When Deterrence Fails: How Improved Hassling Capabilities Produce Worse Outcomes*

Peter Schram

May 22, 2020

Abstract

I formalize interactions between an endogenously rising state and a rival, non-rising state that can accept the rising state's rise, can go to war before the rise comes to fruition, or can degrade the rising state's growth through low-level conflict operations that I call "hassling." The novelty here is that the non-rising state has private information about their hassling capabilities; this implies that the rising state does not know how fast it can rise without invoking the non-rising state to hassle or go to war. I find that when the non-rising state is better able to conduct hassling, it can invoke problematic strategic responses in the rising actor, undermine the non-rising state's ability to use its private information productively, and result in lower utilities for the non-rising state. Empirically, this model provides insight into Saddam Hussein's decision making leading up to the 2003 U.S. invasion, and proxy-wars that occurred during the Cold War.

*Thanks to Brad Smith, Andrew Coe, Brenton Kenkel, Kris Ramsay, William Spaniel, Michael Joseph, and the Vanderbilt Conflict Workshop and Faculty workshop for helpful comments. I am indebted toHaonan Dong, Ethan Sun, and the Vanderbilt ROCCA lab for outstanding research assistance. All errors are my own.

In the lead up to the 2003 U.S. invasion of Iraq, Iraq repeatedly turned away weapons inspectors and made false statements about its chemical and biological weapons programs. These behaviors led Iraq's adversaries to believe that under Saddam Hussein, Iraq was developing weapons of mass destruction (WMDs), when in reality, Iraq's pursuit of WMDs was limited. Saddam's behavior was puzzling. Saddam had reason to keep weapons inspectors out; Saddam was concerned that the weapons inspectors, in searching for WMDs, would document key details on the extent and location of his armaments, and that this information could be used maliciously by Iraq's adversaries (Coe and Vaynman, 2020). But this insight cannot fully explain Saddam's behavior, as a critical question remains: why would Saddam take on the enormous risk of keeping weapons inspectors out in the first place? Through the logic of commitment problems, Iraq's behavior made it appear to be a rising power, which put Iraq at risk for provoking a preventive war (Fearon, 1995; Powell, 2006; Debs and Monteiro, 2014). And, the 1990-91 Gulf War illustrated that the Iraqi Army was no match for U.S. forces. Through any sort of deterrence logic, it is puzzling why Saddam, knowing his army would be defeated, took such bold steps in pursuing policies that eventually produced the end of his regime.

Saddam took on risks in the lead up to 2003 because he was not only considering the Gulf War in his decision-making. 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 they had an ability and willingness to use targeted low-level conflict operations to handle revisionist powers. If Saddam considered Operation Desert Fox in his decision making, then there exists a troubling possibility: by being effective at operations falling below the traditional threshold of a decisive international war, the U.S. may have undermined its own deterrent threat from war and emboldened Saddam to behave as he did. While low-level conflict capabilities can be useful within political crises – Operation Desert Fox was, at the time, a success – possessing effective low-level conflict capabilities may provoke problematic strategic responses from opponents. The availability of technologies like 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 an endogenously rising power, the "rising state," that invests in a "rising technology," that would make the rising state more powerful in the future (like investing in a nuclear program). In response, a rival "non-rising state" can accept the investment, can declare a preventive war to stop the investment from coming to fruition, or can "hassle," which, following Schram (2020), is the use of limited conflict to degrade a rising state's investment in rising technology. What distinguishes this work from substantively similar research (like Bas and Coe (2016), Spaniel (2019, pp. 244-289) or Joseph (2020)) is that the non-rising state has private information about their hassling capabilities. In other words, here the rising state does not know how much it can invest in the rising technology without provoking a destructive response of hassling or war. In this setting, improved hassling capabilities can produce worse outcomes – more war, more hassling, and greater final levels of rising technology, collectively producing a lower utility – for the non-rising state. Thus, all else equal, being better at hassling can encourage a rising state to take undesirable actions in the eyes of the non-rising state, thus producing a deterrence failure.[1]

Improvements in hassling capabilities produce a deterrence failure when they interact with the non-rising state's private information. Two mechanisms can produce this result. First, improvements in the non-rising state's low-level conflict capabilities produce a deterrence failure by *emboldening the rising state*. When Iraq chose to turn away weapons inspectors

[1] As I define formally below, a state experiences a deterrence failure following improvements in hassling capabilities when, across all cases, that state does weakly worse.

and lie about its capabilities,[2] it did so knowing the U.S. was good at hassling (as evidenced by Operation Desert Fox) and Iraq expected hassling as the most likely response (but war was still a possibility). In a counterfactual setting where the U.S. was much worse at hassling, Iraq might expect that expelling weapons inspectors would be met with a higher likelihood of war relative to the case where the U.S. was effective at hassling; in this counterfactual setting, the greater likelihood of war could act as a more effective deterrent and could have convinced Iraq to be more open about their weapons program. In other words, because the U.S. was good at hassling, it emboldened Iraq to gamble by undertaking the actions that served as the justification for war.

As a second mechanism, improvements in the non-rising state's hassling capabilities can *make the non-rising state more predictable.* In the games here, private information is useful because it allows the player with private information to posture to attain better outcomes. The threat of war is an effective deterrent; if a rising state thinks that building a nuclear bomb will provoke their rivals to declare war today, then the rising state will not invest in the reactor. When the non-rising state effectively postures, the non-rising state convinces the rising state that the non-rising state would go to war over an investment in rising technology when, in fact, the non-rising state would not have gone to war. Under some conditions, improvements in the non-rising state's hassling capabilities diminish its ability to use war as a deterrent threat, thus allowing the rising state to calibrate its investments in nuclear technology to extract more bargaining surplus from the non-rising state. I examine this mechanism below in the context of the proxy conflicts that occurred during the Cold War and the Stability-Instability Paradox.

This is not the first paper to suggest that an improved capability to conduct low-level conflict or precision strikes against nuclear facilities can be counter-productive for the state

[2] I elaborate on this case in more detail below, but this model works so long that, in the U.S.'s eyes, Iraq appeared to be investing in rising technology.

making these improvements (Schelling, 1980; Powell, 2015; Bas and Coe, 2016; Joseph, 2020). However, this paper makes three important contributions. First, this game formalizes an important and previously unexplored strategic dynamic,[3] where a rising state wants to invest in a rising technology, but the rising state does not know how much it can invest before its opponents will respond with hassling or war. This dynamic is distinct from that explored in substantively similar work like Debs and Monteiro (2014), Bas and Coe (2016) and Spaniel (2019, pp. 244-289).[4] Second, by considering this previously unexplored interaction, this is the first paper to identify cases of deterrence failure stemming from the non-rising state becoming more predictable and the rising state becoming emboldened. Third, I show, for a broad class of models, that the predictability and emboldening mechanisms are not just two ways deterrence failures can occur, but rather the *only* ways deterrence failures can occur.

Hassling is and has been a pervasive feature of the international system. States have used hassling to degrade power shifts stemming from developing nuclear weapons (Operation Desert Fox, Operation Outside the Box, and the Stuxnet computer worm), future alliances (the Quemoy Crisis and Russian involvement in Ukraine), or possessing geographically valuable territory (U.S. supporting Afghan mujahideen to fight the Soviet Union).[5] And, previous work has demonstrated that hassling can be useful in international crises by offering a less-costly alternative preventive wars (Coe, 2018; Schram, 2020).[6] This paper confirms that hassling can be useful in the moment, but it also shows that building out a strong hassling capability can invoke aggressive strategic responses in adversaries. And, as I describe below,

[3]Given the vast literature on information asymmetry as a driver of international conflict, I claim that this is an important aspect to interactions between a rising state and non-rising state.

[4]Instead, Debs and Monteiro (2014) and Bas and Coe (2016) consider a setting where the non-rising state cannot perfectly observe the investment in rising technology. Put another way, in Debs and Monteiro and Bas and Coe, the rising state invests in rising technology, but is uncertain about the signal that the non-rising state will receive from this investment. In this paper, the rising state invests in rising technology, but is uncertain over what investment levels will be accepted by the non-rising state, and what levels will be met with a destructive response. Additionally, this is the only paper from this set to consider low-level conflict as a distinct policy option.

[5]These examples (and others) are discussed in Schram (2020) and (Benson and Smith, 2020).

[6]As McCormack and Pascoe (2017) shows, similar results can apply to sanctions.

this formalization offers insights into the proxy wars fought during the Cold War, into Saddam's decision making in the lead up to the 2003 U.S. invasion, and into the recent (circa 2020) emergence of "great power competition."

I proceed as follows. In Section 1 I outline the theory. In Section 2 I present the model, and in Section 3 I provide the general results for when improvements in hassling capabilities produce a deterrence failure, and I provide examples. In Section 4 I discuss extensions, in Section 5 I describe the empirical implications, and I conclude in Section 6.

1 Theoretical Background

1.1 Background and Terminology

This paper will reference several types of technologies and political behaviors. Through this subsection, I will Operation Outside the Box as a running example. In Operation Outside the Box, in 2007 Israel discovered that Syria was building a nuclear reactor. In response, Israel used a cyberattack to disable Syrian air defenses and conducted an airstrike on the reactor. When Syria was covertly building a nuclear reactor, it was becoming closer to being able to produce a nuclear weapon, thus it was investing in a "rising technology."

Definition: *States invest in* ***rising technologies*** *by pursuing opportunities – like investing in military technologies or conducting military operations – that strengthen their future military capabilities.*

If investments in rising technologies are allowed to come to fruition, the state making the investments will have improved abilities in future conflicts and more leverage in future negotiations. This dynamic creates commitment problems, and can encourage rival states to declare a preventive war (Fearon, 1995; Powell, 2006; Bas and Coe, 2012; Debs and Monteiro, 2014;

Bas and Coe, 2016). The concept of investing in rising technologies can apply to a range of political behaviors, including investing in military space or cyber technologies (Gartzke and Lindsay, 2017), amassing conventional forces (Copeland, 2001), forming alliances (Benson and Smith, 2020), and securing geopolitically valuable territory (Fearon, 1996; Powell, 2006).

Instead of declaring a preventive war, the non-rising state can "hassle" the rising state. When Israel conducted a cyberattack and airstrike against Syria to degrade Syria's nuclear aspirations, it engaged in "hassling."

