Conflicts that Leave Something to Chance: Establishing Brinkmanship Through Conventional Wars

Peter Schram*

April 14, 2023

Abstract

I present a formal model of nuclear brinkmanship. The novelty here is that the threat of nuclear escalation comes about inadvertently, making it a function of conflict duration with a non-monotonic relationship to conventional capabilities. When two states have similar conventional capabilities and enter into a war, the conventional conflict will be prolonged, thus resulting in high degrees of nuclear risk. When two states have dissimilar conventional capabilities and enter into a war, the conventional conflict will be one-sided and short, thus resulting in low degrees of nuclear risk. The model generates a series of results, including evidence of the nuclear peace and of conflict behavior consistent with the stability-instability paradox. Additionally, the model generates mixed results on the feasibility of using strategic nuclear instability as a substitute for conventional arming.

*Assistant Professor, Department of Political Science, Vanderbilt University. peter.schram@vanderbilt.edu
Thanks to James Fearon, Brad Smith, Peter Bils, Brenton Kenkel, Jennifer Larson, Andrew Coe, Ethan Bueno de Mesquita, James Morrow, Andrew Kydd, Robert Gullotti, Erik Gartzke, Thomas Schere, Scott Tyson, Jessica Sun, and seminar participants at the Center for Peace and Security Studies and the Midwest Political Science Association Annual Conference.

"Discussions of troop requirements and weaponry for NATO have been much concerned with the battlefield consequences of different troop strengths and nuclear doctrines. But the battlefield criterion is only one criterion, and when nuclear weapons are introduced it is secondary. The idea that European armament should be designed for resisting Soviet invasion, and is to be judged solely by its ability to contain an attack, is based on the notion that limited war is a tactical operation. It is not. What that notion overlooks is that a main consequence of limited war, and potentially a main purpose for engaging in it, is to raise the risk of larger war."
-Schelling, Thomas C. Arms and Influence

1 Introduction

The defining feature of international politics since 1945 has been the absence of direct great power conflict (Gaddis, 1986). To explain this historical anomaly, some neorealists classify this "long peace" as the "nuclear peace," where the fear of a nuclear exchange prevents significant conflict among great powers (Waltz, 1981; De Mesquita and Riker, 1982; Mearsheimer *et al.*, 2001; Waltz, 1981).[1] How the nuclear peace functions in practice is subtle. It is not as if states can credibly deter revisionist behavior through the threat of a nuclear first strike. After all, outside of circumstances where a state faces an existential threat, no state would ever launch a significant nuclear strike against a capable nuclear opponent, as doing so would be tantamount to suicide. And, the existence of nuclear weapons does not prevent states from fighting conventional wars. In theory, states with nuclear weapons could forgo these weapons and still fight conventional conflicts with one another, just as they did before the advent of nuclear weapons. Instead, what preserves the nuclear peace is the threat of unintended escalation (Schelling, 1980, 1966; Powell, 2015). In the nuclear era, when states enter into a conventional conflict, that conventional conflict is unstable and could turn into a massive nuclear exchange. Proponents of the nuclear peace suggest that the near absence of large-scale, direct conflict between nuclear states stems from a new and frightening feature of conventional conflicts: they could accidentally spiral out of control (Snyder, 1965; Jervis, 2017; Schelling, 1966).[2]

[1]See Gaddis (1986) footnote 67.

[2]See Kydd (2019) for a recent summary of this debate, as well as an analysis of the expected value of the proliferation of nuclear weapons.

Schelling's insight, as expressed in the quote starting the paper, is that we must consider linkages between aspects of conventional conflict—like troop placement and force posture—and nuclear risk. To begin unpacking this relationship, research has identified several mechanisms for how conventional conflicts could turn into nuclear exchanges. Accidents, decentralized decision making, or inadvertent escalation—all of which are discussed more in Section 2—could transform a conventional conflict into a nuclear one (Perrow, 2011; Sagan, 1994, 2020; Posen, 2014). Critically, across all these mechanisms linking conventional conflict to a nuclear exchange, there is a common underlying factor: time. When international conflicts between nuclear powers are short and decisive, there are fewer opportunities for an unintended escalation leading to a massive nuclear exchange. As conventional conflicts become protracted affairs, there are more opportunities for miscalculations or missteps. Adapting the Schelling (1966) metaphor, if a conventional conflict in the nuclear era is a war of nerves similar to "rocking the boat," then the shorter the time spent rocking, the less likely actors are to get soaked.

The positive relationship between conflict duration and nuclear risk presents a new dimension on what force posture means in the nuclear deterrence setting. Manipulating the conventional force posture can change how protracted a conflict is, which, between two nuclear opponents, changes the likelihood that a conventional conflict turns into a nuclear exchange. This was not an abstract notion. During the Cold War, deterring Soviet aggression against West Berlin did not depend on winning a conventional war; rather, deterrence required ensuring West Berlin had a sufficient conventional force posture to keep any conflict from ending decisively without any escalation risk. In this case, in Western Germany, NATO forces did not seek to win decisively, but rather use conventional forces to draw out a conflict and generate more nuclear risk. In contrast, in 1956, the Soviet Union committed a significant military force to quickly crush the Hungarian Revolution rather than risk letting the conflict spread and escalate. In this case, in Hungary, the Soviet Union committed a considerable force to ensure the conflict ends decisively.

That additional arming could lead to more or less nuclear risk has fundamental implications for how troop placement can deter, yet this has not been formally considered. Overall, this means our understanding of how conventional arming, the placement of bases, and troop deployments are used in nuclear deterrence rests on loose theoretical footing. This paper seeks to offer a more robust grounding.

I present a new model of nuclear deterrence through conventional arming. In the deterrence game, there are two actors, a defender and a challenger, where the defender's resolve is unknown to the challenger.[3] In the game, the defender begins by selecting a level of conventional arms.[4] Arming has several effects: (1) the defender incurs costs from arming, (2) the defender is more likely to win the conventional war at higher arming levels, and (3) should a conventional war occur, the likelihood of a nuclear war is non-monotonic (specifically, increasing then decreasing) in the level of conventional arming, depending on whether additional arms will draw a conflict out or make a conflict more decisive. Next, the challenger observes this level of arming, and chooses whether to challenge or not. Finally, if challenged, the defender can acquiesce and back out of the challenge, or the defender can fight the challenger. When the defender fights the challenger, the actors fight a conventional war where there is a chance of a nuclear exchange.

The model produces a wide range of behavior, where conventional arming can serve as a bluff, a threat, a signal, or a means to perform well in a forthcoming war. And, we establish a series of results that offer a confirmation of past theories of nuclear deterrence, and that offer some (to the best of my knowledge) new perspectives. I establish a series of findings, and highlight three of these here.

First, consistent with the proposed "nuclear peace," as conventional conflicts start carrying the risk of a nuclear exchange, I find that actors will avoid engaging in conventional conflict. This result is re-assuring; while conventional arming here carries additional technical complications—that conventional arming can increase or decrease nuclear risk—I still find that nuclear instability drives actors to engage in less war.

Second, I find that when actors bear greater costs of nuclear war or greater nuclear instability, actors will (generally) arm with the intent of keeping conflicts less protracted. These findings share some features of the "stability-instability paradox" discussed in (Snyder, 1965) and formalized in (Powell, 2015). Consistent with this research, I find that increasing nuclear instability and costs may result in less violent conflicts with lower force postures. However, I also find the opposite to be true: when faced with higher nuclear costs and instability, actors

[3]The structure here has similarities to many other deterrence models, including Fearon (1997), Powell (2015), Gurantz and Hirsch (2017), Baliga *et al.* (2020) Di Lonardo and Tyson (2022) as a few examples. See Huth (1999) and Ramsay (2017) for reviews.

[4]Manipulating p is analogous to deploying troops in preparation for a conflict.

may arm more aggressively, with the intent of conducting a fast, one-sided conflict. Together, this paper suggests it is not just that actors conduct conflicts with diminished force postures in the nuclear era, but rather they will (at times) conduct conflicts with more aggressive force postures, with the intent of quickly ending the conflict. These latter findings can offer insight into Soviet actions during the Hungarian Revolution (1956).

Third, the model also identifies new insights into the use of conventional arming in the nuclear era. In contrast to the logic of "asymmetric response" strategies, it is not possible for defenders to use nuclear instability as a substitute for conventional arming in all cases. Because added nuclear instability dissuades both challengers *and defenders* from engaging in conventional conflict, in an environment with elevated nuclear instability, defenders must invest more in conventional armaments in order to be willing to fight when challenged. In other words, in a world with nuclear risk, a credible deterrent threat might require additional arming on the part of the defender. This also means that adding nuclear risk can have detrimental effects on welfare because it may require parties to invest more in costly armaments to establish a credible deterrent threat.

This paper is most similar to the incomplete information model in Powell (2015) which also considers nuclear brinkmanship multiple levels of conflict. There are several differences, two of which I highlight here. First, these papers differ in how nuclear risk is generated. In this paper, nuclear risk is manipulated indirectly. The defender selects a level of arming that, among other effects, determines the expected conflict duration, which in turn determines the likelihood of a nuclear escalation. In Powell, the defender is able to directly manipulate nuclear risk, where such manipulation is public, credible, flexible, and will not alter the defender's likelihood of winning in a conventional conflict. Because these mechanisms for generating nuclear risk are different, deterrence operates differently across the models (as I discuss more in 7.6). To summarize, I find that it is possible to peacefully signal intent and deter a a challenger, while Powell finds there cannot be deterrence without some likelihood of conflict. Second, these papers differ in how conventional forces translate into nuclear risk. Powell also considers a challenger who selects a conventional force posture, where additional forces *always* results in greater nuclear risk. This paper assumes a more flexibility than Powell; here, adding conventional forces could generate more or less nuclear risk, depending on whether it makes the conflict more or less decisive.

This paper is also distinct from other signalling models with conflict. The perspective that arming or troop placement could reduce nuclear risk—when it leads to more decisive conflicts—or increase nuclear risk—when it drags out conflict—is different from other treatments in costly signalling games. Similar to work like Slantchev (2005) and Slantchev (2011), additional arming could be a productive costly signal, where additional arming leads to less nuclear risk and better expected wartime outcomes for the arming state. Or, additional arming could lead to more nuclear risk and worse expected wartime outcomes, thus making it a "handicap signal," as explored in Reich (2022). While existing research has explored these topics separately, there does not exist a unified, theoretical grounding for how conventional forces can function as a deterrent threat when arming could be productive or handicapping.

While this paper will present and discuss the model in the context of conventional war and nuclear brinkmanship, the formalization below can apply to other cases where actors engage in one level of conflict and there is the possibility of a costly and painful escalation. The "challenger" and "defender" below could be gangs or rival drug cartels that, when they engage in a protracted and bloody fight, run the risk of the government intervening. A similar logic could hold for low-levels of intrastate conflict that run the risk of a third-party actor intervention that is worse for both parties.

2 On Arming and Nuclear Risk

While much of the model below is fairly standard in deterrence games, what is somewhat different here is that conventional conflict could probabilistically escalate into a nuclear exchange. What is unique here is that I assume a non-monotonic, increasing-then-decreasing relationship between conventional force levels and nuclear risk. I make this assumption based on past scholarship that I describe in more detail below. To summarize, the non-monotonic structure is based on three relationships, which relate conventional arming to nuclear risk. First, when the defender increases their conventional force posture, depending on the challenger's capabilities, this could result in greater conventional military parity between the challenger and defender or less military parity. Second, should a conventional conflict arise, more conventional force parity between disputants will result in a more protracted conventional conflict. Third, longer conventional conflicts will generate greater risks of nuclear exchange.

