracketed off. The dashed lines represent the surplus $D(\alpha,\bar{\theta})$ attains from its private information in equilibrium. Parameters are $\kappa_D = 0.1$, $\kappa_C = 0.4$, $\rho = 0.5$, $Pr(\underline{\theta}) = 0.5$, $Pr(\bar{\theta} = 0.5)$, $F(\underline{\alpha},\underline{\theta}) = 0.5$, $F(\underline{\alpha},\bar{\theta}) = 1$, $F(\bar{\alpha},\underline{\theta}) = 0.8$, $F(\bar{\alpha},\bar{\theta}) = 1.1$.

However, when D's public hassling capabilities improve, D's private information can become less valuable. Because C has some uncertainty over how willing D is to hassle, C must benchmark its transgression against type $\underline{\theta}$ if it seeks to avoid war. If improvements in D's known hassling abilities reduce the importance of D's unknown type, then C does not need to scale back as much to prevent war. Comparing the top two and bottom two panels illustrates how this may occur. In the top two panels, when α is low, there is a significant gap between C's selected transgression $t(\underline{\alpha}, \underline{\theta})$ and the point that makes $D(\underline{\alpha}, \bar{\theta})$ indifferent between war and hassling ($t(\underline{\alpha}, \bar{\theta})$). Moving to $\bar{\alpha}$, there is a much smaller gap between $t(\bar{\alpha}, \underline{\theta})$ and $t(\bar{\alpha}, \bar{\theta})$. This implies that as D becomes publicly better at hassling (moving from $\underline{\alpha}$ to $\bar{\alpha}$), C is able to select a t^* that is closer to the point that gives $D(\bar{\alpha}, \bar{\theta})$ its wartime payoff. Put another way, in the bottom two panels it does not matter much if D's private type is revealed because $D(\bar{\alpha}, \bar{\theta})$ does not attain much surplus from its private information; D's public improvements in hassling has made D's private type a smaller factor in determining D's selected hassling levels and has therefore made D more predictable. This case defines a capability failure for D because an increase in hassling capabilities means C can select better tailored transgressions. Proposition 2 defines the necessary conditions under which an improvement in public hassling capabilities makes D more predictable, producing a capability failure.

Proposition 2 (Predictability): *Under the predictability conditions, C avoids war across parameters $\underline{\alpha}$ and $\bar{\alpha}$; formally, $Q(\underline{\alpha}) \geq 0$ (Condition 1) and $Q(\bar{\alpha}) \geq 0$ (Condition 2). Additionally, D's private information plays a diminished role under parameter $\bar{\alpha}$ relative to parameter $\underline{\alpha}$; formally, $(F(\underline{\alpha}, \bar{\theta}) - F(\underline{\alpha}, \underline{\theta}) > F(\bar{\alpha}, \bar{\theta}) - F(\bar{\alpha}, \underline{\theta})$ (Condition 3). When these predictability conditions hold, then improvements from $\underline{\alpha}$ to $\bar{\alpha}$ produce a capability failure.*

Proposition 2 generalizes the intuition in the figure above. The first two predictability conditions imply that across $\underline{\alpha}$ and $\bar{\alpha}$, the equilibrium is characterized in Case 1 of Proposition 1. The third condition, which is decreasing differences in F (Ashworth and Bueno de Mesquita, 2006), captures the differences in how types $\underline{\theta}$ and $\bar{\theta}$ play the game across parameters $\underline{\alpha}$

and $\bar{\alpha}$. Intuitively, across α's, C selects $t(\alpha, \underline{\theta})$ to make $D(\alpha, \underline{\theta})$ indifferent between war and hassling. Because $D(\alpha, \bar{\theta})$ faces lower costs from hassling, $D(\alpha, \bar{\theta})$ selects a greater level of hassling, captured in the optimal hassling level $h^*(\theta) = \frac{F(\alpha,\theta)}{2}$ and attains a greater utility, captured in the $\frac{F(\alpha,\bar{\theta})}{4} - \frac{F(\alpha,\underline{\theta})}{4}$ term in $D(\alpha, \bar{\theta})$'s utility function. When $F(\alpha, \bar{\theta}) - F(\alpha, \underline{\theta})$ is small, $D(\alpha, \underline{\theta})$ and $D(\alpha, \bar{\theta})$ play the game similarly, implying that D's private information is not particularly valuable to D. When the three conditions hold, the improvement in α degrades the value of D's private information, thus producing the capability failure.

4.2 Emboldening and Capability Failure

In the game, C wants to transgress as much as possible but may be deterred from transgressing if further transgressions lead to a greater probability of war. Whether C is willing to transgress to a level that risks war is contingent upon the benefit that C receives from the transgression, which is dependent upon D's willingness to hassle. If C is emboldened, D's shift in capabilities incentivizes C to transgress more aggressively, knowing that, while with some probability D may initiate war, C holds a better bargaining position in the case where D does not initiate war. An improvement in public hassling capabilities emboldens C when the benefit of transgressing more aggressively now outweighs the greater risk of war.[6] Figure 2 visualizes this intuition. C's expected utility is increasing in the selected t (moving right along the x-axis), until it provokes $D(\alpha, \underline{\theta})$ to go to war at $t(\alpha, \underline{\theta})$ (producing the first discontinuity), and is increasing in t again until all types to go to war at $t(\alpha, \bar{\theta})$.