Definition: Hassling *is the limited use of costly and destructive military capital against a targeted state with the intent of blunting power shifts to allow for bargaining between states to occur.*

Hassling was previously defined and discussed in Schram (2020). Fundamentally, hassling operates like a steam valve; in an international system where a rising power may provoke other states to declare a preventive war, hassling can diffuse the situation at a lower cost than a preventive war. Many instances of limited strikes against nuclear facilities (Reiter, 2005; Fuhrmann and Kreps, 2010; Kreps and Fuhrmann, 2011),[7] hybrid-conflict (Lanoszka, 2016; Trenin, 2018) and gray zone conflict (Mazarr, 2015; Votel *et al.*, 2016) could qualify as hassling. Furthermore, hassling operates similarly to how other recent work treats arming, sanctions, or containment regimes (McCormack and Pascoe, 2017; Coe, 2018; Joseph, 2020).

For Israel to conduct Operation Outside the Box, it relied on its cyber- and precision-strike-capabilities, or its "hassling capabilities."

[7]As discussed in Schram (2020), some contemporary definitions of preventive war are broad enough to include hassling (Levy, 2011). However, there is value in treating hassling as a distinct policy choice as there are some technologies that lend themselves better to low-level operations than larger, more conventional forms of escalation.

Definition: Hassling capabilities *are the tools that states use to hassle.*

I examine the equilibrium effects of improvements in hassling capabilities. Bas and Coe (2016) and Joseph (2020) both consider this substantive topic. The key distinction here is that I consider an environment with private hassling capabilities, which implies that the rising state is uncertain over how much it can invest in rising technologies before it invokes hassling or war.[8]

I want to highlight two features of hassling capabilities that I operationlize below. First, hassling technologies have both public and private components. For example, in 2007, Syria knew that Israel had some public latent hassling capability, as Syria could observe the 1981 Operation Opera, Israel's attack against an Iraqi nuclear reactor, as well as more recent Israeli operations into Lebanon and Palestinian territories. But, Israel's hassling capabilities also depended on private capabilities that Israel knew but Syria did not. For example, Syria likely did not know that Israel possessed a cyberweapon that could disrupt their air defenses. Furthermore, even if Syria had some idea that Israel possessed a cyberweapon like the one that was implemented, Syria did not know Israel's willingness to use the weapon in a strike on the Al Kibar reactor.[9] Of course, the feature that a state possesses private information about their ability to engage in conflict is not new, and the claim that states have private information about their hassling capabilities is consistent with claims that states have private information about their ability to wage war (see Fearon (1995), Fey and Ramsay (2011), and Spaniel (2019) as examples).

To further expand on this feature, I will refer to public improvements in hassling capa-

[8] As an additional distinction, this paper treats the set of hassling actions as convex, which is an important technical contribution.

[9] This issue is particularly salient for cybertechnology. Because once a cyberattack is detected the exploit can be fixed, a state like Israel may be unwilling to use an exploit in one case because then they could not use it in another case (**?**).

bilities, or improvements to the non-rising state'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.

As a second feature, I treat hassling capabilities as fixed, meaning I do not allow the non-rising state to tailor their hassling capabilities to the crisis at hand. This feature has empirical justification as it is difficult to vary latent hassling capabilities in the lead up to a political crisis. For example, during Operation Outside the Box, the Israeli Air Force deployed F-15Is and F-16Is, systems that Israel acquired in the 1990s when Israel's primary operations were in Lebanon or against Palestinian militants. Put another way, it is not as if Israel, in the lead up to Assad building a nuclear reactor, could have either destroyed their aircraft and reduced their hassling capabilities, or quickly expanded and acquired new systems to bolster their hassling capabilities.[10] Overall, a state's latent hassling capabilities at the time of a political crisis are subject to a wide range of factors, including past conflicts, delays in weapons development or procurement, or domestic politics, all of which is idiosyncratic to the conflict at hand. This all being stated, if this feature is viewed as incorrect, this paper still produces a surprising result: states may rationally turn down costless opportunities to improve their hassling capabilities due to the problematic strategic response it can induce in adversaries.

Having to conduct Operation Outside the Box represents a kind of deterrence failure. For Israel, ideally, Syria would have never invested in the reactor in the first place, and Israel would not have needed to conduct a costly and risky hassling campaign.

[10] A similar discussion could be made for the weapons systems used during Operation Desert Fox, like the AFM-86C cruise missile, B-1 Bomber, and B-52 Stratofortress, which were developed during the Cold War and later retrofit to deploy non-nuclear payloads (U.S. Air Force, 2019a,c,b).

Definition: *A state experiences a* ***deterrence failure*** *when a state fails to prevent an opponent from undertaking undesirable activities, resulting in the state experiencing overall worse outcomes.*

Following classical deterrence theory, when a state prevents an opponent from undertaking an action through the threat of retaliation, the state has deterred its opponent. When an opponent is undeterred from taking an action that the state dislikes, the state experiences a deterrence failure. This definition is useful because it does not identify a deterrence failure in terms of war. For example, a deterrence failure could also occur if Syria builds a nuclear reactor and then a nuclear bomb, and Israel does nothing to stop this from occurring. I include a further discussion of this concept in the formalization below.

1.2 Theory

I consider the following interaction. One state, a rising power, chooses how much to invest in its rising technologies. If the investments are allowed to come to fruition, the rising state can leverage the better capabilities into better future political outcomes. However, upon observing the investment, a rival non-rising state may go to war to decisively challenge the rising state before the capabilities come to fruition, or hassle and degrade the investments. And, because this response depends on the non-rising state's hassling capabilities (that has private components), the rising state does not precisely know how the non-rising state will respond to a selected level of investment. I formalize this interaction and consider what effect improvements in hassling capabilities have on the non-rising state. I find that better hassling capabilities can hurt the non-rising state through one of two mechanisms: improvements in public hassling capabilities may make the non-rising state more predictable, or may embolden the rising state.

How can improvements in hassling technologies make the non-rising state more predictable, with worse outcomes for the non-rising state? When private hassling capabilities (as opposed to public hassling capabilities) play an outsized role in determining the non-rising state's overall hassling capabilities, the non-rising state's response to a selected investment in rising technology is difficult to predict. This unpredictability can be valuable to the non-rising state; here, because it is difficult to know how the non-rising state will behave, the rising state may scale back their investment in the rising technology to avoid war. This is akin to the non-rising state posturing or bluffing to get the rising state to reduce its investments. However, when the non-rising state becomes publicly better at hassling, this can overshadow the relevance of its private capabilities in such a way that makes it predictable. When public improvements in hassling capabilities reduces the uncertainty in how the non-rising state will respond, the non-rising state cannot effectively posture. Here the rising state may be less deterred from aggressive investments in the rising technology as they would know that these investments would be met with hassling and not war. Essentially, improvements in public hassling capabilities can diminish the non-rising state's benefits from its private information.[11]

How can improvements in hassling technologies embolden a rising state? When choosing an investment level in rising technology, the rising state faces a trade-off: selecting a greater investment in rising technology can be beneficial when the non-rising state does not go to war, but the greater investment increases the likelihood that the non-rising state will declare war. The shape of this trade-off – whether a small increase in investments will produce a small or large increase in the likelihood of war – is dictated by the non-rising state's hassling capabilities, that consists of the interaction of public and private components. Under some trade-off shapes, – say a 10% increase in rising technology produces a 10% increase in the likelihood of war – the rising state is unwilling to increase their investments in rising technology. However, under a different trade-off shape that could be borne out when the

[11] In the Online Appendix, I offer an alternate take for the predictability mechanism through the lens of poker.

non-rising state is better at hassling – a 90% increase in rising technology produces only a 10% increase in the likelihood of war – the rising state would be willing to increase their investments. This latter case – where being better at hassling can encourage problematic behavior – can be worse for the non-rising state to an extent that offsets any of the gains that they have from being better at hassling.

What are the empirical implications of these results? In short, under select conditions, if the non-rising state is better at hassling, then it can incentivize rivals to more aggressively pursue nuclear weapons (or other attempts to grow in power), resulting in a greater likelihood of war, more pervasive and intensive hassling campaigns, or some combination of both. Unpacking this result in a more applied setting, it could be argued that the United States and its allies have systematically over-invested in their abilities to conduct hassling since the decline of the Soviet Union. And, it is possible to imagine a counterfactual universe where all else is equal, but the United States and its allies did not go down that path and were worse at conducting hassling. In this alternate counterfactual universe, fewer revisionist states would have pursued nuclear technologies and the United States and its allies would not have needed to conduct as many military operations or wars to prevent proliferation. As an alternate applied interpretation of these results, it could be argued that today the United States is engaging in a complex strategic optimization by developing weapons systems that could be used for hassling in expectation of future confrontations with Russia and China (Joseph (2020) suggests similar results). If this is the case and the level of hassling capabilities are tailored for great power competition, we might expect the sort of bad behavior outlined above occurring in states for whom the optimization is not tailored to. If instead, the United States pursued a different strategic optimization in developing hassling technologies, new states may emerge as active revisionist threats.

Importantly, this paper shows that becoming better at hassling does not always produce a deterrence failure. One way improvements in hassling can lead to better outcomes would be if the improvements discouraged the rival state from investing in their rising technology. For example, if Israel were so effective at hassling that it could consistently degrade its rival's nuclear programs at almost no cost, then Israel's adversaries may be deterred from investing in costly nuclear facilities that they knew would be destroyed. Of course, achieving this degree of hassling efficacy may be difficult, as rival countries may be willing to absorb huge costs to develop these weapons[12] or could discover countermeasures to shield themselves from hassling. Another way this can occur is when there are linkages between hassling capabilities and wartime capabilities, which is discussed in detail in Section 4.4.

1.3 Related Theory

The result that improvements in hassling or low-level conflict can lead to more conflict is not new (Schelling, 1980; Powell, 2015; Bas and Coe, 2016). The contribution of this paper is two-fold. First, this is the only paper to to examine interactions where an endogenously rising power is uncertain over its opponent's ability or willingness to conduct low-level conflict. Second, this paper does not only classify the predictability and emboldening mechanisms and show where they arise, but also shows, for a broad class of models, that these are the *only* two mechanisms that can produce a deterrence failure.