How I model nuclear risk breaks with how Powell (2015) treats nuclear risk, where adding conventional forces to a conflict always results in greater nuclear risks within the conflict. Of course, for whatever reason, readers might take issue with my assumptions on this relationship; this is where the flexibility of this modeling effort is useful. While the model accommodates the increasing-then-decreasing relationship between conventional arming levels and nuclear risk, the results below still hold if the relationship between conventional arming and nuclear risk was *only* increasing or *only* decreasing.[5] This flexibility allows the model to describe a broader set of possible cases, meaning our results can apply to many different kinds of relationships between arming and nuclear risk.

Adding conventional forces could result in more or less military parity. The first relationship is mechanical. Expanding conventional forces or building out military bases will either lead to greater military parity with a challenger—when the defender's capabilities approaches the challenger's capabilities—or less military parity with an opponent—when the defender's capabilities surpass the challenger's capabilities.

Greater military parity between actors results in longer conflicts. As intuition, if there is a low degree of military parity, then a quick, one-sided war or a rapid surrender is more plausible. On the other hand, if militaries are more evenly matched, then neither side has an immediate reason to stop fighting. Furthermore, because evenly-matched militaries will trade battle victories and defeats, war between equally matched adversaries will be less informative or less clearly decisive, which incentivises adversaries to continue fighting. This logic is illustrated in a series of theoretical models (Smith, 1998; Filson and Werner, 2002, 2004; Langlois and Langlois, 2009, 2012; Slantchev, 2004), with the empirical literature echoing this theoretical result (Bennett and Stam, 1996, 2009; Slantchev, 2004; Krustev, 2006; Chiba and Johnson, 2019).[6]

[5]This can be accommodated by shifting the p_0 and p_1 parameters to only consider regions where the relationship is monotonic.

[6]This result does not always hold in a statistical sense. Bueno de Mesquita *et al.* (2004) finds this relationship within democratic dyads, but for non-democratic dyads, no statistically significant relationship exists. Similarly,

Longer conflicts generate a greater likelihood of a nuclear exchange. Research has identified several mechanisms for how conventional conflicts could escalate to a nuclear exchange. This could come about entirely through accident. Within any complex system—like missile detection or early warning systems—system failures are possible (Sagan, 1994, 2020; Sagan and Waltz, 2003; Perrow, 2011). When states are at war, there is heightened risk that a faulty signal could be interpreted as an act requiring a nuclear response (Sagan, 1994). The possibility of a nuclear exchange could also come about through the course of conventional operations spiraling into a cycle of escalation. Whether through mechanical error (a malfunctioning GPS), human error (mis-reading maps), agency problems, or the fog of war, sometimes soldiers or operators take actions beyond what a fully rational, unitary decision maker would want, which could lead to a need for escalation (Sagan, 1994; Posen, 2014). Or, in a protracted conventional war, states may target their opponent's communication or command and control infrastructures, which could inadvertently undermine the targeted state's credible second-strike capability and necessitate an escalation (Posen, 2014).[7] Across all these different ways a conventional conflict could turn nuclear, time is a common underlying factor. When international conflicts between nuclear powers are short and decisive, there are fewer chances or reasons for system failures, overambitious operations, or the targeting of command and control infrastructure. As conventional conflicts drag on, the likelihood of error increases.

3 Game Form and Assumptions

Two players, a challenger (C) and a defender (D), are in a deterrence game with incomplete information. The game order is as follows.

1. Nature designates D's resolve $v_D \in \{\underline{v}_D, \bar{v}_D\}$ with $0 < \underline{v}_D < \bar{v}_D$. Let π denote the probability that D is type $\bar{v}_D$. D knows D's type, but C does not.

2. D selects a conventional force level that determines $p \in [p_0, p_1]$, which is D's likelihood of

Koch (2009) and Shannon *et al.* (2010) do not identify a statistically significant relationship. To the best of my knowledge, no existing research identifies a negative statistically significant relationship between military parity and conflict duration.

[7] Through the logic of mutually assured destruction, an actor on the verge of losing their second-strike capability might undertake dramatic, escalatory steps in an attempt to degrade their opponent's first-strike capability and thus preserve their second-strike capability.

winning in a conventional conflict. I assume $0 < p_0 < p_1 < 1$.

3. C selects whether to challenge or not. If C does not challenge, the game ends with C receiving payoff 0 and D receiving payoff $v_D - K(p)$, where $K : \mathbb{R}_+ \to \mathbb{R}_+$ is D's costs from the conventional force level. I assume $K(p_0) = 0$, and K is continuous and increasing in p. If C does challenge, the game moves to the next stage.

4. D selects whether to acquiesce or escalate to conflict. If D acquiesces, C receives payoff v_C and D receives payoff $-K(p)$. If D escalates to conflict, then both states receive their conflict payoffs (described below).

Conflict is a stochastic process that could end with a nuclear exchange or with a conventional victory. I let $n \geq 0$ denote the hazard rate for the termination of the conflict through a nuclear exchange. Essentially, n represents the "nuclear instability" to a conflict; n is zero when there is no possibility of a nuclear exchange, and n takes on greater values when C or D are more accident prone (Sagan, 1994) or are fighting near critical nuclear infrastructure (Posen, 2014). I let $\frac{\alpha}{p(1-p)}$ denote the hazard rate for the termination of conflict through conventional means. The choice variable p was defined above; for conflicts between lopsided adversaries ($p \approx 0$ or $p \approx 1$) the hazard rate is large, which is consistent with uneven conventional conflicts ending quickly (Slantchev, 2004; Bennett and Stam, 2009). $\alpha > 0$ represents a scaling parameter that grants more flexibility to the conventional conflict hazard rate.[8] Together, this means that $h(p) = n + \frac{\alpha}{p(1-p)}$ is the hazard rate for conflict ending, $n/h(p)$ is the likelihood that conflict ends in a nuclear exchange, $\frac{\alpha}{h(p)p(1-p)}$ is the likelihood that conflict ends conventionally, and $\frac{1}{h(p)}$ is the expected time to conflict termination. It is worthwhile highlighting here that in the main model I am not treating conventional conflict as a continuous-time, war-of-attrition game where states can choose to drop out at any time, but rather treating conflict as a reduced-form process.

When the conflict ends, players receive their payoffs. When the conflict ends with a nuclear exchange, $-N_D \in \mathbb{R}$ and $-N_C \in \mathbb{R}$ denote D's and C's expected payoffs from a nuclear exchange. When the conflict ends conventionally, D wins with probability p, and C wins the

[8] Without α, the conventional conflict hazard rate $\frac{1}{p(1-p)}$ is always greater than or equal to 4, which limits how costs are accrued (as discussed below).

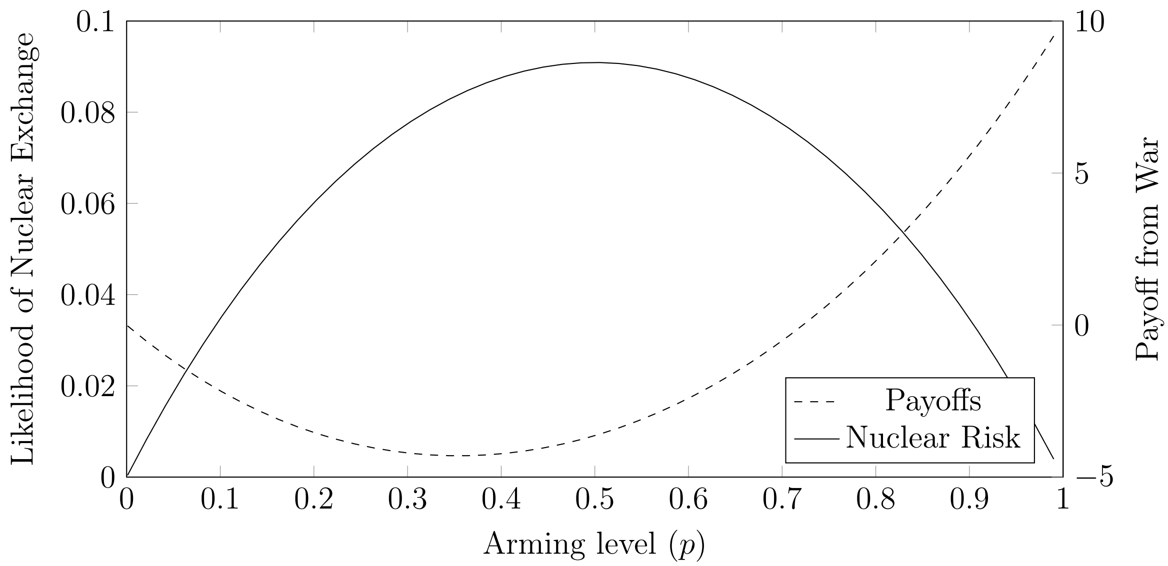

Figure 1: Nuclear risk and payoffs

asset (of value $v_C \geq 0$ to C) with probability $1-p$. Regardless how conflict ends, by fighting, actors accrue conventional conflict costs at rate $c_D \geq 0$ and $c_C \geq 0$.

C's expected utility from conflict is

$$\frac{n}{h(p)} * (-N_C) + \frac{\alpha}{h(p)p(1-p)}((1-p)v_C) - \frac{c_C}{h(p)},$$

and D's expected utility from conflict is [9]

$$\frac{n}{h(p)} * (-N_D) + \frac{\alpha}{h(p)p(1-p)}(pv_D) - \frac{c_D}{h(p)} - K(p).$$

I visualize the likelihood of a nuclear exchange and D's expected utility without arming costs $K(p)$ from a conflict for a range of possible p's under one set of parameters in Figure 1.

First, consider the likelihood of nuclear exchange, which is the solid line in Figure 1. For small

[9]Or, for C and D (respectively), using $h(p) = \frac{\alpha+np(1-p)}{p(1-p)}$,

$$\frac{np(1-p)}{\alpha+np(1-p)} * (-N_C) + \frac{\alpha}{\alpha+np(1-p)}((1-p)v_C) - \frac{c_C p(1-p)}{\alpha+np(1-p)}$$

$$-\frac{np(1-p)}{\alpha+np(1-p)}N_D + \frac{\alpha}{\alpha+np(1-p)}(pv_D) - \frac{c_D p(1-p)}{\alpha+np(1-p)} - K(p)$$

or large conventional arming levels ($p \approx 0$ and $p \approx 1$), $h(p)$ becomes large and $h(p)p(1-p)$ becomes small; thus, when the conventional arming level leads to an unbalanced or one-sided conventional conflict ($p \approx 0$ and $p \approx 1$), there is little risk of a nuclear exchange ($\frac{n}{h(p)}$ is smaller) and there is a greater likelihood of the conflict ending conventionally ($\frac{\alpha}{h(p)p(1-p)}$ is greater). In contrast, in a more-balanced conventional conflict ($p \approx \frac{1}{2}$), there is greater risk of nuclear exchange and a (relatively) lower likelihood the game ends with a conventional victory or defeat.

Now consider D's payoffs from conflict, which are captured in the dashed line in Figure 1. As p increases from 0 to roughly 0.4, the conflict becomes more protracted and the increasing risks of a nuclear exchange results in D's decreasing. For p values greater than 0.4, the defender's benefits from becoming more likely to win the conflict and the nuclear risk increasing more slowly and eventually decreasing makes D's utility increasing. Note that it is not always the case that D's best option is to select the greatest possible arming level: recall that I am not plotting D's costs from arming $K(p)$ in this figure.