The top panel of Figure 2 describes the game under parameter $\underline{\alpha}$, where the gameplay is described in Case 1 in Proposition 1. In this game, C optimally selects $t^* = t(\underline{\alpha}, \underline{\theta})$, and both $D(\underline{\alpha}, \underline{\theta})$ and $D(\underline{\alpha}, \bar{\theta})$ will always hassle. It is worth highlighting what C does not do when $\alpha = \underline{\alpha}$: C does not select $t(\underline{\alpha}, \bar{\theta})$. This "move not taken" captures C's trade-off between a greater transgression and a higher risk of war. The upside to selecting $t(\underline{\alpha}, \bar{\theta})$, is that C

[6] Spaniel and Malone (2019) find a similar mechanism can arise following improved economic interdependence.

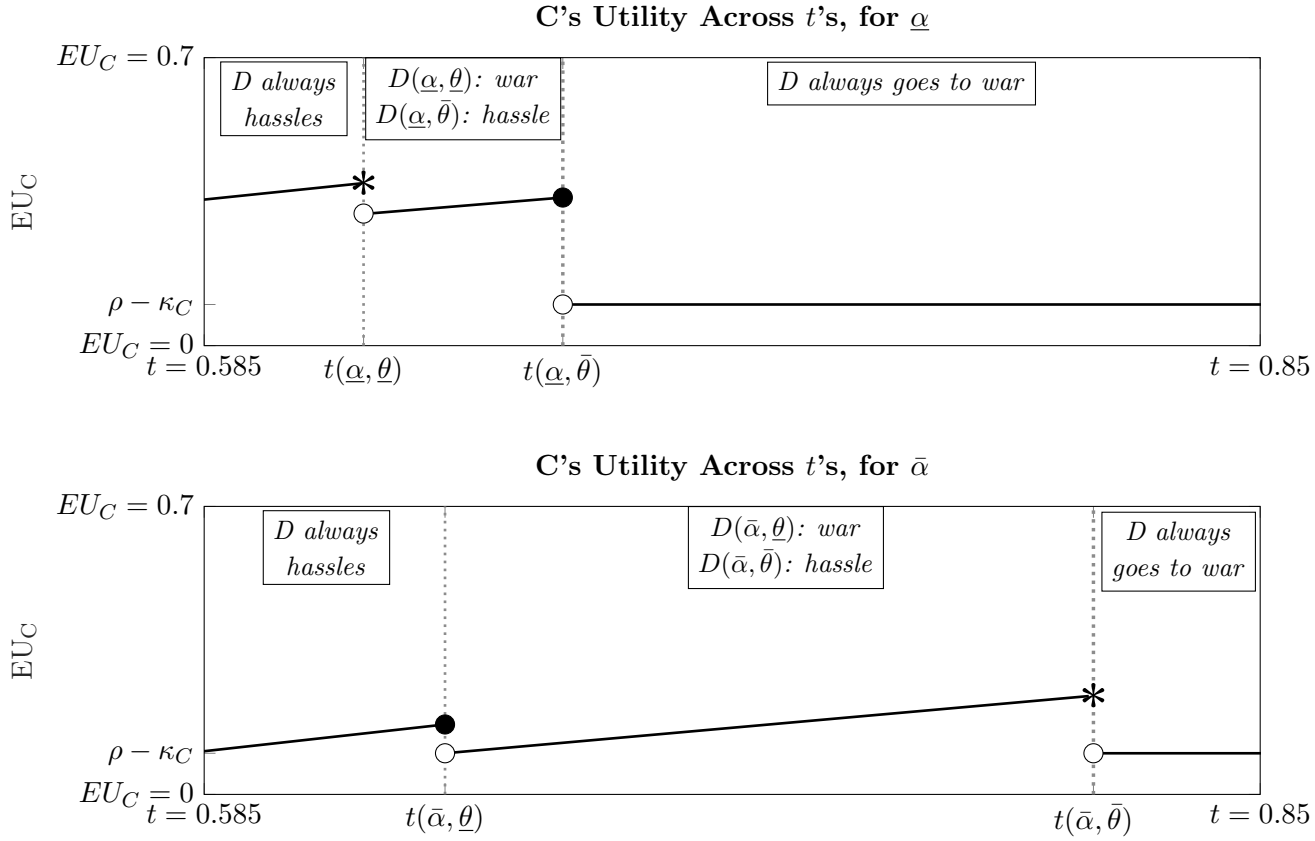


Figure 2: Optimal transgression and C's utility (emboldening).