A now vast literature considers power shifts and preventive wars (Levy, 1987; Fearon, 1995; Powell, 2006), with a smaller set of papers now considering preventive wars with an endogenously rising power (Debs and Monteiro, 2014; Bas and Coe, 2016; Spaniel, 2019; Meirowitz *et al.*, 2019) or a state deliberately making some revisionist action (Schultz, 2010). These models share a common feature: the non-rising state cannot perfectly observe what actions the rising state is undertaking and bases its strategy on a signal. This paper is distinct, treating investments in rising technology as commonly observed. I have several comments

[12] For example, in 1965, Pakistan's Prime Minister Zulfiqar Ali Bhutto said Pakistanis "will eat grass and leaves" to pay for a nuclear bomb (Anderson and Khan, 1998).

on this distinction. First, my work can speak to a wide class of rising technologies outside of nuclear weapons development. For example, when the USSR advanced into geopolitically valuable Afghanistan, this "investment" in the USSR's future strength was common knowledge. Second, in many cases where one state is investing in nuclear technology, the non-investing state is aware of the investment. In Operation Opera, Operation Outside the Box, and the deployment of the Stuxnet Worm, enough was known about the respective nuclear programs to inform what the operations would target and the scope of the operations. Finally, even when investments in rising technology send stochastic signals to rivals, it's not as if the central tension explored here – the state investing in a rising technology is uncertain of how its opponent will respond to the investment – does not exist.

This paper is consistent with a broad literature suggesting that states face uncertainty over their opponent's military capabilities (Morrow, 1989; Fearon, 1995; Fey and Ramsay, 2011; Slantchev and Tarar, 2011; Spaniel and Bils, 2018).[13] This paper models a specific type of uncertainty, namely uncertainty over a state's ability and willingness to use hassling, which is new. This also makes this work distinct from a related set of research that considers how changes in wartime capabilities affects outcomes (like utilities and likelihood of war) (Benson *et al.*, 2016; Spaniel and Malone, 2019).

In all the models cited above, when one state is dissatisfied with their outcome, they declare war. Of course, the real world is not this simple and there are many possible destructive policy responses that decision makers select from, as highlighted in non-formal works like Levy (2008), Gartzke and Lindsay (2017), and Mehta and Whitlark (2017). An emerging branch of formal research considers a policymaker who faces a range of policy outcomes, including limited war (Powell, 2015) or efforts that degrade a rising power's rise (McCormack and Pascoe, 2017; Coe, 2018; Schram, 2020; Joseph, 2020). A common finding within

[13]See Ramsay (2017) for an excellent review.

this literature is that low-level conflict can be useful within a crisis. For example, if a rising state is investing in rising technology, then the non-rising state would want to be better at hassling so that the non-rising state could select the less costly option (hassling) over a preventive war (Schram, 2020). However, when the decision to invest in rising technology is endogenous and there is is uncertainty over the non-rising states willingness to use hassling, then improvements in hassling capabilities can interact with this private information in ways that invokes problematic strategic responses in its opponents (through the predictability or emboldening mechanisms). Put another way, being good at hassling may encourage the rising state to invest more in the rising technology in the first place.

2 Model

2.1 Characterizing Hassling Games

I consider two states, A and D, that are in a "hassling game" over a divisible asset of normalized value 1. The hassling game is based loosely on the standard crisis bargaining framework, but considers a new strategic interaction. In the hassling game, A wants to invest in the rising technology (denoted $t \in \mathcal{T} \subseteq \mathbb{R}_+$) to improve its future wartime capabilities. In response to the selected t,[14] D will either go to war to prevent the rising technology from coming to fruition, accept the rising technology, or engage in hassling (denoted $h \in \mathcal{H} \subseteq \mathbb{R}_+$) to degrade the rising technology. When D does not declare war, A's investment t and D's hassling level h affects A's future likelihood of winning in war, which affects future bargaining between A and D. Critically, D's costs from hassling consist of a public parameter $\alpha \in \mathcal{A} \subseteq \mathbb{R}_+$ and a private (i.e. known only to D) type $\theta \in \Theta \subseteq \mathbb{R}_+$. I explain feature in more detail in Section 1.1, but to offer one justification, α can be thought of as D's observed (i.e. what D has done in the past) hassling capacity, and θ could be thought of as D's willingness to use hassling. With this setup, A does not know how much it can invest before D starts hassling to degrade

[14] I treat t as commonly observed, an assumption that I describe in detail in Section 1.3.

A's investment, or before D goes to war to prevent A's investment from coming to fruition.

I outline the game form below. Regarding analysis, I will compare outcomes to the game across low and high public hassling capabilities, or across set elements $\{\underline{\alpha}, \bar{\alpha}\} \subseteq \mathcal{A}$, with $\underline{\alpha} < \bar{\alpha}$.

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

2. State A selects rising technology level $t \in \mathcal{T} = \mathbb{R}_+$.

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 $h \in \mathcal{H} = \mathbb{R}_+$ (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 $U_A = P(0,0) - \omega_A$ and $U_D = 1 - P(0,0) - \omega_D$ for States A and D (respectively; I characterize the P function below).

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

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

For ease, I summarize all possible game outcomes in Table 1.

Scenario	A's utility	D's utility
D goes to war at Stage 3 (before h and t are realized)	$P(0,0) - \omega_A$	$1 - P(0,0) - \omega_D$
A goes to war at Stage 5 (after h and t are realized)	$P(t,h) - \omega_A$	$1 - P(t,h) - \omega_D - \frac{h^2}{F(\alpha,\theta)}$
A 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 $P : \mathcal{T} \times \mathcal{H} \to [0,1]$ is the likelihood that A wins in a war. I assume functional form $P(t,h) = max\{min\{1, \rho + t - h\}, \rho\}$ with $\rho \in [0,1]$, which implies P is weakly increasing in t and $-h$, and that P falls between ρ and 1 inclusive. Furthermore, the feature that $P(0,0) = \rho$ implies that while hassling can degrade investments in rising technology t, hassling can never make A a declining state, and, at most, hassling returns A to a baseline war victory likelihood of ρ.[15] This functional form may be perceived as undesirable as it contains "kinks" that could drive results; to address this issue, I will demonstrate cases, analysis, and alternate functional forms below and in the appendix where the kinks play no role in any actor's behavior.

When D goes to war in Stage 3, the actors fight over the asset of normalized value 1. This case closely resembles the standard Fearon (1995) costly lottery treatment of war. A's likelihood of winning in war is $P(0,0) = \rho$, and $\omega_A > 0$ and $\omega_D > 0$ are A's and D's costs from war (respectively).

When A goes to war in Stage 5, this represents a war after the investments in rising technology and hassling have come to fruition. The P function, ω_A, and ω_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 $F : \{\mathcal{A} \times \Theta\} \to \{\mathbb{R}_{>0}\}$, and where F is strictly increasing in α and θ. This

[15] While it is possible to imagine low-level conflict being used to turn a state into a declining power, this falls outside of the scope of the analysis.

functional form implies that D faces lower costs to hassling as α and θ increase, and that D pays no costs when $h = 0$ (i.e. from accepting). Throughout the paper, I will define parameters $\underline{\alpha} \in \mathcal{A}$ and $\bar{\alpha} \in \mathcal{A}$ with $\underline{\alpha} < \bar{\alpha}$.

For ease, I do not assume that A faces costs to investing in the rising technology t; I explore a model with these costs in an extension in the Online Appendix, and the substantive results do not change.

When A accepts in Stage 5, this represents a bargained outcome after the rising technology and hassling have come to fruition. The value x denotes the offer made to A, and D still incurs the costs from hassling.

2.2 Equilibrium Concepts and Assumptions

I limit attention to pure strategy perfect Bayesian Nash Equilibria. The actions taken in the game depend on the observed and private components of D's capabilities. A strategy for State D is a mapping from the selected level of rising technology t and its hassling capabilities to its action space (consisting of w_D, h, and x), or $\sigma_D : (\mathcal{T}, \mathcal{A}, \Theta) \rightarrow \{0, 1\} \times \mathcal{H} \times [0, 1]$. Because State A does not know the value of θ, A's strategy is a mapping from the known parameter α to its action space (consisting of t and w_A), or $\sigma_A : \mathcal{A} \rightarrow \mathcal{T} \times \{0, 1\}$. I let σ denote a pair of strategies or $\sigma = (\sigma_A, \sigma_D)$, and a strategy profile $\sigma^* = (\sigma_A, \sigma_D)$ constitutes a Bayesian equilibrium if $\sigma_D(\mathcal{T}, \alpha, \theta)$ is a best response to σ_A, and $\sigma_A(\alpha)$ is a best response to σ_D based on the known capabilities parameter and expectations over D's type. For ease, I limit myself to pure strategy equilibria, I let x^*, h^*, w_A^*, w_d^*, and t^* denote equilibrium values of the various actions, and I let $\sigma^*(\alpha, \theta)$ denote the equilibrium for parameter $\alpha \in \mathcal{A}$ and type $\theta \in \Theta$.

At this point, I can formally define a "deterrence failure." The logic of the definition is that D wants to possess a set of technological capabilities that produces good outcomes (i.e.

greater utilities); thus, if improvements in hassling capabilities produce worse outcomes, it implies that A is undertaking problematic behavior.

Definition: *Improvements in publicly observed hassling capabilities (i.e. moving from* $\underline{\alpha}$ *to* $\bar{\alpha}$ *with* $\underline{\alpha} < \bar{\alpha}$*) produces a* ***deterrence 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* $\theta \in \Theta$.

Under the above definition, only when every possible type of D experiences a lower expected utility is there a deterrence failure. I include a discussion of why I opt for this definition rather than considering deterrence failure as measured by D's ex-ante expected utility in the online appendix.[16]

As a final assumption, I will limit my analysis below to scenarios where the constraints on the P function and final and x do not bind. This ultimately amounts to several technical conditions which result in me assuming, 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.[17] Furthermore, this allows the analysis to be done in a straightforward manner, without needing to consider excessive casework (i.e. defining where the constraints do or do not bind) while still using a benign functional form. In the Appendix I include alternate functional forms and specifications that do not rely on this assumption that illustrates identical findings, but these results are more complicated.

[16] I encourage readers to consult the Appendix for a more complete discussion, but to briefly summarize, using D's ex-ante expected utility is the right metric if D had no better knowledge of their type than A before the game began. In using the ex-ante expected utility definition of a deterrence failure, some types of D could attain strictly greater utilities when a deterrence failure occurs. Thus, if D knew their type before the game began, some types of D would actually prefer to experience a deterrence failure.