4 Modeling Issues

This paper is most similar to the incomplete-information model in Powell (2015), but most clearly differs in the treatment of nuclear risk. In this paper, outside of exogenous parameters, nuclear risk is fully determined through the defender's arming level.[10] In the Powell formalization, after the challenger selects a conventional arming level, the defender is able to publicly and credibly manipulate the level of nuclear risk within a conventional war without altering their likelihood of winning in the conventional war.[11] A natural interpretation of the defender's choice in the Powell model would be manipulating Defense Readiness Condition (DEFCON) levels while in a conflict, which, through organizational and technical channels, would alter levels of nuclear risk (Sagan, 1985).

While Powell is groundbreaking in what it does, the model there may not apply to all settings.

[10] This is represented in the $\frac{n}{h(p)}$ function.

[11] Formally, after a challenger selects some level of arming p, the defender selects function $r(p)$ from a set of allowed functions. p denotes the likelihood the C wins in a conventional conflict, and $r(p)$ denotes the likelihood the conventional conflict escalates to a nuclear exchange.

First, Powell assumes it is possible to publicly and credibly manipulate nuclear risk within a crisis. However, practically, doing so is subject to "cheap-talk" concerns. In Powell, in many cases, a defender would want to signal that they have implemented a high-risk (of nuclear exchange) system when in fact they have implemented a low-risk system. As a second limitation, Powell assumes it is possible to exclusively manipulate nuclear risk without altering the balance of conventional power. This feature means that Powell's model is outside of most standard discussions of tripwires (see Schelling (1966)); for example, the positioning of conventional forces in Western Berlin could be used to generate nuclear risk, but it would also mechanically alter the likelihood of conventional conflict success.[12] Third, in the Powell model, taking the defender's choice as fixed, there is a positive monotonic relationship between conventional force levels the challenger brings and the final probability of a nuclear exchange. In this regard, Powell is not considering the possibility that arming to a level that produces a decisive victory could ever reduce the probability of a nuclear exchange.[13]

Additionally, because nuclear risk is generated differently across the models here and in Powell (2015), signalling resolve and deterrence function differently. This is discussed in Section 7.6.

This paper also has benefited from decades of iterations of models of nuclear deterrence (Schelling, 1980; Nalebuff, 1988; Zagare and Kilgour, 1993; Powell, 1989, 2003). Because this is not a review article I will not cite the entire set of this research on nuclear deterrence, but rather refer readers to several excellent reviews, including see Jervis (1979), Huth (1999), Quackenbush (2011) and Gartzke and Kroenig (2016). Additionally, the model integrates features from the formal literature on endogenous power shifts and deterrence (Fearon, 1997; Schultz, 2010; Tarar, 2016; Debs and Monteiro, 2014; Gurantz and Hirsch, 2017).

Of course, nearly all models cited above only consider two types of outcomes: war or peace. A new branch of research considers conflict that can be more multifaceted. (Zagare and Kilgour,

[12] Powell (2015) discusses the Kremlin's decision whether or not to authorize the use of tactical nuclear weapons in Cuba as an example of manipulating $r(\cdot)$. However, using tactical nuclear weapons would undoubtedly alter the defender's likelihood of winning a conflict, which means that p would also be altered. Additionally, the Kremlin's decision was not known to the United States, meaning it was not publicly manipulated.

[13] I would be remiss to highlight two things Powell does quite well. First, while the Powell risk function is only monotonically increasing in the conventional arming level, this risk function is not assigned any functional form (like it is here) and is allowed to be quite flexible. Second, Powell considers an endogenously arming challenger while I treat the challenger's arming level as exogenous.

1998; Werner, 2000; Schultz, 2010; Slantchev, 2011; Powell, 2015; Tarar, 2016; McCormack and Pascoe, 2017; Coe, 2018; Spaniel, 2019; Joseph, 2020; Baliga *et al.*, 2020; Schram, 2021a,b). The key distinctions from existing work is that (a) I consider uncertainty over the defenders valuation of the good, and (b) to better capture the case of nuclear escalation, I consider a stochastic structure to when conflict escalates from lower-levels to higher-levels. To the best of my knowledge, outside of Powell (2015), no existing paper does both.

5 Complete Information Equilibrium

Before discussing the incomplete-information game, I first establish the equilibrium for a complete-information version of the game. This version of the game no longer has nature setting D's resolve as $v_D \in \{\underline{v}_D, \bar{v}_D\}$, but rather defines D's resolve as fixed v_D.

The key strategic tension in this game is as follows. C will challenge unless two conditions hold: (a) C knows that they will have to fight after challenging, and (b) C knows that fighting is sufficiently bad for them. For (a) to be met, D must be willing to fight, meaning that D has armed to a level where D does well enough in the conflict to be willing to incur its costs. This is cutpoint p^D,[14] which satisfies

$$p^D = 1 - \frac{\alpha v_D}{c_D + nN_D}.$$

For (b) to be met, D's level of arming is at or beyond some (alternate) cutpoint where, should D escalate, C does sufficiently bad in the ensuing conflict. This cutpoint is p^C,[15] which satisfies

$$p^C = \frac{\alpha v_C}{c_C + nN_C}.$$

If D arms to a level at or beyond both p^D and p^C, then C will be deterred.

Also note that if $p^D < p^C$, D could select some $p \in [p^D, p^C]$. When this is the case, C will

[14] Formally, p^D solves $0 \leq \frac{n}{h(p^D)} * (-N_D) + \frac{\alpha}{h(p^D)p^D(1-p^D)}\left(p^D v_D\right) - \frac{c_D}{h(p^D)}$

[15] Formally, p^C solves $0 = -\frac{n}{h(p^C)}N_C + \frac{\alpha}{h(p^C)p^C(1-p^C)}\left((1-p^C)v_C\right) - \frac{c}{h(p^C)}$.

challenge, D will escalate, and the game will end in war.[16] I can then define

$$\hat{p} \in argmax_{p\in[p^D,p^C]} \left\{ \frac{n}{h(p)} * (-N_D) + \frac{\alpha}{h(p)p(1-p)} (pv_D) - \frac{c}{h(p)} - K(p) \right\},$$

where $\hat{p}$ is D's optimal conventional force level when the game ends in war. Note that the set $\hat{p}$ may not be singleton, in which case I abuse notation and let $\hat{p}$ define the smallest element of that set. I define $U_D(\hat{p})$ as D's utility from selecting $\hat{p}$ as defined above.

For ease, I will assume for the complete information case that $p^C > p_0$, $p^D > p_0$, $p^C < p_1$, $p^D < p_1$. Together, these expressions imply that deterrence is non-trivial but still feasible when arming costs are low enough.

In equilibrium, D will select one of three arming values. First, D could select $p^* = max \left\{p^C, p^D\right\}$ which deters C from ever challenging and gives D a final utility of $V_D - K(p^*)$. Second, D could select $p = p_0$ and acquiesce when challenged, giving D a final utility of 0. Finally, it could be that deterring is too costly but fighting is more productive than peace: in this case, D will select some $p = \hat{p}$. Formally, the equilibrium play is as follows.

Case 1: Let $p^C \leq p^D$.
1A. If $V_D - k(p^D) \geq 0$, then D selects $p = p^D$ and C does not challenge.
1B. Otherwise, D selects $p = p_0$, C challenges, and D acquiesces.

Case 2: Let $p^D < p^C$.
2A. If $V_D - k(p^C) \geq 0$ and $V_D - k(p^C) \geq U_D(\hat{p})$, then D selects $p = p^C$ and C does not challenge.
2B. If $0 > V_D - k(p^C)$ and $0 > U_D(\hat{p})$, then D selects $p = p_0$, C challenges, and D acquiesces.
2C. Otherwise, D selects $p = \hat{p}$, C challenges, and D fights.

While the primary analysis occurs below, note that p^D is decreasing in v_D. This implies that as D has a greater resolve, D can deter through lower levels of arming in the complete information model. This is important because in the incomplete information version of the game, the opposite holds: moving from type $\underline{v}_D$ to type $\bar{v}_D$ *always* results in a greater level of arming. As I will discuss, this means that arming is inefficient in the game with incomplete information.

[16]As is common in deterrence models, war is possible in this complete information game when states both value the asset highly enough.

6 Incomplete Information Equilibria

In the incomplete information game, much of the strategic tension is similar to what is described in the complete information game. What's new here is that because C is uncertain of D's resolve, sometimes C is uncertain whether D is willing to fight or not. This in turn can shape arming decisions. In many equilibria, low-types and high-types will behave differently. For example, under some parameters, low-types will acquiesce and high-types will deter. Under different parameters, low-type D's can bluff by mimicking the arming behavior of high-type D's; this can result in C not challenging low-type D's even when low-type D's would not fight if challenged. Also, under an alternate set of parameters, sometimes high-type D's will signal their resolve by selecting levels of arming beyond what was needed in the complete-information game, with the intent of getting low-type D's to drop out and to have C be deterred.

I limit analysis to a perfect Bayesian Nash equilibria satisfying the intuitive criterion (Cho and Kreps, 1987). While multiple equilibria can still exist after applying the intuitive criterion—for example, when C or D is indifferent over their actions, a continuum of strategies could still be supported—the empirical implications of analyzing all the remaining multiple equilibria will be limited.[17] For the rest of this paper, I focus on a single equilibrium where there isn't mixing at the points where actors are indifferent.

Because this model takes a phenomenonalist approach (Paine and Tyson, 2020) where the model's setup corresponds with trade-offs that decision-makers face, I make few restrictive assumptions and the model can generate a large set of possible strategic behavior (deterring, bluffing, signaling, etc). For that reason, I would encourage readers to not get too bogged-down in every equilibrium statement in this section, especially because the next sections discussing general results across all equilibria are the better "punchlines" that should be considered. That being said, having a broad set of equilibrium behavior is a virtue of this paper: it is not difficult to find real-world examples of signaling, bluffing, deterring, or acquiescing within international

[17] For example, all equilibrium actions (i.e. the selected p^*, the decision to challenge, and the decision to fight) listed below could be supported by a range of possible off-path beliefs. Alternatively, multiple equilibria can arise when C or D has indifference over their set of actions. For example, under some parameters, high-type D's will be indifferent between selecting an arming level where they will drop out when challenged or arming up with the intent of fighting when they are challenged. This indifference (or others like it) can support a continuum of mixed strategy equilibria that I am not considering.

relations, so efforts made to further simplify the set of possible equilibria would detract from the generality of the analysis.

6.1 Characterizing Equilibrium Arming Levels

Before presenting the equilibria, I define several of the arming levels that are selected in equilibria, and I offer some intuition for their relevance. I derive all values in the Appendix.

As it was above, I define p^C as the conventional force level that would make C indifferent between challenging or not, conditional on D escalating in stage 4.

$$p^C = \frac{\alpha v_C}{c_C + nN_C}$$

I let $p^D(v_D)$ denote the conventional force level that would make a type $v_D \in \{\underline{v}_D, \bar{v}_D\}$ D indifferent between escalating or not once C has challenged.

$$p^D(v_D) = 1 - \frac{\alpha v_D}{c_D + nN_D}$$

Additionally, I define $\tilde{p}$ as the smallest feasible level of arming that would both (a) make type $\bar{v}_D$ D escalate when challenged and type $\underline{v}_D$ D acquiesce when challenged, and (b) make C at-least indifferent between challenging or not conditional on D's expected behavior. As intuition, if a $\tilde{p}$ exists, then there can exist a pooling equilibrium where both types of D select $\tilde{p}$ and C never challenges. To that end, because for arming level $\tilde{p}$ type $\underline{v}_D$ D's will not fight, here $\underline{v}_D$ D is bluffing. To define $\tilde{p}$, I first characterize the set $\tilde{P}$ as all values of p that satisfy conditions (b) and (a) in the order listed.

$$\tilde{P} = \left\{ p : \begin{cases} 0 \geq & \pi\left(-\frac{n}{h(\tilde{p})}N_C + \frac{\alpha}{h(\tilde{p})\tilde{p}(1-\tilde{p})}\left((1-\tilde{p})v_C\right) - \frac{c_C}{h(\tilde{p})}\right) + (1-\pi)v_C,\ and \\ p \in & \left[max\left\{p^D(\bar{v}_D), p_0\right\}, min\left\{p^D(\underline{v}_D), p_1\right\}\right] \end{cases} \right\}$$

Second, I define $\tilde{p}$ as $\tilde{p} = min\left\{\tilde{P}\right\}$ whenever $\tilde{P}$ is non-empty.