C's selected level of transgression under parameters $\underline{\alpha}$ and $\bar{\alpha}$ are denoted by the asterisks. D's response to the selected t's are labeled. The parameters are $\kappa_D = 0.1$, $\kappa_C = 0.4$, $\rho = 0.5$, $Pr(\underline{\theta}) = 0.2$, $Pr(\bar{\theta}) = 0.8$, $F(\underline{\alpha}, \underline{\theta}) = 0.5$, $F(\underline{\alpha}, \bar{\theta}) = 0.7$, $F(\bar{\alpha}, \underline{\theta}) = 0.6$, $F(\bar{\alpha}, \bar{\theta}) = 1.3$.

would have an advantage when facing $D(\underline{\alpha}, \bar{\theta})$.[7] However, the downside for C of selecting a $t(\underline{\alpha}, \bar{\theta})$ is that now $D(\underline{\alpha}, \underline{\theta})$ will go to war, increasing the likelihood of war from zero to $Pr(\underline{\theta})$. The difference between $t(\underline{\alpha}, \underline{\theta})$ and $t(\underline{\alpha}, \bar{\theta})$ can be thought of as the potential upside if C were willing to go to war with likelihood $Pr(\underline{\theta})$. When $\alpha = \underline{\alpha}$, C's upside to doing better against $D(\underline{\alpha}, \bar{\theta})$ does not outweigh the costs of going to war with $D(\underline{\alpha}, \underline{\theta})$. In the figure, this is shown by the asterisk at $t(\underline{\alpha}, \underline{\theta})$ giving C greater utility than the point at $t(\underline{\alpha}, \bar{\theta})$.

When $\alpha = \bar{\alpha}$, C faces a new trade-off, and the gameplay is described in Case 2 in Proposition 1. Selecting $t(\bar{\alpha}, \bar{\theta})$ will provoke $D(\bar{\alpha}, \underline{\theta})$ to go to war. However, as the bottom panel illus-

[7]For the selected parameters, when nature sets $\theta = \bar{\theta}$, if C selects $t = t(\underline{\alpha}, \underline{\theta}) = 0.625$, C attains utility 0.375, and if C selects $t = t(\underline{\alpha}, \bar{\theta}) = 0.675$, C attains utility 0.425.

trates, C now faces a greater upside to selecting $t(\bar{\alpha}, \bar{\theta})$ over $t(\bar{\alpha}, \underline{\theta})$ because these cutpoints are so different, representing a large opening for a greater transgression if C is willing to risk some likelihood of war.[8] Now, C's upside to doing better against $D(\bar{\alpha}, \bar{\theta})$ outweighs the increased likelihood of war, as shown by the asterisk at $t(\bar{\alpha}, \bar{\theta})$ giving C greater utility than $t(\bar{\alpha}, \underline{\theta})$. This case defines a capability failure for D because an increase in hassling capabilities emboldened C to pursue a riskier transgression strategy and sometimes provoke war, both to D's detriment. Proposition 3 defines the necessary conditions for improvements in public hassling capabilities to embolden C and produce a capability failure.

Proposition 3 (Emboldening): *Under the emboldening conditions, C avoids war under parameter $\underline{\alpha}$ and goes to war under parameter $\bar{\alpha}$; formally, $Q(\underline{\alpha}) \geq 0$ (Condition 1) and $Q(\bar{\alpha}) < 0$ (Condition 2). When these emboldening conditions hold, then improvements from $\underline{\alpha}$ to $\bar{\alpha}$ produce a capability failure.*

Proposition 3 generalizes the intuition in the example above. The $Q(\underline{\alpha}) \geq 0$ condition implies that C will not risk war under parameter $\underline{\alpha}$, which will give $D(\underline{\alpha}, \underline{\theta})$ its wartime utility and $D(\underline{\alpha}, \bar{\theta})$ a utility above its wartime payoff. The $Q(\bar{\alpha}) < 0$ condition implies that C will risk war under parameter $\bar{\alpha}$, which will result in both $D(\bar{\alpha}, \underline{\theta})$ and $D(\bar{\alpha}, \bar{\theta})$ attaining their wartime utilities. The capability failure arises because $D(\bar{\alpha}, \bar{\theta})$ does strictly worse than $D(\underline{\alpha}, \bar{\theta})$.

4.3 When (and Only When) Improved Capabilities Produce Capability Failures

The aim of this paper is to define the relationship between capabilities and outcomes when there is uncertainty over the willingness to use low-level conflict. I showed that an improvement in α when the predictability or emboldening conditions hold represent two ways capability failures can occur. But are there other ways for capability failures to occur in the model? Proposition 4 and the corollaries that follow show that there is not.

[8] For the selected parameters, when nature sets $\theta = \bar{\theta}$, if C selects $t = t(\bar{\alpha}, \underline{\theta}) = 0.65$, C attains utility 0.1, and if C selects $t = t(\bar{\alpha}, \bar{\theta}) = 0.825$, C attains utility 0.275.

Proposition 4: *Improvements in α produce a capability failure if and only if the emboldening or predictability conditions hold.*

Proposition 4 shows that the two mechanisms characterized above represent all the ways that improvements in public hassling capabilities can produce a capability failure. But can similar capability failures arise by considering private hassling capabilities or war capabilities?

Corollary 1: *Improvements in private hassling capabilities θ (e.g. moving from $\underline{\theta}$ to $\bar{\theta}$) cannot produce a capability failure.*

Corollary 2: *Improvements in war capabilities (e.g. decreasing ρ or κ_D) cannot produce a capability failure.*

Both Corollaries follow Proposition 1 and suggest that public hassling capabilities are unique in their ability to cause capability failures. And, analysis in the Appendix demonstrates that these results are not simply theoretical anecdotes driven by a game form, but rather can apply to a broad class of games and bargaining protocols with discrete or continuous type spaces. On Corollary 1, the logic follows from standard mechanism design results where higher types attain an increasing information rent (Myerson, 1979; Banks, 1990; Salanié, 2005; Fey and Ramsay, 2011). On Corollary 2, when some types attain their wartime utility, improvements in wartime payoffs always produce better outcomes. Of course, it would be incorrect to say that public hassling capabilities are exclusively responsible for capability failures.