[17] Note that this game is fundamentally different from typical multi-stage games illustrating commitment problems, where the inability for the rising state to make a large enough concession in the first round produces a preventive war – see Fearon (1995). To give a sense of what this assumption does, I start the section below by defining $x^* = \rho + t^* - h^* - \omega_A$. Without this assumption, I must define $x^* = min\left\{max\left\{min\left\{max\left\{\rho + t^* - h^*, \rho\right\}, 1\right\} - \omega_A, 0\right\}, 1\right\}$ as part of D's optimization.

3 Equilibria

To offer some intuition, in the fourth and fifth stages, D does weakly better always avoiding war, which D can always accomplish by making A an offer equal to their wartime utility or $x^* = \rho + t^* - h^* - \omega_A$. In the third stage, D reacts to A's selected level of rising technology (t^*) by either going to war (setting $w_D^* = 1$) or not going to war (setting $w_D^* = 0$) and selecting some optimized hassling level h^* to counter the selected t^*. In the second stage, A runs their own optimization in selecting t^*, contingent on how A expects D will respond to the t^* conditional on D's known parameter and the distribution of D's types

I can further describe A's behavior. Because A faces no costs to investing in the rising technology,[18] A wants to invest in high levels of rising technology, but knows that if they select too great a t, D will respond with war. Thus, A's expected utility from increasing investments in rising technology is increasing, unless it provokes D to go to war, where D's cutoff strategy is determined on D's hassling capabilities. Because α is known and because A knows D is either type $\underline{\theta}$ or $\bar{\theta}$, for $\alpha \in \{\underline{\alpha}, \bar{\alpha}\}$, A will either select a t that would make a parameter α type $\underline{\theta}$ D indifferent between war and hassling, or select a t a parameter α type $\bar{\theta}$ D indifferent between war and hassling. In the former case, A avoids war altogether. In the latter case, A will go to war if D is a type $\underline{\theta}$.

Because the selected t's described above play a role throughout the rest of the analysis, I define them here.

Definition: *I define* $\boldsymbol{t(\alpha, \theta)}$ *as the level of rising technology that would make a D with parameter* $\alpha \in \{\underline{\alpha}, \bar{\alpha}\}$, *type* $\theta \in \{\underline{\theta}, \bar{\theta}\}$ *indifferent between war and hassling. Formally,* $t(\alpha, \theta) = \omega_D + \omega_A + \frac{F(\alpha,\theta)}{4}$

[18] I examine the case where A does face costs to investing in the Online Appendix and find substantively similar results.

3.1 Equilibria

Through the logic described above, there are two types of equilibria: a type where A optimally avoids war, and a type where war sometimes occurs. Proposition 1 lists some features of the pure strategy Perfect Bayesian Nash Equilibrium, and the appendix contains the full equilibrium.

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.*

- ***Case 1.*** When $Q(\alpha) \geq 0$ holds, *A selects technology level* $t^* = t(\alpha, \underline{\theta})$*, which results in both types of D hassling, setting* $h^* = \frac{F(\alpha,\theta)}{2}$ *for all* $\theta \in \{\underline{\theta}, \bar{\theta}\}$*. Type* $\underline{\theta}$ *D's attain utility* $U_D(\sigma^*(\underline{\theta}, \alpha)) = 1 - \rho - \omega_D$*, and type* $\bar{\theta}$ *D's attain utility* $U_D(\sigma^*(\bar{\theta}, \alpha)) = 1 - \rho - \omega_D + \frac{F(\alpha,\bar{\theta})}{4} - \frac{F(\alpha,\underline{\theta})}{4}$.
- ***Case 2.*** *When* $Q(\alpha) < 0$ holds,*A selects technology level* $t^* = t(\alpha, \bar{\theta})$*, which results in type* $\underline{\theta}$ *D's declaring war, and type* $\bar{\theta}$ *D's hassling, setting* $h^* = \frac{F(\alpha,\bar{\theta})}{2}$*. Both types of D attain their wartime utility, or* $U_D(\sigma^*(\theta, \alpha)) = 1 - \rho - \omega_D$ *for all* $\theta \in \{\underline{\theta}, \bar{\theta}\}$.

With $Q(\alpha) = Pr(\underline{\theta})\left(\omega_A + \omega_D\right) + \left(Pr(\bar{\theta}) - Pr(\underline{\theta})\right)\frac{F(\alpha,\underline{\theta})}{4} - Pr(\bar{\theta})\frac{F(\alpha,\bar{\theta})}{4}$.

Proof: See appendix.

In Case 1, A will seek to avoid war altogether. In Case 1, A will select a level of rising technology that makes the type of D that is least willing to hassle – type $\underline{\theta}$ – indifferent between hassling and war (as defined earlier as $t(\alpha, \underline{\theta})$), thus giving type $\underline{\theta}$ the wartime utility $\rho - \omega_A$. Also within Case 1, because A selects $t(\alpha, \underline{\theta})$, the type D's that are more willing to hassle – type $\bar{\theta}$ – will also hassle and will attain a utility above their wartime utility because hassling is cheaper for them. In the Case 2, A will sometimes risk war. In Case 2, A will select a level of rising technology that makes the type of D that is most willing to hassle

– $\bar{\theta}$ – indifferent between hassling and war $t(\alpha, \bar{\theta})$, thus giving type $\bar{\theta}$ their wartime utility $\rho - \omega_A$. Also within Case 2, because A selects $t(\alpha, \bar{\theta})$, the type D's that are less willing to hassle – type $\underline{\theta}$ – will want to go to war and will attain their wartime utility. Thus, $Q(\alpha)$ is simply the cutpoint where, for values below the cutpoint, A does better selecting a greater t^* (selecting $t(\alpha, \bar{\theta})$ over $t(\alpha, \underline{\theta})$) and sometimes risking war.

4 Results

In this section, I will first discuss how improvements in public hassling capabilities can lead to D becoming more predictable or A becoming emboldened (which occur under distinct conditions), thus producing a deterrence failure. I then discuss how these two mechanisms are the *only* way a deterrence failure can occur in the model. I then describe some additional results on improvements in private hassling capabilities, welfare (pending), and when improvements in hassling capabilities also change wartime capabilities. For a further discussion on results, see Section 6.

4.1 Predictability and Deterrence Failure

The key idea behind improvements in public hassling capabilities making D more predictable and producing a deterrence failure is as follows: in the game, D (weakly) benefits from their private information.[19] If a shift in D's public hassling capabilities makes D more predictable across their private types – meaning that different types of D select a more similar level of hassling – A can select an investment in rising technology that better extracts the benefit that D attains from their private information. This occurs despite war occurring weakly less following the change. Figure 1 visualizes this intuition using the results from a fully

[19] In an equivalent setup but with full information, A would always select the level of rising technology that makes D indifferent between war and hassling, thus always giving D (regardless of their private type) their wartime utility. In the setting with private information, in Case 1 of Proposition 1, type $\bar{\theta}$ D's attain a utility above their wartime utility because A selects the t that makes a type $\underline{\theta}$ D indifferent between war and hassling.

parameterized example. For the selected parameters, under both $\underline{\alpha}$ and $\bar{\alpha}$, the equilibrium is described in Case 1 in Proposition 1, where A always avoids war.

Under the selected parameters, across $\underline{\alpha}$ and $\bar{\alpha}$, A selects the largest investment in rising technology (t) that will keep D from ever declaring war. Because A observes D's parameter α but not D's type θ, the largest t that A can select (while still avoiding war) is the t that would keep a parameter α type $\underline{\theta}$ indifferent between hassling and going to war, or $t(\alpha, \underline{\theta})$. In the top two panels of Figure 1, A's optimal investment level $t^* = t(\underline{\alpha}, \underline{\theta})$ is indicated by an asterisks. In the top panel, the selected t^* results in a D with capabilities $(\underline{\theta}, \underline{\alpha})$ made indifferent between war and hassling, implying that a $(\underline{\theta}, \underline{\alpha})$ capabilities D will attain its wartime utility. In the second panel, the space between the selected level of rising technology and the level of rising technology that would have made a capabilities $(\bar{\theta}, \underline{\alpha})$ D indifferent between hassling and going to war (that is marked with a dashed line) represents D attaining some surplus. Instead of always receiving its wartime payoff, types $\bar{\theta}$ are better off *due to their private information*; in other words, so long that types $\bar{\theta}$ D's are able to keep their type private in the lead-up to A's selection of t – in other words, they are able to successfully posture[20] – they can attain some bargaining surplus.

However, D can become more predictable when becoming better at hassling (improvements in α) infringes on D's ability to use its private information. Because A has some uncertainty over how willing D is to hassle, A must scale back its investment in rising technology and benchmark it against type $\underline{\theta}$ if they seek to avoid war. If improvements in D's known hassling abilities reduce the salience of D's unknown type, then A does not need to scale back as much to prevent war. Comparing the top two and bottom two panels of Figure 1 suggests how this may occur. Whereas in the top two panels there is a significant gap between A's selected investment $t(\underline{\alpha}, \underline{\theta})$ and the point that would make a $(\underline{\alpha}, \bar{\theta})$ D indifferent

[20]The "posture" can be thought of as a $(\underline{\alpha}, \bar{\theta})$ D telling A that they are actually a $(\underline{\alpha}, \underline{\theta})$ D, and that, in order to avoid war, A must select a low level of rising technology.