Next, I define $\hat{p}(v_D)$ as the level of arming level that is optimal for a type v_D D conditional

Symbol	Value
p^C	$p^C = \frac{\alpha v_C}{c_C + nN_C}$
$p^D(\bar{v}_D)$	$p^D(\bar{v}_D) = 1 - \frac{\alpha \bar{v}_D}{c_D + nN_D}$
$p^D(\underline{v}_D)$	$p^D(\underline{v}_D) = 1 - \frac{\alpha \underline{v}_D}{c_D + nN_D}$
$\tilde{p}$	$0 \geq \pi \left(-\frac{n}{h(\tilde{p})} N_C + \frac{\alpha}{h(\tilde{p})\tilde{p}(1-\tilde{p})} \left((1-\tilde{p}) v_C \right) - \frac{c_C}{h(\tilde{p})} \right) + (1-\pi) v_C$
$\hat{p}(v_D)$	$\hat{p}(v_d) \in \underset{p \in [max\{p^D(\bar{v}_D), p_0\}, min\{p^C, p_1\}]}{\arg\max} \left\{ \frac{(-n * N_D - c_D)}{h(p)} + \frac{\alpha}{h(p)p(1-p)} (p v_D) - K(p) \right\}$
$\bar{p}$	$\underline{v}_D - k(\bar{p}) = 0$

Table 1: D's selected arming levels.

on C challenging and D fighting.[18] I let $\hat{U}_D(p, v_D)$ denote a type v_D D's utility from selecting arming level p, C challenging, and D fighting.[19]

$$\hat{p}(v_D) \in \underset{p \in [max\{p^D(v_D), p_0\}, min\{p^C, p_1\}]}{\arg\max} \left\{ \hat{U}_D(p, v_D) \right\}$$

Note that the set $\left\{ p : \, p \in \left[max \left\{ p^D(v_D), p_0 \right\}, min \left\{ p^C, p_1 \right\} \right] \right\}$ may be empty, meaning no such $\hat{p}(v_D)$ exists.

Finally, I implicitly define $\bar{p}$ as the level of arming where a low-type D is indifferent between (a) arming to level $\bar{p}$, not being challenged, and attaining the asset, and (b) arming to level p_0, always being challenged, and acquiescing.

$$\underline{v}_D - K(\bar{p}) = 0.$$

I summarize the arming levels in Table 1.

For simplicity, I make the following parameter assumptions.

Parameter Assumptions: I assume $\bar{v}_D - K(max \left\{ p^D(\bar{v}_D), p_0 \right\}) > 0$ and $p^1 > p^D(\underline{v}_D)$.

The Parameter Assumptions imply that type $\bar{v}_D$ D's are able to fight over the asset (that arming level $p^D(\bar{v}_D)$ can be reached because $p^D(\bar{v}_D) < p_1$),[20] and that they are willing to arm to this level if it means that they attain the asset ($\bar{v}_D - K(max \left\{ p^D(\bar{v}_D), p_0 \right\}) > 0$). Essentially, if

[18]Formally, for both C and D to be willing to fight, it must be that D selects a p such that p is greater than $p^D(\bar{v}_D)$ and less than p^C, while also subject to the bounds that $p \in [p_0, p_1]$.

[19]Formally, $\hat{U}_D(p, v_D) = \frac{n}{h(p)} * (-N_D) + \frac{\alpha}{h(p)p(1-p)} (p v_D) - \frac{c_D}{h(p)} - K(p)$.

[20]Note that if $p^1 > p^D(\underline{v}_D)$ holds, $p^1 > p^D(\bar{v}_D)$ also holds.

Equilibrium Behavior	**Equilibrium Type**	**Type $\bar{v}_D$ arming**	**Type $\underline{v}_D$ arming**
Acquiesce-Acquiesce	Pooling	p_0	p_0
Deter-Bluff	Pooling	$max\left\{p^D(\bar{v}_D), \tilde{p}\right\}$	$max\left\{p^D(\bar{v}_D), \tilde{p}\right\}$
Deter-Deter	Pooling	$max\left\{p^D(\underline{v}_D), p^C\right\}$	$max\left\{p^D(\underline{v}_D), p^C\right\}$
Signal-Acquiesce	Separating	$\bar{p}$	p_0
Deter-Acquiesce	Separating	$max\left\{p^D(\bar{v}_D), p^C\right\}$	p_0
Fight-Acquiesce	Separating	$\hat{p}(\bar{v}_D)$	p_0
Fight-Fight	Separating	$\hat{p}(\bar{v}_D)$	$p^D(\underline{v}_D)$
Deter-Fight	Separating	p^C	$p^D(\underline{v}_D)$

Table 2: Equilibrium Summary. In the "Equilibrium Behavior" column, I list type $\bar{v}_D$'s behavior first and type $\underline{v}_D$'s behavior second. Note the values here assume $p_0 < p^D(\bar{v}_D)$.

these assumptions did not hold, then high-type D's would never be willing to deter C. I allow more flexibility for low-types. I also assume that type $\underline{v}_D$ D's are able to deter (that arming level $p^D(\underline{v}_D) < p_1$), but I do not assume that low-type D's are always willing to deter (I do not assume that $\underline{v}_D - K(max\left\{p^D(\underline{v}_D), p_0\right\}) > 0$).

6.2 Discussion

Table 2 presents the selected arming levels and whether or not war occurs. See the Appendix for a complete characterization. For the values in Table 2, I assume $p_0 < p^D(\bar{v}_D)$; I express the full equilibria without this assumption in the Appendix.

Within Table 2, I characterize eight kinds of equilibrium based on the strategic play of high- and low-types, as described in the first column. The second column denotes whether the equilibrium behavior is part of a pooling or separating equilibrium. The third and forth column denote the high-type's and low-type's arming levels.

The equilibrium behavior is standard to typical deterrence games. Regarding equilibrium behavior, either type D can "acquiesce," meaning they select the minimum arming level $p^* = p_0$ and acquiesce when challenged. Either type D can "deter," or select a level of arming that would keep C from challenging and where they would fight if challenged. To deter, high-types must select at least $p^* \geq max\left\{p^D(\bar{v}_D), p^C\right\}$, and low-types must select at least $p^* \geq max\left\{p^D(\underline{v}_D), p^C\right\}$. Either type D can "fight," or select a level of arming knowing that C will challenge and they will escalate and fight; here D will select their optimal arming level $p^* = \hat{p}(v_D)$. Only low-types "bluff." Bluffing describes behavior within a pooling equilibrium where low-types select a level of arming where they (a) mimic the high-type's arming level and (b) would not fight if challenged ($p^* \leq p^D(\underline{v}_D)$). Only high-type players "signal." When high-types signal, high-types select a level of arming beyond what they would need to deter in a complete information game as a way to distinguish themselves from low-type D's. In the signalling equilibrium space, high-types will select $p^* = \bar{p}$, where $\bar{p}$ is the arming level that makes low-type D indifferent between arming and acquiescing. In the Signal-Acquiesce equilibrium space, high-types will arm and deter C while low-types will not arm and acquiesce.

Figure 2 illustrates one example of equilibrium behavior. I vary $\underline{v}_D$ on the x-axis, and I vary v_C on the y-axis. In the region where low-type D's have the lowest resolve and v_C is low (bottom left), high-type D's can arm to a level that will keep C from challenging, and low-type D's are unwilling to arm to this level: this is the "Deter-Acquiesce" equilibrium space.[21] As v_C grows (moving up), it either becomes too costly for high-types to deter C ($K(p^C)$ is too high) or it no longer becomes feasible to deter C ($p^C > p_1$). When this is the case, high-type D's will switch from deterring either to going to war or to acquiescing. Under the parameters used here, the high-type D's prefer going to war over acquiescing. Note that moving to the right of the War-Acquiesce space—increasing $\underline{v}_D$ to a point that approaches $\bar{v}_D$— moves into the War-War equilibrium space, where both types select an arming level that will result in C challenging, and both types fighting. In the upper right corner, because low-types look more like high-types (i.e. they care almost as much about the issue), they are also willing to fight.

Returning to low values of v_C, in the "Deter-Bluff" equilibrium space, low-type D's care more about the issue because $\underline{v}_D$ has increased, but C does not. In this space, both D's are willing to select some arming level where high-type D's would be willing to fight if challenged ($p^* \geq p^D(\bar{v}_D)$), where low-type D's would be unwilling to fight if challenged ($p^* \leq p^D(\underline{v}_D)$), and where C, knowing this, isn't resolved enough to challenge. Here low-type D's can successfully bluff without the bluff ever being called. As C cares more about the issue (moving up), one of two things can happen. When low-type D's could have fairly low resolve (i.e. the Signal-Acquiesce region), high-type D's will signal to distinguish themselves from low-type D's by selecting an

[21] This space has the curve above the Signal-Acquiesce equilibrium space because at the start of the curve $p^C = p^D(\bar{v}_D)$. Then, as v_C increases, p^C increases while $p^D(\bar{v}_D)$ stays the same, making deterrence more costly; thus, under more values of $\underline{v}_D$, low-type D's are unwilling to fight.