Corollary 3: *Improvements in public hassling capabilities cannot produce a capability failure without private information (e.g. when $\underline{\theta} = \bar{\theta}$).*

Corollary 3 follows from Propositions 1 and 4. When when $\underline{\theta} = \bar{\theta}$, C optimally selects a transgression level that makes D indifferent between war and hassling, thus granting D their wartime utility. While the specific predictability and emboldening conditions are products of modeling choices,[9] in every iteration that is considered, interactions between public and

[9]For example, I show in the Appendix when C's transgressions affects D's hassling costs, the predictability

private capabilities following improvements in public capabilities are needed for capability failures to occur.

4.4 Additional Results

What if improved hassling capabilities also affect wartime capabilities? Following the logic from Corollary 2, if improvements in publicly observed hassling capabilities also improve payoffs from war, then the improvements in public hassling capabilities do not produce a capability failure. Because type $\underline{\theta}$ D's always attain their wartime payoff, if a shift from $\underline{\alpha}$ to $\bar{\alpha}$ resulted in D attaining a greater wartime payoff, then $D(\bar{\alpha}, \underline{\theta})$ would always do better than $D(\underline{\alpha}, \underline{\theta})$. As discussed in Section 1.1, whether or not improved hassling capabilities affects war capabilities is an empirical question, grounded in how different war operations and hassling operations are. At a minimum, this observation suggests a useful heuristic for developing hassling capabilities: so long as improvements in hassling also improve wartime outcomes, then these improvements will not produce a capability failure.

What is the effect of hassling capabilities on the likelihood of war? While not the main focus of the paper, the relationship between wartime capabilities and the likelihood of war has received considerable attention (Wittman, 1979; Powell, 1999; Kydd, 2000; Benson *et al.*, 2016; Spaniel, 2020). Similar to recent findings on war capabilities (see Benson *et al.* (2016)), I show that improvements in public hassling capabilities could lead to more, less, or no change in the final likelihood of war.[10] Thus, the relationship between low-level capabilities and the outbreak of war is nuanced, suggesting that, for example, third-party assistance to a protege's hassling capabilities could be stabilizing or destabilizing.

What if transgressions come with costs? I examine a model with transgression costs in the Appendix, showing that while the results on capability failures remain substantively

mechanism can be satisfied under increasing differences.

[10]Following the improvement in α, when $Q(\underline{\alpha}) \geq 0$ and $Q(\bar{\alpha}) < 0$ there is more war; when $Q(\underline{\alpha}) < 0$ and $Q(\bar{\alpha}) \geq 0$ there is less war; and when $Q(\underline{\alpha}) \geq 0$ and $Q(\bar{\alpha}) \geq 0$ or $Q(\underline{\alpha}) < 0$ and $Q(\bar{\alpha}) < 0$ there is no change in the likelihood of war.

similar, this alternate model imposes new, stricter predictability and emboldening conditions. These new conditions are needed because improving public hassling capabilities can produce a new effect: it can deter a would-be challenger from transgressing in the first place, as they would know that their costly transgressions could be largely degraded. Whenever improving public hassling capabilities produces this effect, then there cannot be a capability failure.

5 Emboldening: Saddam's Gamble

Within the ceasefire terms of the 1991 Persian Gulf War and following United Nations Security Council Resolutions, Iraq was forbidden from possessing nuclear, biological, or chemical weapons (WMDs hereafter) and was subject to inspections from a United Nations Special Commission and the International Atomic Energy Agency. Between 1991-1998, Iraq repeatedly violated the terms of the resolutions by engaging in the following: preventing inspectors from searching suspected vehicles (1991); preventing inspectors from searching government buildings (1992); submitting incomplete disclosures of weapons programs (1995-1996); and eventually terminating all cooperation with inspectors (1998). These actions (and others) led U.S. and British forces to conduct Operation Desert Fox, a bombing raid against Iraqi military targets in 1998. Between 1999-2002, Iraq continued keeping weapons inspectors out, despite intensifying rhetoric within the United States. And, while in late 2002 Iraq agreed to resume inspections, after over a decade of friction with inspectors, Iraq's change proved insufficient, and the United States invaded Iraq in March 2003. In 2004 the Iraq Survey Group released the "Duelfer Report," which described how Iraq's WMD programs had essentially been destroyed in 1991 (Duelfer, 2004).

To claim that Iraq was emboldened in the lead up to the 2003 U.S. invasion, three questions must be addressed: Was Iraq "transgressing" as defined above? Did the U.S. demonstrate a robust hassling capability? Finally, did U.S. hassling capabilities shape Iraqi behavior?

Saddam's transgression was keeping the weapons inspectors out. As Koblentz (2018) and

Coe and Vaynman (2020) discuss, states may be weakened by allowing external observers to inspect their military capabilities because this information could leak and be used by adversaries. In breaking the terms of the U.N. resolutions, Saddam put Iraq in a better military position. Moreover, keeping inspectors out led the United States to grow concerned that Iraq was developing WMDs and was therefore a rising state. While the United States and Saddam interpreted the decision to keep weapons inspectors out in different ways, Saddam ultimately knew that his actions could provoke a military response and deliberated over how much to reveal about the state's security forces to weapons inspectors (see Woods *et al.* (2006, 15-16, 30, 91, 96-97, 125) and Woods *et al.* (2011, 257-258)).