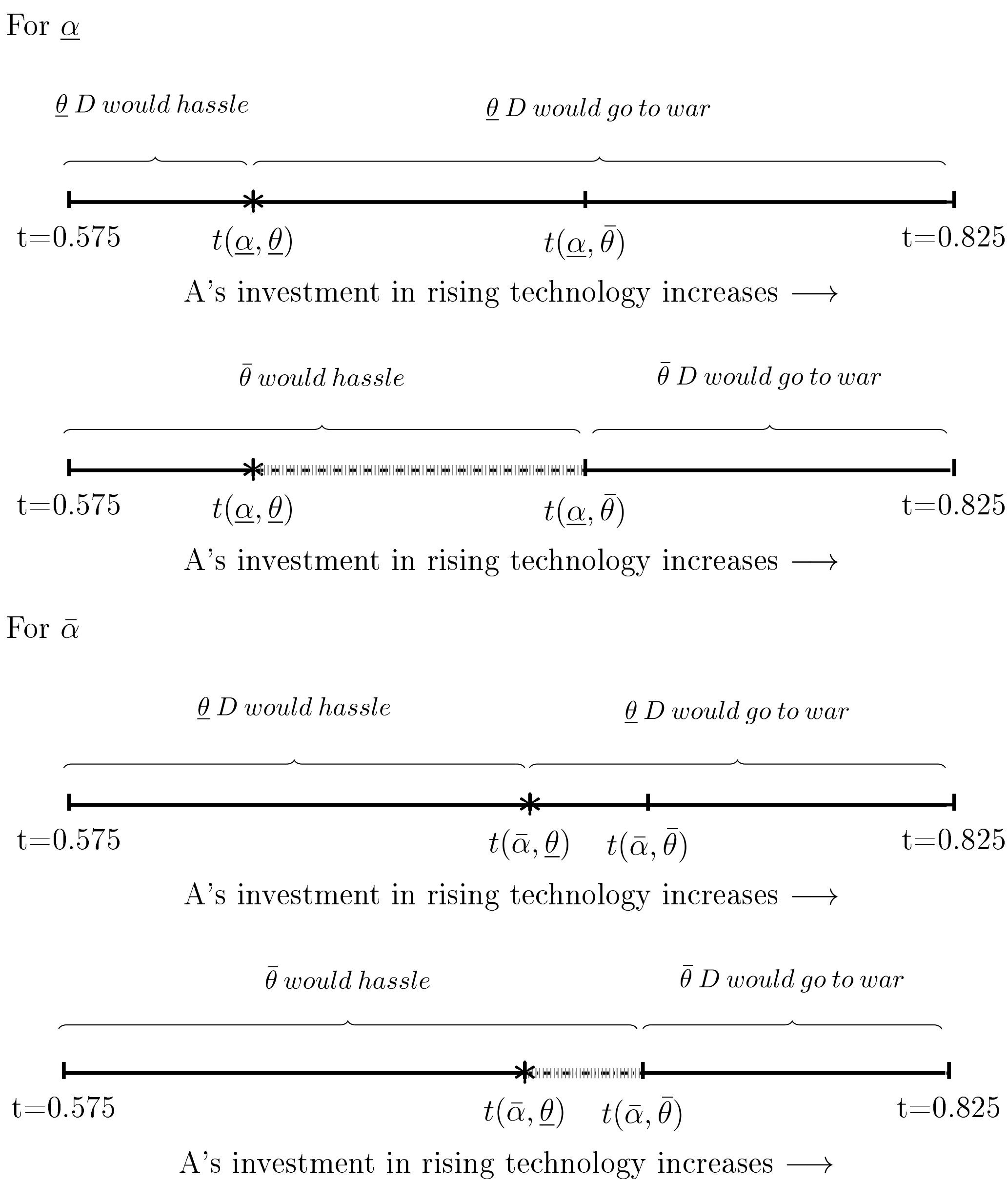

Figure 1: Optimal rising technology levels and D's response (Predictability).

A's selected level of investment under parameters $\underline{\alpha}$ and $\bar{\alpha}$ are denoted by the asterisks. D's response to given t's are bracketed off. The dashed lines represent D's the surplus type $\bar{\theta}$ D's attain from their private information in equilibrium. Parameters are $\omega_D = 0.1$, $\omega_A = 0.4$, $\rho = 0.5$, $Pr(\underline{\theta}) = 0.5$, $Pr(\bar{\theta} = 0.5)$, $F(\underline{\alpha}, \underline{\theta}) = 0.5$, $F(\underline{\alpha}, \bar{\theta}) = 1$, $F(\bar{\alpha}, \underline{\theta}) = 0.8$, $F(\bar{\alpha}, \bar{\theta}) = 1.1$.

between war and hassling ($t(\underline{\alpha}, \bar{\theta})$), for a paramter $\bar{\alpha}$ D there is now a much smaller gap. This implies that as D becomes better at hassling (moving from $\underline{\alpha}$ to $\bar{\alpha}$), A is able to select a level of investment in rising technology closer to the point that would give D its wartime payoff. Put another way, in the bottom two panels, it would not matter much if D's private type were revealed when D has parameter $\bar{\alpha}$ because types $\bar{\theta}$ do not attain much surplus from its private information; here D's improvements in $\bar{\alpha}$ have made D more predictable. As depicted, it is possible that as D becomes better at hassling, A can better tailor its level of investment in rising technologies across D's types, resulting in D's private information being less valuable for extracting surplus from the game.

Proposition 2 defines, for the hassling game characterized above, the necessary conditions for improvements in public hassling capabilities to make D more predictable and produce a deterrence failure.

Proposition 2 (Predictability): *Under the "Predictability Conditions," A avoids war across parameters $\underline{\alpha}$ and $\bar{\alpha}$ (formally $Q(\underline{\alpha}) \geq 0$ and $Q(\bar{\alpha}) \geq 0$), and D's private information plays a diminished role under parameter $\bar{\alpha}$ relative to parameter $\underline{\alpha}$ ($F(\underline{\alpha}, \bar{\theta}) - F(\underline{\alpha}, \underline{\theta}) > F(\bar{\alpha}, \bar{\theta}) - F(\bar{\alpha}, \underline{\theta})$. If the Predictability Conditions hold, then improvements from $\underline{\alpha}$ to $\bar{\alpha}$ produce a deterrence failure.*

Proof: See Appendix.

Proposition 2 generalizes the intuition in the numerical example above. The first two Predictability Conditions imply that across parameters $\underline{\alpha}$ and $\bar{\alpha}$, the behavior is characterized in Case 1 of Proposition 1. The third condition captures the differences in how types $\bar{\theta}$ and $\underline{\theta}$ play the game across parameters $\underline{\alpha}$ and $\bar{\alpha}$. Intuitively, for a given α, A selects a fixed t^* to make type $\underline{\theta}$ D's indifferent between war and hassling. Because $\bar{\theta}$ types face lower costs from hassling, type $\bar{\theta}$ D's select greater levels of hassling – as captured in the optimal hassling level $h^*(\theta) = \frac{F(\alpha,\theta)}{2}$ – and therefore attain a greater utility – as captured in the $\frac{F(\alpha,\bar{\theta})}{4} - \frac{F(\alpha,\underline{\theta})}{4}$ term in type $\bar{\theta}$'s utility function. When $F(\alpha,\bar{\theta}) - F(\alpha,\underline{\theta})$ is small, it implies, that types $\bar{\theta}$ and $\underline{\theta}$ do not play the game particularly differently (the h^* are similar), implying that D's private information is not particularly valuable to D (formally, $\frac{F(\alpha,\bar{\theta})}{4} - \frac{F(\alpha,\underline{\theta})}{4}$ is smaller). Therefore, when the bottom condition holds and the F function exhibits decreasing differences, the improvement in α degrades the value of D's private information.

4.2 Emboldening and Deterrence Failure

The key idea behind improvements in public hassling capabilities emboldening A and producing a deterrence failure is as follows: in the game, A wants to invest in the rising technology, but may be deterred from investing if further investment leads to a greater probability of war. If a shift in D's public hassling capabilities emboldens A, then A's upside to selecting greater investments in rising technology grows, overshadowing the downside of a greater probability of war. Together, because A is risking war more and selecting greater levels of investment, D suffers when A is emboldened.[21] Figure 2 visualizes this intuition using the results from a fully parameterized example. As can be seen, A's expected utility is increasing in the selected t (moving east along the x-axis), until this selection goes too far and provokes first types $\underline{\theta}$ (at $t(\cdot,\underline{\theta})$) to go to war (thus producing the first discontinuity), and then all types (at $t(\cdot,\bar{\theta})$) to go to war (producing the second discontinuity). For the selected parameters, under $\underline{\alpha}$ the equilibrium is described in Case 1 in Proposition 1 (where A always avoids war), and under $\bar{\alpha}$ the equilibrium is described in Case 2 (where A sometimes goes to war). Because A is choosing to go to war with a greater likelihood following the change in α, I plot A's expected utility on the y-axis.[22]

[21] It is worthwhile highlighting that the emboldening result is borne out through a similar mechanism that is explored in Spaniel and Malone (2019).

[22] The Online Appendix also contains a figure that is analagous to Figure 1 above, only that describes the emboldening case.