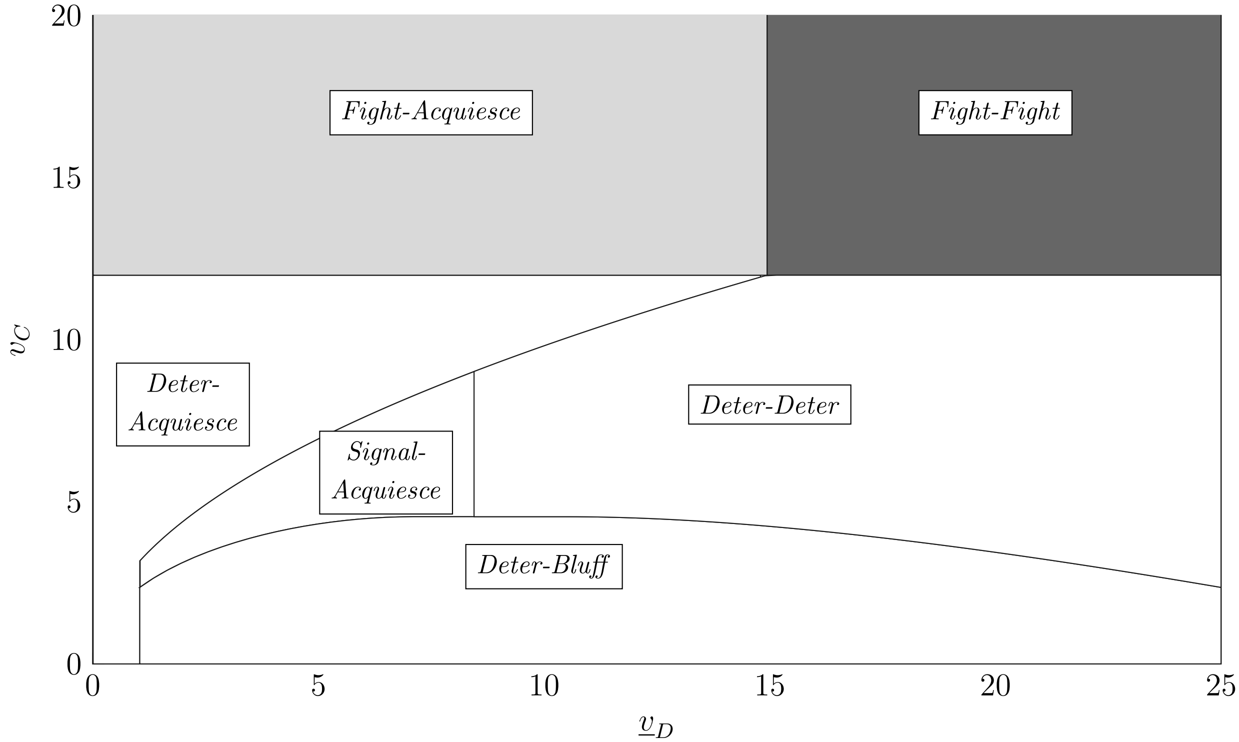

Figure 2: Equilibrium behavior. The x-axis varies $\underline{v}_D$ and the y-axis varies v_C. The intensity of shading represents the the likelihood of war. In the non-shaded (white) equilibrium spaces, war never occurs. In the light-gray "War-Acquiesce" equilibrium space, war occurs when D is type $\bar{v}_D$, or with probability π. In the dark-gray "War-War" equilibrium space, war always occurs.

arming level $p^* = \bar{p}$ that low-type D's are not willing to arm up to ($\underline{v}_D - K(\bar{p}) \leq 0$). This costly signal allows C to know that only high-types would select $p^* = \bar{p}$, and this will motivate C to not challenge when observing $\bar{p}$ because then C would always have to fight. When low-type D's care more about the issue (moving to the right), this is the "Deter-Deter" equilibrium space where both types of D's are willing to arm up to level p^* such that low-type D's would fight (i.e. $p^* \geq p^D(\underline{v}_D)$) and where C would be deterred from challenging $p^* \geq p^C$.

Two equilibrium spaces, "Acquiesce-Acquiesce" and "Deter-War," are listed in Table 2 but not graphed. These are fairly straightforward. Under different assumed parameters (or functional forms), these are possible, and would occur in the upper-regions of the graph.

7 Results

All proofs are in the appendix.

7.1 Arming Levels are Increasing in Private Type

Remark 1 (Arming increases in type): *High-types always select greater arming levels ($p^*(\bar{v}_D) \geq p^*(\underline{v}_D)$).*

In order for D to be willing to fight when challenged, D must arm to at least level $p^D(v_D) = 1 - \frac{\alpha v_D}{c_D + nN_D}$. Were this a complete information game and both low- and high-type D's were trying to deter C from challenging, high-type D's could deter C with weakly lower levels of arming than low-type D's (because $p^D(v_D)$ is decreasing in v_D).[22]

But, in the incomplete information game, arming is *always* increasing in private type. Why? Some of the logic for Remark 1 can be illustrated within the Deter-Deter equilibrium space when both types select $p^D(\underline{v}_D)$, which is beyond what high-types need to be willing to fight. Within this equilibrium space, high-types cannot deviate and select lower levels of arming due to the standard costly-signalling logic.[23] Suppose all high-types switched and selected $p' = p^D(\bar{v}_D)$ (while assuming that $p^D(\bar{v}_D) > p^C$). If this were the case, then C would not challenge, as C would know that the high-types who selected arming level p' would fight if challenged. However,

[22] D arms to weakly lower levels because sometimes D can deter by arming to p^C, which is unchanging in v_D.

[23] See (Cho and Kreps, 1987) for a discussion.

this cannot be an equilibrium because low-type D's would have an incentive to mimic high-types by selecting arming level p' (which is less costly) and not get challenged.

There are two subtleties to Remark 1 worth mentioning. First, recall that in some cases D will arm with the intent of deterring C, and in other cases D will arm with the intent of fighting D in a conventional war. Because deterring C requires a greater arming level than fighting with C,[24] Remark 1 implies that equilibria can exist where low-type D's fight and high-type D's do not. This is consistent with other models of deterrence like Slantchev (2005). Figure 3 illustrates one example of this.

The top graph in Figure 3 illustrates low-type D's equilibrium arming levels (p, on the y-axis) over a range of C's valuation of the asset (v_C, on the x-axis). The bottom graph illustrates high-type D's equilibrium arming levels over the same range of C's valuations. The key take-away from this figure is that, under this parameterization, there is a range of v_C's where low-type D's fight but high-type D's do not (roughly 30.85 to 31.25). This equilibrium behavior is consistent with Remark 1, as every low-type D selects a weakly lower level of arming for a given v_C. To elaborate on what is happening, consider what would happen if D was facing a Challenger with valuation $v_C = 31$. A low-type D has three arming options: (1) they could select some optimal arming level in preparation for fighting ($p = 0.65$); (2) they could arm to deter C and attain the asset with certainty ($p = 0.85$); or (3) they could not arm and acquiesce when challenged ($p = p_0$). For these parameters and all possible v_C, the not-arming option (3) is always worse than fighting (1) or deterring (2). Additionally, at $v_C = 31$, deterring C is too costly for low-type C's, and low types prefer arming to a lower-level and fighting relative to maintaining the high level of arming needed to actually deter C. In contrast, when $v_C = 31$, high-type D's prefer deterring to fighting . Because high-type D's value the asset more than low-types, entering into a "costly lottery" over the asset is less appealing for high-types than getting the asset outright; thus high-types are willing to arm to high levels to deter. Of course, even high-type D's have their limits. At $v_C = 31.25$, the costs of arming are too high even for high-type D's, and high-types switch and prefer to fight.

As the second subtlety, Remark 1 rules out a range of possible equilibria behavior. Recall

[24] Fighting is only possible when $p^D(v_D) < p^C$ and D selects some $p \in [p^D(v_D), p^C)$; meanwhile, when feasible, D can deter C by setting $p = p^C$.

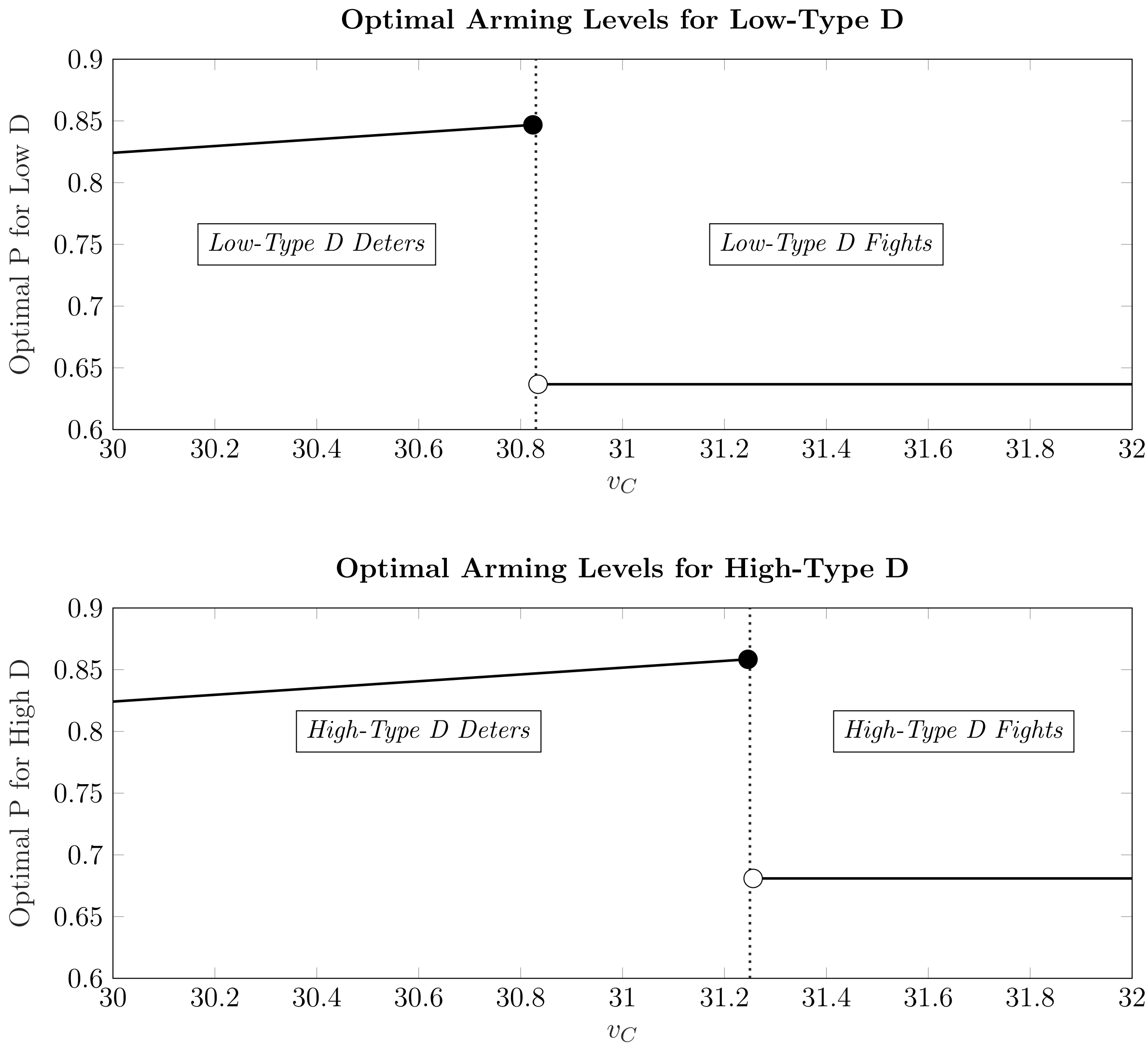

Figure 3: Optimal Arming Levels for D at Different Values of C's Resolve

The x-axis varies the values of v_C, and the y-axis displays D's optimal arming levels for low-types (top graph) and high-types (bottom graph). The labels indicate where D deters or where D fights. In addition to confirming Remark 1 (the optimal level of arming is weakly increasing in private type), the key take-away is that under some parameters, low-types fight while high-types do not. Parameters are included in the Appendix.

that D "acquiescing" involves the lowest level of arming (p_0), fighting involves a greater level of arming ($\hat{p}(v_D)$), and deterring involves the greatest level of arming ($max \left\{p^D(v_D), p^C\right\}$. Remark 1 implies that it will never be the case that high-type D's acquiesce when low-type D's deter or fight, and high-type D's will never fight when low-type D's deter.