I do not want to gloss over the fact that Saddam was not actually building WMD, and that if the U.S. had all the correct information, then the U.S. plausibly would not have invaded Iraq. The reason the U.S. did not know the extent of Iraq's WMD programs was because Saddam kept weapons inspectors out. As Coe and Vaynman (2020) suggests, "[i]n principle, Iraq could be blanketed with inspectors," producing transparency in Iraq's weapons development, if not for Saddam's handling of the inspectors. Saddam's decisions created uncertainty around Iraq's WMD programs that the U.S. grappled with in the lead up to the 2003 invasion, setting the stage for the strategic tensions explored in research like Baliga and Sjöström (2008), Debs and Monteiro (2014), Bas and Coe (2016), Spaniel (2019), and others to operate. Ultimately, while Saddam stood to gain by turning away weapons inspectors, he did not know the U.S.'s true willingness to accept uncertainty over Iraq's WMD programs and was too bold in his handling of weapons inspectors.

The second question to answer is whether the U.S. had publicly demonstrated a robust hassling capability. This was indeed the case. After the Cold War, the U.S. transformed much of its technical know-how in building and deploying missiles and bombers for nuclear strikes into precision strike capabilities (U.S. Air Force, 2019). Additionally, the efficacy of these capabilities and U.S. public willingness to use them was revealed in a series of conflicts,

including, but not limited to, Operation Desert Fox.

Third, there is ample evidence that Saddam's decision to refuse entry to the inspectors was influenced by U.S. hassling capabilities. In internal discussions before the invasion, Saddam commented frequently on the U.S. reliance on air strikes—not ground invasion—a sentiment echoed by his advisers (Woods *et al.*, 2006, 15-16, 30, 96-97, 125). At one point, an adviser was recorded in a conversation with Saddam claiming, "I believe if any incident occurs, the Americans will utilize their air strike methods, which they prefer and used recently, instead of sending troops, based on their horrific experience in Somalia" (Woods *et al.*, 2006, 30). Similarly, in interviews after Saddam was captured, he further stressed his belief that the U.S. would respond in 2003 as it had in Operation Desert Fox (Battle, 2009).

Overall, Saddam knowingly chose to refuse entry to weapons inspectors to bolster his future military strength. In doing so, Saddam recognized that he risked confrontation with the United States. But, because the U.S. was effective at hassling, Saddam believed that the benefits of keeping weapons inspectors out outweighed the expected response of hassling and the small risk of war. However, Saddam did not know the U.S.'s true willingness to accept his treatment of weapons inspectors, and he ultimately behaved too aggressively. While there is no way to know how Saddam would have behaved if the U.S. possessed a less robust hassling ability, in this counterfactual setting, it is plausible that Saddam would have taken the threat of war more seriously and would have been more open to weapons inspections.

That Saddam may have changed his behavior in response to improved U.S. hassling capabilities (i.e., improvements that facilitate low-level conflict) in such a way that ultimately led to a war (i.e., an escalated level of conflict) has not previously been formalized or explained. While using existing theoretical frameworks—like treating improved hassling capabilities as a "failure to burn bridges" (see Appendix) or applying models like Bas and Coe (2016) and Joseph (2020)—can show that improved hassling capabilities lead to more hassling, existing research cannot explain how being better at hassling led to an escalation to war in this

setting. Thus, this case presents a new perspective on the 1990s "revolution in military affairs," which viewed developments in precision strike and cyber capabilities as fundamentally altering the way conflict would be fought (Cordesman, 1999). This paper does not dispute that states have successfully used these technologies for hassling, plausibly in lieu of more costly and destructive wars. However, here Saddam observed the U.S.'s increased reliance on hassling technologies and was emboldened to turn away weapons inspectors, believing that his decision would likely be met with a limited response. Ultimately, this case embodies the theoretical finding that improvements in hassling capabilities can shift the strategic decisions that revisionist states make and may ultimately result in devastating and destabilizing wars.

6 Predictability and the Stability-Instability Paradox

During the Cold War, while mutually assured destruction prevented the United States and Soviet Union from engaging in nuclear war, the superpowers did compete in multiple conflict theaters at low-levels, for example, by providing sympathetic rebels or governments with support and fighting proxy wars. These observations led scholars to posit a "stability-instability paradox," where stability at the nuclear level opened the possibility for competition (instability) at lower-levels (Snyder, 1965; Jervis, 1984). While there is some empirical evidence of its existence (Rauchhaus, 2009; Early and Asal, 2018), the scope conditions for the paradox outside of nuclear-level stability is still an open topic (O'Neill, 2019). Notably, even Snyder (1965, 189–190) hedges on the claim that nuclear stability *causes* low-level instability, emphasizing that the opposite could be true if fear of linkages or escalation between levels of conflict are too salient. This paper can offer two points on these theoretical underpinnings.