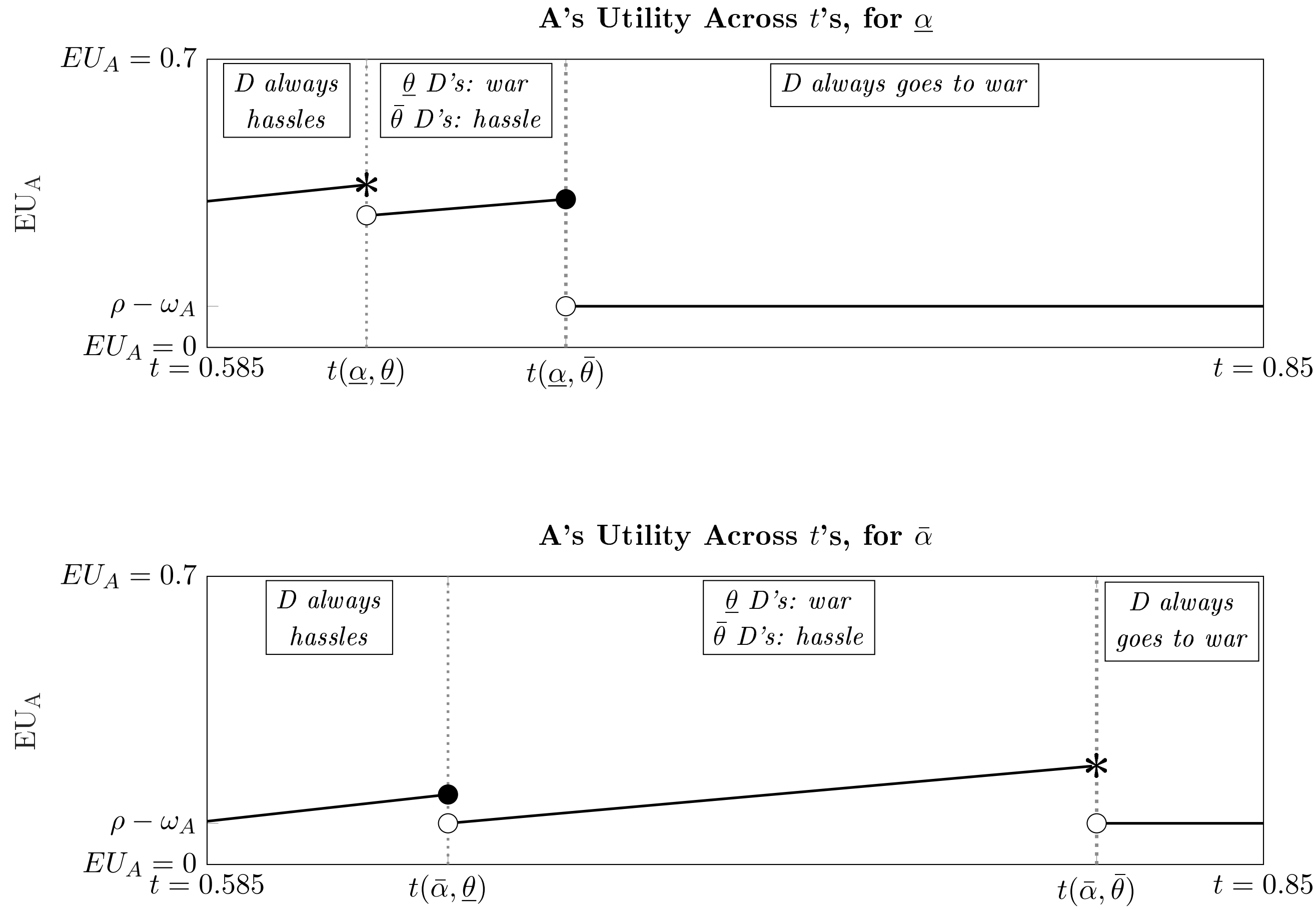

Figure 4. Emboldening and State A's Utilities Across Values of t. A's selected level of investment under parameters $\underline{\alpha}$ and $\bar{\alpha}$ are denoted by the asterisks. The parameters are $\omega_D = 0.1$, $\omega_A = 0.4$, $\rho = 0.5$, $Pr(\underline{\theta}) = 0.2$, $Pr(\bar{\theta}) = 0.8$, $F(\underline{\alpha}, \underline{\theta}) = 0.5$, $F(\underline{\alpha}, \bar{\theta}) = 0.7$, $F(\bar{\alpha}, \underline{\theta}) = 0.6$, $F(\bar{\alpha}, \bar{\theta}) = 1.3$.

The top panel of Figure 2 describes the game under parameter $\underline{\alpha}$. In this game, A optimally selects $t^* = t(\underline{\alpha}, \underline{\theta})$, which is characterized by the asterisks. As a response, both types D will always hassle. It is worthwhile highlighting what A does not do: A could have selected $t(\underline{\alpha}, \bar{\theta})$.[23] This "move-not-taken" captures A's trade-off between greater investments in rising technology and more war. As an "upside" to selecting $t(\underline{\alpha}, \bar{\theta})$, A's increased investment level would put A at an advantage when facing a type $\bar{\theta}$ D, thus granting A a larger final

[23]Note that A would never select a t greater than the value that would make a type $\bar{\theta}$ indifferent between hassling and war because this would always result in war, which is strictly worse for A.

utility.[24] However, A's "downside" to selecting a $t(\underline{\alpha}, \bar{\theta})$ is that now all type $\underline{\theta}$ D's will go to war, increasing the likelihood of war from zero to the likelihood that D is type $\underline{\theta}$. Thus, the difference between $t(\underline{\alpha}, \underline{\theta})$ and $t(\underline{\alpha}, \bar{\theta})$ can be thought of as the amount of rising technology that A is could invest in, if A was willing to go to war with type $\underline{\theta}$ D's. Under the parameter $\underline{\alpha}$, A's upside to doing better against type $\bar{\theta}$ D's does not outweigh the increased likelihood of war and costs that come with it, as visualized by the asterisks at $t(\underline{\alpha}, \underline{\theta})$ being taller than the point at $t(\underline{\alpha}, \bar{\theta})$.

Under parameter $\bar{\alpha}$, A faces a new trade-off, as represented by the greater difference between $t(\bar{\alpha}, \underline{\theta})$ and $t(\bar{\alpha}, \bar{\theta})$ (relative to the difference between $t(\underline{\alpha}, \underline{\theta})$ and $t(\underline{\alpha}, \bar{\theta})$). Under parameter $\bar{\alpha}$, if A were to move from selecting $t(\bar{\alpha}, \underline{\theta})$ to selecting $t(\bar{\alpha}, \bar{\theta})$, A faces the same downside of provoking type $\underline{\theta}$ D's to go to war. However, as can be seen, now A faces a much greater upside to selecting $t(\bar{\alpha}, \bar{\theta})$, because these cutpoints are so different.[25] Under the parameter $\bar{\alpha}$, A's upside to doing better against type $\bar{\theta}$ D's outweighs the increased likelihood of war, as visualized by the asterisks at $t(\bar{\alpha}, \bar{\theta})$ being taller than the point at $t(\bar{\alpha}, \underline{\theta})$. Intuitively, the improvement in D's public hassling capabilities offered A incentives to pursue a riskier investment strategy in their rising technologies and sometimes invoke war. Thus here, the improvements in α emboldened A in its pursuit of rising technology.

Proposition 3 defines, for the hassling game characterized above, the necessary conditions for improvements in public hassling capabilities to embolden A and produce a deterrence failure.

Proposition 3 (Emboldening): *Under the "Emboldening Conditions," A avoids war under parameter $\underline{\alpha}$ and goes to war under parameter $\bar{\alpha}$ (formally $Q(\underline{\alpha}) \geq 0$ and $Q(\bar{\alpha}) < 0$). If the following Emboldening Conditions hold, then improvements from underlineα to $\bar{\alpha}$ produce a*

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

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

deterrence failure.

Proof: See Appendix.

Proposition 3 generalizes the intuition in the numerical example above. The $Q(\underline{\alpha}) \geq 0$ conditions implies that A will not risk war and select $t^* = t(\underline{\alpha}, \underline{\theta})$, which will grant type $\bar{\theta}$ D's a utility above their wartime payoff. The $Q(\bar{\alpha}) < 0$ conditions implies that A will risk war with type $\underline{\theta}$ D's and select $t^* = t(\bar{\alpha}, \bar{\theta})$, which will result in all types of D attaining their wartime utilities. Because the $\bar{\theta}$ types do worse following the shift from $\underline{\alpha}$ to $\bar{\alpha}$, this produces a deterrence failure for D.

4.3 When (and Only When) Improved Capabilities Create Deterrence Failures

Within the model and model assumptions characterized above, it is only when the Predictability Conditions or the Emboldening Conditions hold that a deterrence failure is possible.

Proposition 4: Improvements *in α produce a deterrence failure if and only if the Emboldening or Predictability Conditions hold.*

Proof: See Online Appendix

In the proof of Proposition 4, I examine every possible outcome to the model described above, and I show that the only way a deterrence failure can occur is when D becomes more predictable or A is emboldened. Thus, for this model and under these assumptions, I have comprehensively characterized when improved public hassling capabilities produce worse outcomes for the state making these improvements.

5 Empirical Implications

In this section, I describe how the model offers insights into the 2003 U.S. Invasion of Iraq and political behavior occurring during the Cold War that would fall under the scope of the Stability-Instability Paradox.

5.1 Emboldening: Saddam's Gamble

Before discussing the case, it is worthwhile highlighting an empirical implication of the above theory. The theory suggests that being better at hassling can invoke rivals to change their behavior, which can, through the emboldening mechanism, lead to an escalation to war. Put another way, improvements in the technological capabilities for low-level conflict (hassling) can lead to more instances of high-level conflict (war).[26] This can explain previously unexplained features surrounding Saddam's behavior in the lead up to the 2003 U.S. Invasion of Iraq.

For the theory to apply to the case, three questions must be addressed: Was Iraq "investing in rising technology" as defined above? Did the U.S. demonstrate a robust hassling capability? And, did the U.S.'s hassling capabilities shape Iraqi behavior?

As defined above, by turning away weapons inspectors, Iraq was positioning itself as a rising power. In the eyes of the U.S., Iraq was keeping weapons inspectors out in an attempt to hide their WMD program, which suggested that Iraq could possess a greater military capability in the future. And, while Iraq was not actually investing in WMDs, activities like keeping out weapons inspectors would qualify as "investing in rising technology" so long that Iraq's actions strengthen their future military capabilities. As discussed in Koblentz (2018) and Coe and Vaynman (2020), state typically are concerned with allowing external observers to

[26] Critically, this result arises without any kind of assumptions that low-level conflict can probabalistically escalate, like that in Powell (2015).

inspect their military capabilities because this information could eventually be used against them by the U.S. or regional adversaries. Thus, in keeping weapons inspectors out, Iraq was insuring a stronger future wartime capability, which is in-line with the formalization above.

It is worthwhile elaborating on the investments in rising technology that Iraq made. In the model above, I treat investments in rising technology as commonly observed. Of course, I am not claiming that the U.S. observed Iraq's investments in WMD technology, as this clearly was not the case! Instead, Saddam's investment in rising technology was keeping the weapons inspectors out. The Saddam regime deliberated over how open to be with respect to revealing aspects of the state's security forces to weapons inspectors, knowing that less-open behavior could be met with a destructive response.[27] Essentially, Saddam's handling of weapons inspections in the decade preceding the 2003 invasion set the stage for the U.S. to be uncertain over Iraq's WMD programs and concerned that Iraq was developing the weapons; the eventual decision to invade was determined by U.S. willingness (or unwillingness) to tolerate the uncertainty that Saddam's actions created, or to use hassling or war to undermine the suspected program.[28] Overall, Saddam's decision to turn away weapons inspectors benefited his regime, was observed by the U.S. and its allies, and was eventually internalized and informed the decision to invade (not hassle) the regime in 2003, all making it consistent with the treatment of investments in rising technology above.

In the lead up to the 2003 invasion, the U.S. both possessed and previously demonstrated a robust hassling capability. In the aftermath of the Cold War, the U.S. successfully transformed much of its technical know-how in building and deploying missiles and bombers used for nuclear strikes into precision strike capabilities (U.S. Air Force, 2019a,c,b). The efficacy of U.S. bombing strikes and the willingness to rely on airstrikes was revealed in a series of

[27] For examples, see Woods *et al.* (2006, 15-16, 30, 91, 96-97, 125) and Woods *et al.* (2011, 257-258).