7.2 Introducing Nuclear Risk Requires More or Less Conventional Arming for Deterrence

Remark 2 (Nuclear and Conventional Forces are Imperfect Substitutes). *Increasing nuclear instability (n) has ambiguous effects on the equilibrium level of conventional arming (p).*

Increasing the level of nuclear instability could result in D optimally selecting greater or lower levels of conventional arming. This ambiguity is cleanly seen when D is trying to deter C (for example, in the Deter-Acquiesce equilibrium). Within this equilibrium space, $p^* = max \left\{p^D(\bar{v}_D), p^C\right\}$, meaning that to deter C, D must arm to a level that will make D willing to escalate and will keep C from challenging. Consider the case when the relevant constraint is D's own willingness to fight ($p^D(\bar{v}_D) \geq p^C$), meaning D selects arming level $p^* = p^D(\bar{v}_D)$. Based on how $p^D(\bar{v}_D)$ is defined, $p^D(\bar{v}_D)$ is increasing in n. Intuitively, as nuclear instability increases, for D to be willing to escalate, D must attain a better outcome in the conventional conflict to compensate for the greater nuclear risk from fighting. Next, consider when the relevant constraint is C's willingness to fight ($p^D(\bar{v}_D) < p^C$), meaning D selects arming level $p^* = p^C$. By how p^C is defined, p^C is decreasing in n. Intuitively, as nuclear instability increases within a conventional war, C becomes less willing to challenge and end up in this conflict. Thus, depending on underlying model parameters, the effect of increasing nuclear instability on D's selected arming level is ambiguous.

Remark 2 captures the difficulties in attempting to use nuclear risk as a substitute for conventional capabilities. Waltz (1981) describes several reasons why states may want nuclear weapons, including the following: "[S]ome countries may find nuclear weapons a cheaper and safer alternative to running economically ruinous and militarily dangerous conventional arms races. Nuclear weapons may promise increased security and independence at an affordable price." Within the scope of this paper—where a strategic nuclear exchange is a background risk within a conventional war—only sometimes can nuclear weapons serve as an alternate to conventional arming.[25] In some cases, increasing nuclear risk makes D less willing to fight over issues, which means that unless D invests in a robust conventional capability to ensure a conventional war with C goes more in their favor, D will not be willing to fight when challenged.[26]

Result 2 also has clear policy implications. Consider the 2010 Nuclear Posture Review, which states that "fundamental changes in the international security environment in recent years—including the growth of unrivaled US conventional military capabilities [and] major improvements in missile defenses...enable us to fulfill...objectives at significantly lower nuclear force levels and with reduced reliance on nuclear weapons...without jeopardizing our traditional deterrence and reassurance goals" (Leah and Lowther, 2017). On one hand, the model partially supports this assessment, as a robust conventional force posture can sometimes deter in settings where nuclear instability is lowered. On the other hand, we have not observed the counterfactual world where nuclear instability has been lowered; if this instability was pivotal in deterring rivals, reducing reliance on nuclear weapons could require an expansion in conventional forces. Thus, it remains an open question how the scope of challenges would change when nuclear instability is lowered.

7.3 Evidence of a "Nuclear Peace"

Remark 3 (Nuclear Peace). *Increasing nuclear instability results in fewer instances of war. Formally, we define nuclear instability parameters* $n', n'' \in \mathbb{R}_+$ *with* $n' < n''$*. If* n' *shifts to* n''*, then the set of parameters where war occurs shrinks, and the likelihood of war weakly decreases.*

Remark 3 implies that moving from a lower to higher nuclear instability parameter will shrink the parameter set under which a conventional war will occur. This has empirical implications. One way to formalize the introduction of nuclear weapons is increasing the nuclear instability parameter from $n = 0$ to some $n > 0$. If this is the case, Remark 3 implies that if we compared how history played out from 1950-present ($n > 0$) to a counterfactual history without the

[25] In an alternate setting, where one state is faced with an existential threat and nuclear weapons are used as a last resort, the Waltz (1981) comments seem highly plausible.

[26] Admittedly, Waltz (1981) may be discussing tactical nuclear weapons. I discuss this in depth below.

development of nuclear weapons ($n = 0$), then we would observe weakly more conventional conflicts in the counterfactual history. And, if we observed a third counterfactual history—where greater nuclear instability existed among states with nuclear weapons—then we would observe even fewer conventional conflicts.

Two forces drive the nuclear peace. First, as nuclear instability increases, a conventional war becomes a worse option for the defender because the risk of a catastrophic nuclear exchange during the conventional war grows. Second, as nuclear instability increases, the set of possible arming levels that could result in war—in other words, arming levels where C would be willing to challenge and D would be willing to escalate if challenged—is shrinking. Essentially, when D arms in preparation to fight a war, D is undertaking a constrained optimization problem. Specifically, when D wants to fight, D selects a level of arming that optimizes how they do in a conventional war over the set of arming levels that results in C being willing to challenge and D being willing to escalate when challenged. Together, as nuclear instability increases, D's objective function produces categorically worse options—because now conventional war is less stable—and the set over which D optimizes shrinks—because some arming levels that previously would have motivated D or C to fight now, under high nuclear instability, no longer motivate D or C to fight.

To give a sense of what these results look like visually, in Figure 4, I include three plots, each with fixed parameters other than n which increases moving down. *Draft note: I apologize, I am working on improving this image.

In the top plot, $n = 0$ (i.e. there is no risk of a nuclear exchange) and there is a large range of values where the game ends with one type or both types of D declaring war. In the second (middle) plot, $n = 0.02$, the parameter space where war occurs shrinks. As n increases, it becomes easier to deter C by arming to level p^C (which is decreasing in n), and it makes war worse; this shifts the Deter-Acquiesce and Deter-Deter regions upwards and grows them while shrinking the the War-Acquiesce and War-War regions (the latter is now only a sliver) In the third plot $n = 0.05$; for the highest values of v_C, both types of D optimally select $p = p_0$ and acquiesce. In addition to arming to level p^C becoming cheaper (growing Deter-Acquiesce and Deter-Deter regions), here war becomes more costly for D, and for the greatest values of p^C,

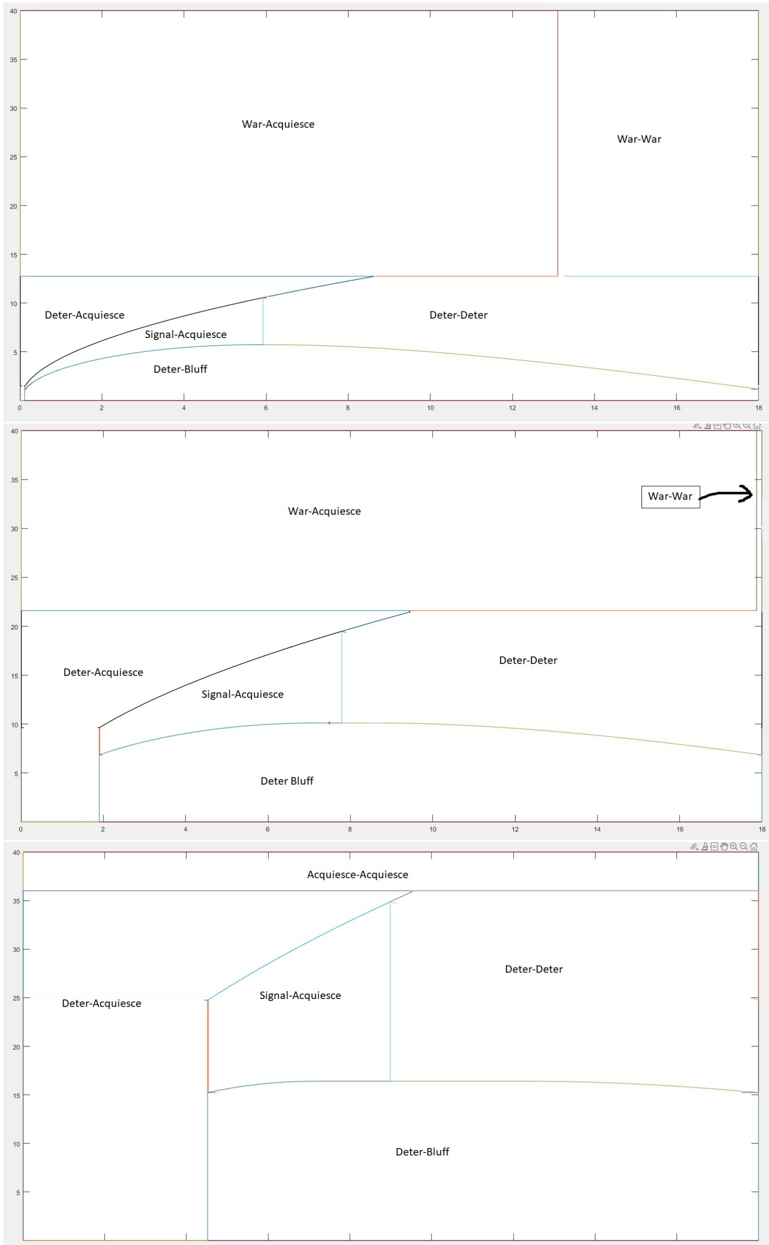

Figure 4: On the x-axis, I vary the values of $\underline{v}_D$, and on the y-axis I vary the values of v_C. In the top figure, $n = 0$. In the middle figure, $n = 0.02$. In the bottom figure, $n = 0.05$. Apologies the figures are small and crummy, I will fix this in the next iteration

both types of D will end up acquiescing to avoid war.

7.4 On Welfare

Remark 4 (Welfare Analysis). *Increasing nuclear instability can lower overall welfare, even in equilibria where war does not occur.*

While increasing nuclear instability (n) does lead to weakly less conventional war (Remark 3), adding nuclear instability is not inherently welfare improving. Increasing nuclear instability can make both actors weakly worse off through two channels. The first channel is straightforward: while adding nuclear instability can reduce the likelihood of war, when it does not rule out war altogether, war can be more costly with greater nuclear escalation risk.[27] This idea is emphasized in work discussing the inherent risks of a world with nuclear instability (Sagan, 1985, 1994). Second, when states do not go to war, increasing nuclear instability can make deterrence more expensive. When D arms to level $p^D(v_D)$, increasing n requires D to arm to a greater level, thus resulting in D attaining a strictly lower utility and C's utility remaining unchanged. This latter point is not well examined. Nuclear optimists point to the decrease in great power conflict as a virtue of the nuclear era (De Mesquita and Riker, 1982). This paper suggests that even if there has been a decline in conventional conflict, this decline may have come with added costs of conventional arming. To the best of my knowledge, the possibility of these added costs is new; they have not been highlighted by nuclear pessimists (Snyder, 1965; Jervis, 2017; Schelling, 1966), nor have they been discussed in the context of the expected value of nuclear weapons (Kydd, 2019).

7.5 Nuclear Instability, Costs, and Arming Incentives

Remark 5 (Stability Instability Paradox): *Under select conditions, as nuclear instability increases (n increases) or D's costs of a nuclear exchange increases (N_D increases), D arms in such a way that conflicts will be shorter and more decisive.*

Formally, consider parameters N_D, N'_D, n and n', where $N_D < N'_D$ and $n < n'$. Suppose under the parameters n or N_D, type $v_D \in \{\underline{v}_D, \bar{v}_D\}$ D goes to war (i.e. $p^* = \hat{p}(v_D)$).

[27] Naturally, as n increases, D may select a new p^*. However, sometimes escalation risk will still increase.

(a) If $p^*(N_D) \leq \frac{1}{2}$, $p^*(N_D') \leq \frac{1}{2}$ and $p^C > \frac{1}{2}$,[28] then $p^*(N_D) \geq p^*(N_D')$. And, if $p^*(N_D) \geq \frac{1}{2}$, $p^*(N_D') \geq \frac{1}{2}$, and $p^D(v_D) < \frac{1}{2}$, then $p^*(N_D) \leq p^*(N_D')$.

(b) If $p^*(n)$ and $p^*(n')$ are small enough, and $p^C > \frac{1}{2}$,[29] then $p^*(n) \geq p^*(n')$. And, if $p^*(n)$ and $p^*(n')$ are large enough, $p^D(v_D) < \frac{1}{2}$, and $p^C \geq p_1$, then $p^*(n) \geq p^*(n')$.