First and most simply, for low-level instability to occur, states must be able and willing to engage at those low-levels. In terms of the model, consider the U.S. as the defender and the U.S.S.R. as the challenger.[11] At the onset of the Cold War, funding for the newly formed

[11] The same logic exists if these labels are flipped.

Central Intelligence Agency grew rapidly, allowing the organization to participate in a wide portfolio of covert operations beyond intelligence gathering (Gaddis, 2005; Snider, 2015).[12] In other words, the U.S. expanded its hassling capabilities, and the U.S. engaged in hassling operations throughout the globe. But of course, as the model with transgression costs (see Appendix) suggests, the expanded hassling capability could have deterred transgressions. After all, if the challenger knows only well-funded and aggressive transgressions can survive a defender's robust hassling ability, then a challenger may decide transgressing is not worth the costs. Instead, the U.S.S.R. was was resolved enough to transgress and expanded its own low-level conflict abilities, even while facing a rival that had invested great capital and efforts into its hassling capabilities. Essentially, the expansion in low-level conflict capabilities on both sides in the early Cold War opened the door to greater low-level instability.

Second, the predictability mechanism frames the topics of linkages and escalation within a utility analysis. The existence of low-level instability does not imply that the defender is experiencing low utility. Following improvements in hassling capabilities, the defender could be more successfully degrading transgressions at low cost while avoiding war, all of which are good outcomes. Rather, the defender experiences worse utilities when the defender becomes better at engaging in hassling, and this results in the linkages between the challenger's actions and the defender's responses becoming clearer and more predictable; when this happens, the challenger's fear of a possible escalation is undermined. If the U.S. expansion in hassling abilities led to the U.S. becoming more predictable, then the U.S.S.R. could more effectively "design around" the defender's deterrent threat from war and undertake aggressive, calibrated transactions without fear of escalation (George *et al.*, 1974). While a complete analysis of the Cold War is beyond the scope of this paper, in some cases, there appeared to be such calibrated actions. For example, the U.S.S.R. invasion of Afghanistan led to international condemnation and hassling, but there seemed little risk that the U.S.

[12]Even under Eisenhower's "asymmetric response" doctrine, the C.I.A. expanded rapidly, and a wide range of covert operations occurred (Gaddis, 2005, 125–197).

would commit conventional forces or engage in brinkmanship over the issue (Coll, 2005). If the Soviet treatment of Afghanistan was informed by decades of U.S. hassling operations, then it is possible the U.S. had become predictable and was experiencing a capability failure.

7 Conclusion

In this paper, I present a game theoretic model where a rising power chooses how much to transgress, knowing that it faces a defending power who can respond at multiple conflict escalation levels and who has private information about its willingness to engage at different levels. How do the defender's capabilities affect its outcomes? When the defender improves their ability to engage in low-level conflict, two mechanisms can cause worse outcomes for the defender: the rival state may be emboldened or the defender may become predictable. In short, these results arise when improved low-level conflict capabilities undermine the defender's ability to effectively use its private information.

The results here suggest that political scientists and policymakers need to take a harder look at low-level conflict capabilities. The logic of having "many tools in the policy toolbox" can be counterproductive because of the strategic responses these tools can produce in rivals. While this paper does not suggest that having many policy options is always bad, under some conditions, more tools can lead to systemically worse outcomes. More research is needed on this topic, especially work that considers uncertainty over willingness to engage in multiple kinds of conflict. Future work should take a harder, systematic look at the relationship between war and hassling capabilities, especially when hidden types across conflict options are correlated. Future work should also consider multiple dimensions of low-level responses.

References

Acharya, A., Grillo, E. *et al.* (2015) War with crazy types, *Political Science Research and Methods*, **3**, 281–307. 1.3

Ashworth, S. and Bueno de Mesquita, E. (2006) Monotone comparative statics for models of politics, *American Journal of Political Science*, **50**, 214–231. 4.1

Baliga, S., De Mesquita, E. B. and Wolitzky, A. (2020) Deterrence with imperfect attribution, *American Political Science Review*, **114**, 1155–1178. (document), 1.3

Baliga, S. and Sjöström, T. (2008) Strategic ambiguity and arms proliferation, *Journal of political Economy*, **116**, 1023–1057. 1.3, 5

Baliga, S. and Sjöström, T. (2020) The strategy and technology of conflict, *Journal of Political Economy*, **128**, 3186–3219. 1.3

Banks, J. S. (1990) Equilibrium behavior in crisis bargaining games, *American Journal of Political Science*, pp. 599–614. 4.3

Bas, M. A. and Coe, A. J. (2016) A dynamic theory of nuclear proliferation and preventive war, *International Organization*, **70**, 655–685. (document), 1.1, 1.3, 5

Battle, J. (2009) *Saddam Hussein Talks to the FBI*, National Security Archive. 5

Benson, B. V., Meirowitz, A. and Ramsay, K. W. (2016) Changing capabilities, uncertainty, and the risk of war in crisis bargaining, *Research & Politics*, **3**. 1.3, 4.4

Blainey, G. (1988) *Causes of war*, Simon and Schuster. (document)

Buchanan, B. (2020) *The Hacker and the State: Cyber Attacks and the New Normal of Geopolitics*, Harvard University Press. 1.1

Chassang, S. and Miquel, G. P. I. (2010) Conflict and deterrence under strategic risk, *The Quarterly Journal of Economics*, **125**, 1821–1858. 1.3