[28] This willingness to tolerate Saddam's treatment of inspections is analogous to the private information parameter θ variable in the model, as in his decisions surrounding weapons inspectors, Saddam could not know the precise level of openness to select (the t action) that would lead to acceptance, hassling, or war.

conflicts, including the NATO bombing of Bosnia and Herzegovina, the NATO bombing of Yugoslavia, and U.S. strikes against Iraq in 1998 in Operation Desert Fox.

Finally, Iraq was both aware of U.S. capabilities, and based their decisions regarding inspectors on this fact. After Saddam was captured by Coalition forces, he was interviewed by FBI agent George L. Piro (Battle, 2009). In one such interview, Piro asked Saddam why he turned away weapons inspectors, which resulted in the following retelling:

> *Even though Saddam claimed Iraq did not have WMD, the threat from Iran was the major factor as to why he did not allow the return of the UN inspectors. Saddam stated he was more concerned about Iran discovering Iraq's weaknesses and vulnerabilities than the repercussions of the United States for his refusal to allow UN inspectors back into Iraq. [...] Saddam indicated he was angered when the United States struck Iraq in 1998. Saddam stated Iraq could have absorbed another United States strike for he viewed this as less of a threat than exposing themselves to Iran.*

This quote suggests that Saddam considered the U.S.'s robust hassling capabilities as the more likely possibility when making his decision to turn away UN inspectors. And, while it could be believed that Saddam possessed incentives for lying about this point, in internal discussions leading up to the 2003 invasion, Saddam commented frequently on the U.S.'s reliance on air strikes for international conflict, a sentiment echoed by Saddam's advisers (Woods *et al.*, 2006, 15-16, 30, 96-97, 125). At one point, one such (unidentified') 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).

That Saddam changed his behavior in response to improved U.S. hassling capabilities (i.e.

improvements that facilitate a low-level of escalation) in such a way that led to a war (i.e. a high-level of escalation) has not previously been formalized or explained. Somewhat similar results exist in the literature – both Bas and Coe (2016) or Joseph (2020) suggest that being better at low-level operations like preventive strikes can lead to more destructive operations – but the literature cannot currently explain how being better at hassling can lead to an escalation to war. For example, the Joseph result suggests that being better at hassling can lead to more proliferation and more hassling occurring in the world, but never finds that it can produce more war.[29]

The theory and case above suggest a new perspective on the 1990's "revolution in military affairs," where improvements in precision strike technologies ultimately resulted in the U.S. using these technologies more within low-level operations. Here Saddam observed the U.S.'s increased reliance on hassling technologies and took a gamble in turning away weapons inspectors, believing that his decision would be met with, at worst, a limited response and not a full invasion aimed at overthrow. While there is no way to know how Saddam would have behaved in the absence of the technologies that allowed for Desert Fox to occur, it is plausible that Saddam would have been less willing to risk provoking the U.S. if the U.S. were less reliant on low-level operations and the threat of war served as a stronger deterrent.[30]

5.2 Predictability and the Stability-Instability Paradox

Before discussing this theory in the context of the stability-instability paradox, it is worthwhile highlighting more implications of the theory above. When the predictability mechanism is in play, then improvements in hassling capabilities are met with opponents more aggressively investing in rising technology, knowing that the more aggressive investments will be met with hassling rather than war. This suggests that if the non-rising state were worse at

[29] Bas and Coe (2016) find substantively similar results, where being better at preventive attacks, through strategic affects, can make preventive attacks more likely.

[30] This discussion echoes concerns in Cordesman (1999).

hassling, the threat of escalation to war could serve as a more successful deterrent against the investments in the rising technology.

During the Cold War, a series of scholars observed a kind of paradox: while mutually assured destruction kept the global superpowers from engaging in a nuclear exchange, below the threshold of nuclear war, the superpowers competed in a range of conflict theaters. These observations informed a new theory in international politics, the "stability-instability paradox," which describes how parity (or stability) at one level of conflict opens the opportunity for lower-levels of conflict (or instability) (Snyder, 1965; Jervis, 1984). Consistent with the theory, throughout the Cold War the U.S. and U.S.S.R. maintained a second-strike capability that insured stability at the nuclear level, and the U.S. and U.S.S.R. engaged in territorial power grabs (like the U.S.S.R invading Afghanistan) or backed pro-communist or pro-western rebels and governments across the globe (like the U.S. providing arms, funds, and training to Guatemalan rebels, who overthrew Jacobo Árbenz).

The topics covered in this paper similarly apply to the the proxy wars that occurred during the Cold War, where the U.S. and U.S.S.R. spent considerable effort through covert and overt measures to bolster their own allies and to frustrate the other side, all while to improve their own global standing. For example, the U.S.S.R. invasion of Afghanistan was conducted to insure the Afghan government remained a client of the Soviet Union. By invading, the U.S.S.R.'s operation attempted to secure future geopolitical strength, thus making the invasion a form of "investing in rising technology." In response, the U.S. degraded their attempts at future strength by supporting mujahideen rebels, making their response a form of "hassling." Because the many military operations that arose during the Cold War took a similar form of one side attempting to establish an international environment amenable to their own interests, while the other side attempted to deny this, the stability-instability paradox and the theory discussed here cover similar conflict phenomena (Gaddis *et al.*, 2005;

O'Brien, 1995).

While much of the behavior in the Cold War fell within the theorized behavior described in the stability-instability paradox, in the initial presentation of the paradox, Snyder (1965) leaves open the possibility for a different response to stability at the nuclear (or as he puts, "strategic") level. In the initial text, Snyder states the following:

> *Thus firm stability in the strategic balance tends to destabilize the conventional balance and also to activate the lesser nuclear "links" between the latter and the former. But one could argue precisely the opposite – that the greater likelihood of gradual escalation due to a stable strategic equilibrium tends to deter both conventional provocation and tactical nuclear strikes – thus stabilizing the overall balance. The first hypothesis probably is dominant, but it must be heavily qualified by the second, since nations probably fear the possibility of escalation "all the way" nearly as much as they fear the possibility of an "all-out" first strike.*

This paper can speak to Snyder's initial uncertainty over the existence of the paradox by presenting a critical scope condition: instability at lower-levels relies on each state possessing a capacity to engage at these lower levels, thus reducing the risk for a possible escalation (which in turn could produce stability). The predictability mechanism demonstrates how being highly capable at low-level conflict can lead to more investments in rising technology, having to use hassling more, and overall worse outcomes for all parties involved. Thus, the theory here suggests that it was not just stability at the nuclear level that created an abundance of conflict between the superpowers at lower levels, but it was also that the U.S. and U.S.S.R. expanded their ability to conduct these low-level operations. Both the CIA and the KGB embraced covert operations as a means for projecting or securing U.S. or Soviet (respectively) influence (O'Brien, 1995). In the case of the United States, after WWII, the funding for the CIA grew rapidly, both explicitly (Gaddis *et al.*, 2005, 154-156) and with the backing of congressional leadership (Snider, 2015, 161-163). The expansion of limited capabilities meant that a power-grab by one side could be met with a limited response. Put anther way, when the U.S.S.R. invaded Afghanistan, the considered response was not to put nuclear warheads on alert, but rather to covertly support rebels, like the U.S. had been doing for the duration of the Cold War. The theory here suggests that by becoming better at low-level operations, the U.S. and U.S.S.R. opened the door to more aggressive power grabs because the would-be rising power could predict that these attempts would, at worst, be met with a low-level response.

Today, several scholars and practitioners have observed a new emergence of great power competition and a new stability-instability paradox (Lindsay *et al.*, 2016; DoD, 2018). What is unique about this period's paradox is that the expansion in low-level engagement across the globe is not just conducted by states who have established a credible nuclear (or conventional conflict) deterrence, but also involves unbalanced dyads like the U.S. and Iran. The theory here suggests that this new competition could be conflated by the proliferation of precision, drone, and cyber technologies, which can cheaply and easily be used to hassle. The prevalence of these technologies being used at low-levels of conflict today offer some support of this theory.

Further on this topic, I include, in the Online Appendix, a further discussion on what these results mean for best-practices in developing hassling capabilities. To summarize, the key take-away is that insuring that improving hassling capabilities also comes with improved capabilities at higher levels of escalation can successfully prevent deterrence failures.

6 Additional Modeling Results and Extensions

6.1 Additional Results

6.1.1 Improvements in Private Hassling Capabilities

This paper has demonstrated that improvements in *public* hassling capabilities (α) can produce a deterrence failure. Can improvements in *private* hassling capabilities (θ) ever produce worse outcomes for D? Observation 1 speaks to this question.

Observation 1: *For a fixed $\alpha \in \mathcal{A}$, if D experiences an improvement in private hassling capabilities (i.e. moving from $\underline{\theta}$ to $\bar{\theta}$), then D attains a weakly greater utility.*

Proof: Follows from Proposition 1.

Observation 1 suggests that improvements in private capabilities are theoretically distinct from improvements in public capabilities, as only improvements in public capabilities can produce deterrence failures. This result is consistent with Fey and Ramsay (2011). Of course, the take-away from this section is not that states should only make improvements in their private hassling capabilities. Practically speaking, it is not necessarily clear that private hassling capabilities could not become public capabilities through leaks or espionage. Thus, it is possible that an improvement in a private hassling capability becomes public, and this now public improvement leads to a deterrence failure. Importantly, Observation 1 holds for every model extension examined in the Online Appendix.

6.1.2 Deterrence Failure and Welfare

If an improvement from $\underline{\alpha}$ to $\bar{\alpha}$ produces a deterrence failure, what can be said about the aggregate welfare? Observation 2 speaks to this point.

Observation 2: *If improvements in publicly observed hassling capabilities (i.e. moving from* $\underline{\alpha}$ *to* $\bar{\alpha}$ *with* $\underline{\alpha} < \bar{\alpha}$*) produces a deterrence failure, then it produces a welfare loss.*

Proof: See Online Appendix.