The stability-instability paradox remains a caveat on the concept of a nuclear peace (Snyder, 1965; Jervis, 1984).[30] Empirically, there is some evidence that while the great powers avoided direct, large-scale conventional conflict, they did engage frequently at lower levels of conflict (Rauchhaus, 2009; Early and Asal, 2018). That said, the scope conditions for the paradox outside of nuclear-level stability is still an open topic (O'Neill, 2019), though (Powell, 2015) offers one such analysis. Powell finds that greater nuclear instability (approximated by the risk that large conventional conflicts turn nuclear) results in a more occurrences of smaller conflicts using low-levels of violence, and reduced occurrences of larger conflicts using high-levels of violence. Powell suggests that nuclear instability results in actors bringing less force to bear.

This model presents a new set of results. While under some parameters these results echo Powell,[31] at other times, I find that defenders commit more force to a conflict when faced with greater nuclear instability and greater nuclear costs. Why? Because prolonged conflicts bear a greater risk of a nuclear exchange, the defender is incentivised to avoid these drawn-out conflicts. Here, the defender can avoid prolonged conflicts by reducing military parity through greater or lower levels of arming. Put another way, this paper and Powell offer different answers to how the nuclear era changes conventional conflict. Whereas Powell finds that conventional conflicts with nuclear risk should exhibit lower-levels of force, I find that these conflicts will be more decisive and less prolonged, which sometimes involves more aggressive force postures (with some technical caveats, discussed below). I visualize this result, then discuss the empirical implications.

[28]Recall if $p^*(N_D)$ and $p^*(N_D')$ may denote sets; in this case, we abuse notation and assume that every element of both $p^*(N_D)$ and $p^*(N_D')$ are less than or equal to $\frac{1}{2}$.

[29]We will clarify what "small enough" and "large enough" mean in the Appendix.

[30]The paradox suggests that "stability" at the nuclear level between adversaries (i.e. adversaries each possessing a capable nuclear arsenal and where there is little risk of escalation) can breed "instability" (i.e. conflict) at lower-levels.

[31]I find that sometimes defenders will commit less conventional arms to a conflict when faced with higher degrees of nuclear instability or higher costs from a nuclear exchange.

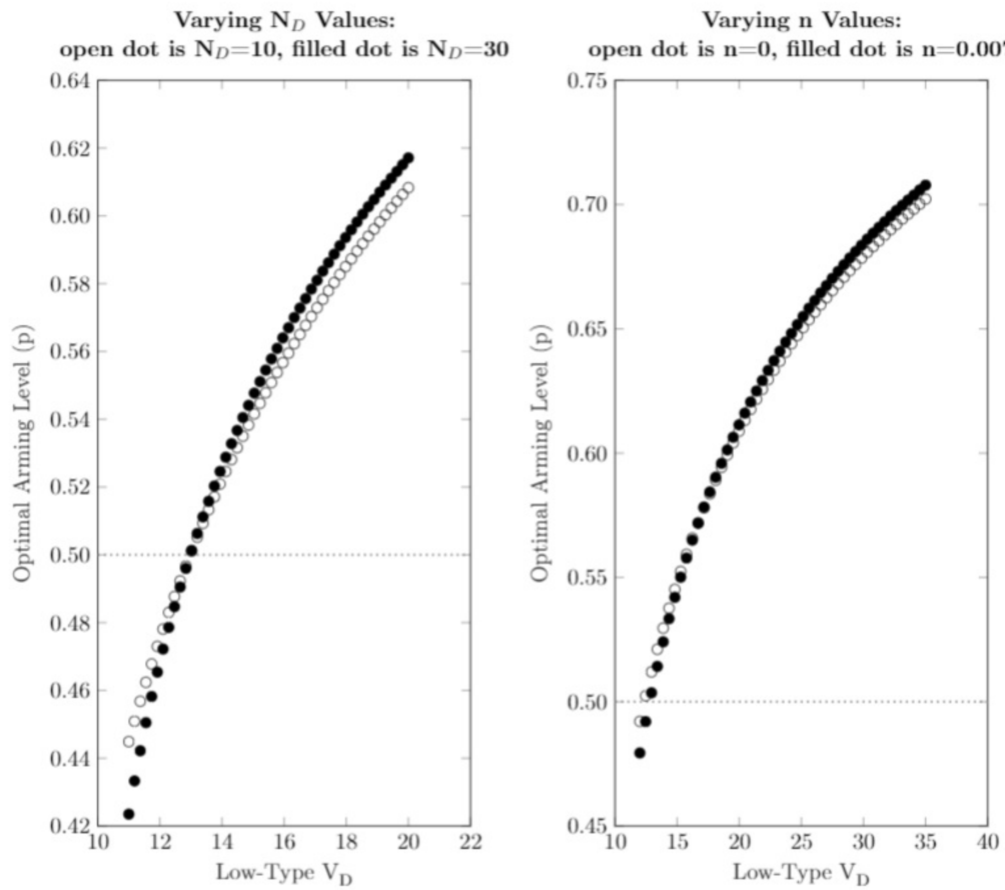

Figure 5: Optimal Arming Levels for D at Different Values of Nuclear Risk. In the left figure, increasing D's costs from a nuclear exchange N_D from $N_D = 10$ to $N_D = 30$ results in D selecting greater or lower levels of arming, but the selection is such that conflict becomes less protracted. In the right figure, increasing nuclear instability parameter n from $n = 0$ to $n = 0.007$ also can result in D arming in such a way to make the conflicts less protracted, but this result is not as clean as the effects on N_D. Here, D arms to make conflict less protracted when, under $n = 0$, $p \leq 0.5$ or $p \geq 0.59$, but not within that range.

Figure 5 visualizes how the defender responds to growing nuclear costs and risks. The left graph in Figure 5 illustrates the equilibrium arming level (y-axis) for a range of $\underline{v}_D$ values (x-axis) under a set of parameters where low-type D's always fight. The different dot styles capture changes in D's nuclear cost N_D. The open dots plot the arming levels when the costs of a nuclear exchange are relatively low ($N_D = 10$), and the closed dots plot arming levels when the costs of a nuclear exchange are relatively high ($N_D = 30$). The right graph in Figure 5 considers changes in nuclear instability n. The open dots are the arming levels at when $n = 0$, and the closed dots are for $n = 0.07$.

Take the open dots on the left graph of Figure 5, representing the equilibrium arming levels for various $\underline{v}_D$ for $N_D = 10$. As N_D increases (moving from open to closed dots at a given $\underline{v}_D$), equilibrium arming levels below the initial arming level of $p^* = 0.5$ decrease and equilibrium arming levels above the initial arming level of $p^* = 0.5$ increase. For example, for $v_D = 11$ and $N_D = 10$, D will arm to level $p^* = 0.445$. Then, as the nuclear costs increase to $N_D = 30$, D is incentivised to make the conflict more decisive, which D does by selecting a lower arming level $p^* = 422$. This decrease in arming represents D moving towards a less protracted conflict. In contrast, for all arming levels $p^* \geq 0.5$, the way to create a more decisive conflict is inverted. At $\underline{v}_D = 20$, under low nuclear instability D will select $p^* = 0.610$, and under high instability ($N_D = 30$) D will select $p^* = 0.619$. Together, the effect of increases in N_D results in a move towards more decisive conflicts, rather than a move towards lower or higher force postures.

Now consider the open dots on the right graph of Figure 5, representing the equilibrium arming levels for various $\underline{v}_D$. As n increases (moving from open to closed dots), equilibrium arming levels shift in a similar manner to those on the left, but no longer around the same inflection point of $p^* = 0.5$. Why? In short, this is a more complicated relationship, as n factors into a wider range of conflict payoff terms. In addition to influencing the cost function from conventional war (because a high n can cut the costs from a conventional war short), it also changes the likelihood of a conventional versus nuclear conclusion to the conflict. This is not to say this relationship cannot be defined: I express a formula for when increasing n increases or decreases the optimal arming level in the appendix. However, this is a more complicated expression, and depends on a wide range of factors outside of the initial p^*.

Of course, in both of these graphs, as nuclear instability and the costs of a nuclear war increased, we (generally) observed a move towards more decisive conflict. Remark 5 goes beyond this: as these variables increase, it's not only that the defender arms in preparation for a more decisive war, the defender may also arm with the intent of avoiding war altogether by deterring C or backing down when challenged.

(**Draft note: what follows is an 11th hour addition and needs additional work.)

These results suggest that, in the nuclear era, defenders select more extreme force postures when they engage in crises. To this end, our results echo Powell (2015) in suggesting that nuclear risk drove Indian troops to limit their forces in the Kargil War.[32] What is new here is that our results can offer an explanation for aggressive military operations conducted with the intent to shorten conflicts in the nuclear era. As an example, consider the Hungarian Revolution (1956). In late October that year, following a series of clashes between student protesters and government forces, a group of anti-Soviet revolutionaries ousted or killed Hungarian communist leaders and members of the Hungarian secret police, eventually installing Imre Nagy as the Prime Minister (on October 27). While at first Soviet leadership considered negotiating with Nagy and the new Hungarian Government, the Soviet Union quickly changed course and invaded Hungary. By November 3, "Operation Whirlwind" was underway, where between 30,000 and 60,000 Soviet troops invaded Hungary and circled Budapest.[33] By November 11, the revolutionaries had been decisively defeated, the revolutionary government was deposed, and Soviet backed forces resumed control of Hungary.

Within the broader Cold War context, Soviet activities in Hungary were not without broader risk. Until that point, the Eisenhower administration had vocally advocated the "rolling back" of Soviet influence in Eastern Europe, even if it required using armed forces.[34] And, when the Soviet Union invaded Hungary, it made deliberations all the while considering the possibility of Western forces stepping in and there being a direct conflict. But, instead of this convincing the Soviet Union to apply less force, these circumstances incentivised the Soviet leadership to act aggressively and crush the revolution before it could spread. Internal documents and analyses of Soviet decision makers highlighted the need to act quickly, as the United States was (at the time) occupied by what was happening in the Suez Crisis; additionally these leaders were concerned that a failure to act fast could result in other revolutions spreading to other Eastern European states. In response to these pressures, the Soviet leadership deployed a robust force posture, thus mitigating the risk of a broader, protracted war with nuclear risk.

In short, our model and analysis suggest that nuclear instability can incentivise both less

[32] We will not discuss this case, because it is discussed at length in Powell and our results are consistent with what is discussed there.

[33] There are mixed reports of this. Some Soviet archives suggest the numbers were closer to 30,000 while other sources suggest the number was closer to 60,000. Additionally, it was documented that 17 Soviet Divisions were in Hungary; to the best of my understanding, Soviet Divisions had roughly 10,000 troops.

[34] See "Rollback, Liberation, Containment, or Inaction? U.S. Policy and Eastern Europe in the 1950s" by László Borhi for a review.

aggressive military maneuvers—as it was in the Kargil War—and more aggressive military maneuvers—as it was in the Soviet invasion of Hungary. While we cannot observe a counterfactual world, what the theory here suggests is that in addition to creating more instances of peace, added nuclear instability and costs generally lead to more decisive conflicts.

7.6 Deterrence Without War

Remark 6: *Peaceful signalling of resolve is possible.*

A key distinction between equilibrium behavior in this model and the model in Powell (2015) is that here it is possible for the defender to signal their resolve without ever having to go to war.