Chassang, S. and Padró i Miquel, G. (2009) Defensive weapons and defensive alliances, *American Economic Review*, **99**, 282–86. 1.3

Coe, A. J. (2018) Containing rogues: A theory of asymmetric arming, *The Journal of Politics*, **80**, 000–000. (document), 1.1, 1.3

Coe, A. J. and Vaynman, J. (2020) Why arms control is so rare, *American Political Science Review*, pp. 1–14. (document), 5

Coll, S. (2005) *Ghost wars: The secret history of the CIA, Afghanistan, and Bin Laden, from the Soviet invasion to September 10, 2001*, Penguin. 6

Copeland, D. C. (2001) *The origins of major war*, Cornell University Press. 1.1

Cordesman, A. H. (1999) The lessons of desert fox: A preliminary analysis. 5

Debs, A. and Monteiro, N. P. (2014) Known unknowns: power shifts, uncertainty, and war, *International Organization*, **68**, 1–31. (document), 1.1, 1.3, 5

Di Lonardo, L. and Tyson, S. (2017) Political instability and the failure of deterrence, *Available at SSRN 3094929*. 1.3

Duelfer, C. (2004) *Comprehensive Report of the Special Advisor to the DCI on Iraq's WMD*, vol. 1, Central Intelligence Agency. 5

Early, B. R. and Asal, V. (2018) Nuclear weapons, existential threats, and the stability–instability paradox, *The Nonproliferation Review*, **25**, 223–247. 6

Fearon, J. D. (1994) Domestic political audiences and the escalation of international disputes., *American Political Science Review*, **88**, 577–592. 1.3

Fearon, J. D. (1995) Rationalist explanations for war, *International organization*, **49**, 379–414. (document), 1.1, 1.3, 2

Fearon, J. D. (1996) Bargaining over objects that influence future bargaining power. 1.1

Fearon, J. D. (1997) Signaling foreign policy interests tying hands versus sinking costs, *Journal of Conflict Resolution*, **41**, 68–90. 1.3

Fey, M. and Ramsay, K. W. (2011) Uncertainty and incentives in crisis bargaining: Game-free analysis of international conflict, *American Journal of Political Science*, **55**, 149–169. (document), 1.3, 4.3

Fuhrmann, M. and Kreps, S. E. (2010) Targeting nuclear programs in war and peace: A quantitative empirical analysis, 1941-2000, *Journal of Conflict Resolution*, **54**, 831–859. 1.1

Gaddis, J. L. (2005) *Strategies of Containment*, Oxford University Press. 6, 12

Gannon, A. J., Gartzke, E., Lindsay, J. R. and Schram, P. (2020) The shadow of deterrence, *Manuscript*. 1.1, 1.3

Gartzke, E. and Lindsay, J. R. (2017) Thermonuclear cyberwar, *Journal of cybersecurity*, **3**, 37–48. 1.3

George, A. L., Smoke, R. *et al.* (1974) *Deterrence in American foreign policy: Theory and practice*, Columbia University Press. 6

Gurantz, R. and Hirsch, A. V. (2017) Fear, appeasement, and the effectiveness of deterrence, *The Journal of Politics*, **79**, 1041–1056. (document), 1.1, 1.3, 2

Horowitz, M. and Reiter, D. (2001) When does aerial bombing work? quantitative empirical tests, 1917-1999, *Journal of Conflict Resolution*, **45**, 147–173. 1.1

Huth, P. K. (1999) Deterrence and international conflict: Empirical findings and theoretical debates, *Annual Review of Political Science*, **2**, 25–48. 1.1, 3

Jervis, R. (1984) The illogic of american nuclear strategy. 6

Joseph, M. (2020) The secret failure of nuclear containment, *Manuscript*. (document), 1.1, 1.3, 5

Katz, Y. (2010) And they struck them with blindness, *The Jerusalem Post*. 1.1

Koblentz, G. D. (2018) Saddam versus the inspectors: the impact of regime security on the verification of iraq's wmd disarmament, *Journal of Strategic Studies*, **41**, 372–409. (document), 5

Kreps, S. E. and Fuhrmann, M. (2011) Attacking the atom: Does bombing nuclear facilities affect proliferation?, *Journal of Strategic Studies*, **34**, 161–187. 1.1

Kydd, A. (2000) Arms races and arms control: Modeling the hawk perspective, *American Journal of Political Science*, pp. 228–244. 4.4

Kydd, A. H. and McManus, R. W. (2017) Threats and assurances in crisis bargaining, *Journal of conflict resolution*, **61**, 325–348. 1.3

Lanoszka, A. (2016) Russian hybrid warfare and extended deterrence in eastern europe, *International Affairs*, **92**, 175–195. 1.1, 1.3

Levy, J. S. (2011) Preventive war: Concept and propositions, *International Interactions*, **37**, 87–96. 1.1

Lindsay, J., Gartzke, E., Greenhill, I. K. M. and Krause, P. J. (2016) Coercion through cyberspace: The stability-instability paradox revisited. 1.1

Mazarr, M. J. (2015) Struggle in the gray zone, *War on the Rocks*. 1.1

McCormack, D. and Pascoe, H. (2017) Sanctions and preventive war, *Journal of Conflict Resolution*, **61**, 1711–1739. (document), 1.1, 1.3