This observation adds nuance to deterrence failures. When a deterrence failure occurs following improvements in public hassling capabilities, it is not just that utility is transferred from D to A. Rather, when the Emboldening or Predictability Conditions hold, actors do in aggregate worse following improvements in hassling capabilities. Importantly, this is not to say that improvements in hassling capabilities always produce a welfare loss – I show in the Appendix, under some conditions, improvements in hassling capabilities can produce a welfare improvement when the improvement deters A from declaring war (when $Q(\underline{\alpha}) < 0$ and $Q(\bar{\alpha}) \geq 0$).

6.1.3 Linkages Between Hassling and Wartime Capabilities

What if improved hassling capabilies also affects wartime capabilities? Observation 3 speaks to this question.

Observation 3: *If improvements in publicly observed hassling capabilities also improve wartime payoffs, then the improvements in public hassling capabilities can never produce a deterrence failure.*

Proof: Follows from Proposition 1.

In the two cases in Proposition 1, type $\underline{\theta}$ D's attained their initial wartime payoff $\rho - \omega_D$. If a shift from $\underline{\alpha}$ to $\bar{\alpha}$ also resulted in D attaining a greater wartime payoff, then type $\underline{\theta}$ D's would always do better following the change, thereby undermining the conditions for deterrence failure. It is worthwhile mentioning that it is not obvious that an improvement in hassling capabilities would always produce improvements in wartime outcomes. For example, if increased investments in precision strike capabilities shifted funds in such a way that it undermined the army's ability to conduct a counterinsurgency, then improvements in hassling could lead to worse wartime outcomes depending upon what the war outcome entailed. At a minimum, this point suggests a possibly useful heuristic for investing in hassling capabilities: so long that improvements in hassling are made that also improve wartime outcomes, then these improvements will not produce a deterrence failure.

6.2 Extensions

Above I found that if and only if the "Predictability Conditions" and "Emboldening Conditions" held, there could be a deterrence failure following improvements in public hassling capabilities. How general are these results? The "Predictability Conditions" and "Emboldening Conditions" are specific both to this model and these assumptions; however, it is worth exploring how central the concepts of predictability and emboldening are to creating the outcome of D doing worse following improvements in hassling capabilities.[31] Overall, I find, across a wide range of alternate assumptions and models, that deterrence failure only arises through A becoming emboldened or D becoming more predictable. I describe my results below, then present the full results in the Online Appendix. All sections referenced below are those in the Online Appendix.

In the model above, I did not have A incur costs from investing in the rising technology. In Section 5, I examine a model where this occurs and I find, once again, that deterrence failure only arises from D becoming more predictable or A becoming emboldened.

[31] While the specifics of the conditions for when these occurs are model specific, generally, when D becomes more predictable, it implies that high-types ($\theta > \underline{\theta}$) play the game more like low-types ($\theta = \underline{\theta}$), resulting in high types attaining a lower utility despite not going to war more. And, when A becomes emboldened, it implies that A selects a level of rising technology resulting in more type D's going to war and with the types of D that do not go to war attain lower utilities.

In the model above, I assume that in equilibrium the final realized P function and offer x have the properties $P \in (p, 1)$ and $x^* \in (0, 1)$. In Section 7, I examine a model that utilizes a different peace and wartime payoff structure where there are no such kinks or bounds and I find substantively similar results. Additionally, these results persist outside of this specific modeling technology, as I show in the general analysis in Section 8.

Additionally, the model above is a two-type model. In a model in Section 7 and the general analysis in Section 8, I examine models with a continuum of types, which produces substantively similar results.

Finally, the model above utilizes a specific game structure to bargaining after the rising technology and hassling comes to fruition. In Section 7 and in the general analysis in Section 8, I consider both a more flexible structure that allows for a wide range of bargaining structures, or I use mechanism design to illustrate key intuitions from equilibria with almost no structure to the bargaining protocols. Overall, I show in the most general sense, predictability and emboldening following improvements in public hassling capabilities create a deterrence failure.

7 Conclusion

This paper explores a common occurrence in international affairs. One state considers undertaking revisionist activities, knowing that its rivals may respond with a range of policy levers. How do the rival state's capabilities affect how the revisionist state will behave? While I emphasize this question in the context of a revisionist state building nuclear weapons as this topic has and likely will continue to have policy relevance, this can be explored for any number of possible revisionist moves, ranging from building new weapons systems to seizing strategically valuable territory. Despite the widespread occurrence of this scenario, this is the first paper to formally examine it.

To understand this interaction, I present a new game theoretic model where an endogenously rising power faces a non-rising power who can engage at multiple conflict escalation levels and who has private information about their willingness to engage at different levels. I find that when the non-rising power is better at low-level conflict, it can, though multiple mechanisms, lead to overall worse outcomes for the non-rising power. These results arise when improved low-level conflict capabilities negatively interact with the non-rising state's ability to effectively use their private information.

The results here suggest that political scientists and policymakers need to take a harder look at low-level conflict capabilities. As I show here, 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, it does suggest that having more tools can, under some conditions, lead to systemically worse outcomes. More research is needed on this topic.

I highlight two directions for future research. First, this paper only considered a single type of information asymmetry regarding the costs of hassling. Just as models of war and peace consider multiple types of information asymmetry (Fey and Ramsay, 2011), future research of low-level conflict must also consider more types of uncertainty. Second, while this paper does disaggregate low-level conflict from war, this paper adopts a reductionist form of war (as is consistent with the existing literature). In practice, the formal framework here could be extended to examine how technological developments for different types of war (i.e. wars that emphasize one domain over another) or different conflict theaters affect ultimate outcomes.

References

Anderson, J. W. and Khan, K. (1998) Pakistan politicians' rallying cry: 'let them eat grass', *The Washington Post*. 12

Bas, M. A. and Coe, A. J. (2012) Arms diffusion and war, *Journal of Conflict Resolution*, **56**, 651–674. 1.1

Bas, M. A. and Coe, A. J. (2016) A dynamic theory of nuclear proliferation and preventive war, *International Organization*, **70**, 655–685. (document), 4, 1.1, 1.3, 5.1, 29

Battle, J. (2009) *Saddam Hussein Talks to the FBI: Twenty Interviews and Five Conversations with "High Value Detainee# 1" in 2004*, National Security Archive. 5.1

Benson, B. and Smith, B. C. (2020) Do alliances cause commitment problems and war?, *Working Paper*. 5, 1.1

Benson, B. V., Meirowitz, A. and Ramsay, K. W. (2016) Changing capabilities, uncertainty, and the risk of war in crisis bargaining, *Research & Politics*, **3**, 2053168016657687. 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), 5.1

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. 30

Debs, A. and Monteiro, N. P. (2014) Known unknowns: power shifts, uncertainty, and war, *International Organization*, **68**, 1–31. (document), 4, 1.1, 1.3

DoD (2018) Summary of the 2018 national defense strategy of the united states of america, Tech. rep., Department of Defense Washington United States. 5.2

Fearon, J. D. (1995) Rationalist explanations for war, *International organization*, **49**, 379–414. (document), 1.1, 1.3, 2.1, 17

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

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. 1.1, 1.3, 6.1.1, 7

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. *et al.* (2005) *Strategies of containment: a critical appraisal of American national security policy during the Cold War*, Oxford University Press. 5.2

Gartzke, E. and Lindsay, J. R. (2017) Thermonuclear cyberwar, *Journal of cybersecurity*, **3**, 37–48. 1.1, 1.3

Jervis, R. (1984) The illogic of american nuclear strategy. 5.2

Joseph, M. (2020) The secret failure of nuclear containment, *Working Paper*. (document), 1.1, 1.2, 1.3, 5.1

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. 5.1

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

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

Levy, J. S. (1987) Declining power and the preventive motivation for war, *World Politics*, **40**, 82–107. 1.3

Levy, J. S. (2008) Preventive war and democratic politics: Presidential address to the international studies association march 1, 2007, chicago, *International Studies Quarterly*, **52**, 1–24. 1.3

Levy, J. S. (2011) Preventive war: Concept and propositions, *International Interactions*, **37**, 87–96. 7

Lindsay, J., Gartzke, E., Greenhill, I. K. M. and Krause, P. J. (2016) Coercion through cyberspace: The stability-instability paradox revisited. 5.2

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. 6, 1.1, 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

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

O'Brien, K. A. (1995) Interfering with civil society: Cia and kgb covert political action during the cold war, *International Journal of Intelligence and Counter Intelligence*, **8**, 431–456. 5.2

Powell, R. (2006) War as a commitment problem, *International organization*, **60**, 169–203. (document), 1.1, 1.3

Powell, R. (2015) Nuclear brinkmanship, limited war, and military power, *International Organization*, **69**, 589–626. (document), 1.3, 26

Ramsay, K. W. (2017) Information, uncertainty, and war, *Annual Review of Political Science*, **20**, 505–527. 13

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

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

Schram, P. (2020) Hassling: How states prevent preventive wars, *Working Paper*. (document), 5, 1.1, 7, 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. 1.3

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

Snider, L. B. (2015) *The Agency and the Hill*, Government Printing Office. 5.2

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. 5.2

Spaniel, W. (2019) *Bargaining over the bomb: the successes and failures of nuclear negotiations*, Cambridge University Press. (document), 1.1, 1.3

Spaniel, W. and Bils, P. (2018) Slow to learn: Bargaining, uncertainty, and the calculus of conquest, *Journal of Conflict Resolution*, **62**, 774–796. 1.3

Spaniel, W. and Malone, I. (2019) The uncertainty trade-off: Reexamining opportunity costs and war, *International Studies Quarterly*, **63**, 1025–1034. 1.3, 21

Trenin, D. (2018) Despite the helsinki summit, the hybrid war is here to stay, *Carnegie Moscow Center*. 1.1

U.S. Air Force (2019a) Agm-86b/c/d missiles, Tech. rep. 10, 5.1

U.S. Air Force (2019b) B-1b lancer, Tech. rep. 10, 5.1

U.S. Air Force (2019c) B-52h stratofortress, Tech. rep. 10, 5.1

Votel, J. L., Cleveland, C. T., Connett, C. T. and Irwin, W. (2016) Unconventional warfare in the gray zone, *Joint Forces Quarterly*, **80**. 1.1

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. 27

Woods, K. M., Pease, M. R., Stout, M. E., Murray, W. and Lacey, J. G. (2006) Iraqi perspectives project. a view of operation iraqi freedom from saddam's senior leadership, Tech. rep., United States Joint Forces Command, Norfolk, VA. 27, 5.1