McMahon, R. B. and Slantchev, B. L. (2015) The guardianship dilemma: Regime security through and from the armed forces, *American Political Science Review*, **109**, 297–313. 1.3

Mehta, R. N. and Whitlark, R. E. (2017) The benefits and burdens of nuclear latency, *International Studies Quarterly*, **61**, 517–528. 1.3

Meirowitz, A., Morelli, M., Ramsay, K. W. and Squintani, F. (2019) Dispute resolution institutions and strategic militarization, *Journal of Political Economy*, **127**, 378–418. 1.3

Meirowitz, A. and Sartori, A. E. (2008) Strategic uncertainty as a cause of war, *Quarterly Journal of Political Science*, **3**, 327–352. 1.3

Morrow, J. D. (1989) Capabilities, uncertainty, and resolve: A limited information model of crisis bargaining, *American Journal of Political Science*, pp. 941–972. 1.3

Myerson, R. B. (1979) Incentive compatibility and the bargaining problem, *Econometrica*, pp. 61–73. 4.3

O'Neill, B. (2019) A game theory analysis of the stability–instability paradox, *Game Theory and Nuclear Stability in Northeast Asia*, p. 68. 6

Paine, J. (2020) The dictator's power-sharing dilemma: Countering dual outsider threats, *American Journal of Political Science*. 1.3

Pape, R. A. (1996) *Bombing to win*, Cornell University Press. 1.1

Post, A. (2019) Flying to fail: Costly signals and air power in crisis bargaining, *Journal of Conflict Resolution*, **63**, 869–895. 1.1

Powell, R. (1989) Nuclear deterrence and the strategy of limited retaliation, *The American Political Science Review*, pp. 503–519. (document), 1.3

Powell, R. (1999) *In the shadow of power: States and strategies in international politics*, Princeton University Press. 4.4

Powell, R. (2006) War as a commitment problem, *International organization*, **60**, 169–203. (document), 1.1

Powell, R. (2015) Nuclear brinkmanship, limited war, and military power, *International Organization*, **69**, 589–626. (document), 1.3

Ramsay, K. W. (2017) Information, uncertainty, and war, *Annual Review of Political Science*, **20**, 505–527. (document), 1.1, 1.3

Rauchhaus, R. (2009) Evaluating the nuclear peace hypothesis: A quantitative approach, *Journal of Conflict Resolution*, **53**, 258–277. 6

Reiter, D. (2005) Preventive attacks against nuclear programs and the "success" at osiraq, *Nonproliferation Review*, **12**, 355–371. 1.1

Salanié, B. (2005) *The economics of contracts: a primer*, MIT press. 4.3

Schelling, T. C. (1966) *Arms and influence*, Yale University Press. (document), 1.3

Schelling, T. C. (1980) *The strategy of conflict*, Harvard university press. 1.1, 1.3

Schram, P. (Forthcoming) Hassling: How states prevent preventive wars, *American Journal of Political Science*. (document), 1.1, 1.3

Schultz, K. A. (2010) The enforcement problem in coercive bargaining: Interstate conflict over rebel support in civil wars, *International Organization*, **64**, 281–312. 1.1, 1.3

Slantchev, B. L. and Tarar, A. (2011) Mutual optimism as a rationalist explanation of war, *American Journal of Political Science*, **55**, 135–148. (document), 1.3

Smith, A. (1998) International crises and domestic politics, *American Political Science Review*, pp. 623–638. 1.3

Snider, L. B. (2015) *The Agency and the Hill*, Government Printing Office. 6

Snyder, G. H. (1965) The balance of power and the balance of terror, in *World in Crisis* (Ed.) F. H. Hartmann, The Macmillan Company, chap. 20, pp. 180–191. 6

Spaniel, W. (2019) *Bargaining over the bomb: the successes and failures of nuclear negotiations*, Cambridge University Press. (document), 1.3, 5

Spaniel, W. (2020) Power transfers, military uncertainty, and war, *Journal of Theoretical Politics*, **32**, 538–556. 1.3, 4.4

Spaniel, W. and Malone, I. (2019) The uncertainty trade-off: Reexamining opportunity costs and war, *International Studies Quarterly*, **63**, 1025–1034. (document), 1.3, 6

Trager, R. F. (2010) Diplomatic calculus in anarchy: How communication matters, *American Political Science Review*, pp. 347–368. 1.3

U.S. Air Force (2019) Agm-86b/c/d missiles, Tech. rep. 5

Wagner, R. H. (2000) Bargaining and war, *American Journal of Political Science*, pp. 469–484. 1.3

Waltz, K. N. (1979) *Theory of international politics*, Waveland Press. (document)

Wittman, D. (1979) How a war ends: A rational model approach, *Journal of Conflict Resolution*, **23**, 743–763. 4.4

Woods, K. M., Palkki, D. D. and Stout, M. E. (2011) *The Saddam Tapes: The Inner Workings of a Tyrant's Regime, 1978–2001*, Cambridge University Press. 5

Woods, K. M., Pease, M. R., Stout, M. E., Murray, W. and Lacey, J. G. (2006) Iraqi perspectives project., Tech. rep., United States Joint Forces Command, Norfolk, VA. 5