Within the Signal-Acquiesce region, high-type D's demonstrate their resolve by arming to a level beyond what is needed to make themselves willing to fight, and beyond what is needed to make a conflict sufficiently damaging for C. Instead, the relevant constraint is that high-type D's must arm to a level where low-resolved D's would be unwilling to mimic high-type D's due to the costs of arming. As a result, in equilibrium, only high-type D's arm to level $\bar{p}$, C knows upon seeing $p = \bar{p}$ that D is a high-type, and C will never challenge. Arming as a costly signal of resolve works, and this equilibrium is peaceful.

This is distinct from how costly signalling functions in Powell (2015). In Powell, manipulating nuclear risk is costless unless a war breaks out. And, following the standard signalling logic, unless the signal is costly, low-types are incentivised to mimic high-types and this undermines the informative value of the signal. As a result, in Powell, the defender can only signal their resolve by actually fighting sometimes because only through conventional conflict are the signal's costs realized.[35] In this regard, arming as a costly signal does effectively separate low-types from high-types, but the equilibrium is not always peaceful.

These distinctions have real-world implications. The model here suggests that resolved defenders can deter challengers and entirely prevent conflict through the signal of a costly robust

[35]In the equilibrium in Powell, the resolved defender must select a level of nuclear risk that makes the challenger indifferent between challenging and not challenging, and the resolved defender must fight with positive probability.

conventional force posture. In contrast, Powell (2015) suggests that manipulating nuclear risk cannot function as a fully effective deterrent against challengers, and that conflict must sometimes break out. There are two ways to interpret these results. First, the model here presents a more optimistic perspective on nuclear deterrence than Powell (2015). In Powell, war is an inevitable part of signalling resolve. In contrast, I find it is possible to signal resolve and deter an opponent without ever having to resort to conflict. Second, from a practical perspective, if a defender wanted to signal resolve and avoid conflict, manipulating nuclear risk (ala Powell) is not as effective as committing troops.

7.7 Additional Results

7.7.1 Deterrence Failure and Nuclear Instability

I classify a "deterrence failure" as any equilibrium where C challenges D. As nuclear instability increases, deterrence failures could become more or less common. Two competing effects drive this. First, sometimes D deters C by arming to level $p = p^D(v_D)$. As n increases, $p^D(v_D)$ also increases, possibly to the point where D is unwilling to undertake the (costly) arming needed to deter C from challenging. If D is no longer willing to deter C from challenging, then the increase in n produces a deterrence failure. Second, sometimes D deters C by arming to level $p = p^C$. As n increases, p^C decreases. In some cases, D may have been unwilling to arm to level p^C and deter C from challenging under under a low level of nuclear stability; however, under greater levels of nuclear instability, which leads to a lower p^C, D would be willing to arm to a level that prevents a deterrence failure. Ultimately, whether D experiences more or fewer deterrence failures following increases in nuclear instability depends on the underlying conditions of the case.

These findings pose a useful counterpoint to Mearsheimer (1990), which states that "[d]eterrence is most likely to hold when the costs and risks of going to war are unambiguously stark."[36] Using the definition of deterrence failures here, the model suggests this is not necessarily the case. When there is a high nuclear risk, it could undermine D's deterrent threat and encourage C to

[36]The model also confirms that deterrence is more likely to hold when the costs of a nuclear war for the challenger (N_C) are high, but deterrence is more likely to fail when the costs of a nuclear war for the defender (N_D) are high.

challenge.

7.7.2 Caveats to the Nuclear Peace: Effects Beyond Nuclear Instability

Understanding the true effect of the nuclear revolution on conventional conflict is potentially more complex than what is presented above. This model finds that increasing the nuclear instability parameter n always results in less conventional conflict (Remark 3). A natural interpretation of the nuclear revolution is that the world changed from $n = 0$ to $n > 0$ and, as a result, engaged in less conflict.

However, the nuclear revolution has also shaped conventional forces in ways that influence the likelihood of conventional conflict. Today, several states deploy nuclear powered submarines and aircraft carriers. Additionally, it is possible that one day tactical nuclear weapons are deployed on the battlefield as part of a "conventional" conflict. Together, it could be claimed that technologies like nuclear submarines, because they are more efficient and capable than their non-fission powered counterparts, could lower the costs of conducting a conventional war (i.e. it reduces c_D and c_C). Lowering c_D and c_C have the effect of increasing the parameter space under which conventional conflict occurs. The rationale is essentially the opposite of what is shown in Remark 3: as D's (or C's) costs from war decrease, conditional on fighting, D's objective function increases at all points, and the set over which D optimizes expands.

To the extent that the nuclear revolution both increased nuclear instability and lowered the costs of conventional war, the true effect of the nuclear revolution requires disentangling competing effects. Consider, for example, the possibility that the challenger develops and would deploy tactical nuclear weapons, and the deployment of these technologies increases nuclear instability. When this occurs, evaluating the nuclear peace hypothesis becomes more complex. On one hand, increases to the nuclear instability parameter would still, at a fixed arming level, make a conventional war more costly. On the other hand, if the challenger could conduct conflict more easily, this expands the set of arming levels that would eventually result in a conventional crisis. Without additional structure, I am unable to claim that a nuclear peace result would hold following the development of nuclear weapons.

8 Extensions

8.1 War-of-Attrition Subgame

The model is set up in such a way where fighting could occur as a continuous-type war-of-attrition. I have some notes on how this would work, but this modification may be excessive (at least for this paper).

8.2 Endogenously Arming Challenger

Pending

8.3 Covert Operations

Pending.

9 Conclusion

Every day, every human on earth lives with the background risk of a catastrophic nuclear exchange (Sagan, 1985). But, as this paper also suggests, that neorealist logic may be correct: adding latent nuclear instability comes with the benefit of reducing the likelihood of a conventional war. To this end, I have demonstrated that the observed "long peace" could be a "nuclear peace," where the nuclear great powers are less willing to engage in large conventional wars and more willing to engage in small, regional contests with small risk of nuclear escalation.

Of course, as this paper has demonstrated, this latent nuclear risk does not come without costs. While much attention has been paid to the underlying risks that nuclear weapons hold (Sagan and Waltz, 1995), this paper demonstrates that nuclear weapons can reduce welfare through other channels as well. In a nuclear world, deterrence may become more difficult and costly: states might find themselves investing more in conventional armaments to make their threat to fight a conventional war that could end in disaster. And, even the prospect of a "nuclear peace" is subject to caveats; to the extent that nuclear technologies make conventional conflict cheaper (through, for example, the development of nuclear-powered naval vessels), the connection between nuclear technologies and peace is open to debate.

There are still many research avenues on this topic to consider. This paper treats several variables as exogenous when, in reality, they are strategic choices. For example, while I believe it is difficult to credibly manipulate nuclear instability within a crisis, bigger picture, this paper does not consider how states optimally design nuclear instability based on what crises they expect with deterrence in mind. Additionally, this paper treats the challenger's arming levels as fixed when these are plausibly endogenous. Finally, this paper adopted a specific functional form for wartime payoffs. Future research could either (a) generalize this, or (b) use the framework here to work within a continuous time conflict framework.

References

Baliga, S., De Mesquita, E. B. and Wolitzky, A. (2020) Deterrence with imperfect attribution, *American Political Science Review*, **114**, 1155–1178. 3, 4

Bennett, D. S. and Stam, A. C. (1996) The duration of interstate wars, 1816–1985, *American Political Science Review*, **90**, 239–257. 2

Bennett, D. S. and Stam, A. C. (2009) Revisiting predictions of war duration, *Conflict Management and Peace Science*, **26**, 256–267. 2, 3

Bueno de Mesquita, B., Koch, M. T. and Siverson, R. M. (2004) Testing competing institutional explanations of the democratic peace: The case of dispute duration, *Conflict Management and Peace Science*, **21**, 255–267. 6

Chiba, D. and Johnson, J. C. (2019) Military coalitions and crisis duration, *The Journal of Politics*, **81**, 1466–1479. 2

Cho, I.-K. and Kreps, D. M. (1987) Signaling games and stable equilibria, *The Quarterly Journal of Economics*, **102**, 179–221. 6, 23

Coe, A. J. (2018) Containing rogues: A theory of asymmetric arming, *The Journal of Politics*, **80**, 000–000. 4

De Mesquita, B. B. and Riker, W. H. (1982) An assessment of the merits of selective nuclear proliferation, *Journal of Conflict Resolution*, **26**, 283–306. 1, 7.4

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

Di Lonardo, L. and Tyson, S. A. (2022) Political instability and the failure of deterrence, *The Journal of Politics*, **84**. 3

Early, B. R. and Asal, V. (2018) Nuclear weapons, existential threats, and the stability–instability paradox, *The Nonproliferation Review*, **25**, 223–247. 7.5

Fearon, J. D. (1997) Signaling foreign policy interests tying hands versus sinking costs, *Journal of Conflict Resolution*, **41**, 68–90. 3, 4

Filson, D. and Werner, S. (2002) A bargaining model of war and peace: Anticipating the onset, duration, and outcome of war, *American Journal of Political Science*, pp. 819–837. 2

Filson, D. and Werner, S. (2004) Bargaining and fighting: The impact of regime type on war onset, duration, and outcomes, *American Journal of Political Science*, **48**, 296–313. 2

Gaddis, J. L. (1986) The long peace: Elements of stability in the postwar international system, *International security*, **10**, 99–142. 1, 1

Gartzke, E. and Kroenig, M. (2016) Nukes with numbers: empirical research on the consequences of nuclear weapons for international conflict, *Annual review of political science*, **19**, 397–412. 4

Gurantz, R. and Hirsch, A. V. (2017) Fear, appeasement, and the effectiveness of deterrence, *The Journal of Politics*, **79**, 1041–1056. 3, 4

Huth, P. K. (1999) Deterrence and international conflict: Empirical findings and theoretical debates, *Annual Review of Political Science*, **2**, 25–48. 3, 4

Jervis, R. (1979) Deterrence theory revisited, *World Politics*, **31**, 289–324. 4

Jervis, R. (1984) The illogic of american nuclear strategy. 7.5

Jervis, R. (2017) *Perception and misperception in international politics*, Princeton University Press. 1, 7.4

Joseph, M. (2020) The secret failure of nuclear containment, *Manuscript*. 4

Koch, M. T. (2009) Governments, partisanship, and foreign policy: The case of dispute duration, *Journal of Peace Research*, **46**, 799–817. 6

Krustev, V. L. (2006) Interdependence and the duration of militarized conflict, *Journal of Peace Research*, **43**, 243–260. 2

Kydd, A. H. (2019) The sturdy child vs. the sword of damocles: Nuclear weapons and the expected cost of war, *Security Studies*, **28**, 645–676. 2, 7.4

Langlois, C. C. and Langlois, J.-P. P. (2009) Does attrition behavior help explain the duration of interstate wars? a game theoretic and empirical analysis, *International Studies Quarterly*, **53**, 1051–1073. 2

Langlois, J.-P. P. and Langlois, C. C. (2012) Does the principle of convergence really hold? war, uncertainty and the failure of bargaining, *British Journal of Political Science*, **42**, 511–536. 2

Leah, C. and Lowther, A. B. (2017) Conventional arms and nuclear peace, *Strategic Studies Quarterly*, **11**, 14–24. 7.2

McCormack, D. and Pascoe, H. (2017) Sanctions and preventive war, *Journal of Conflict Resolution*, **61**, 1711–1739. 4

Mearsheimer, J. J. (1990) Why we will soon miss the cold war, *The Atlantic Monthly*, **266**, 35–50. 7.7.1

Mearsheimer, J. J. *et al.* (2001) *The tragedy of great power politics*, WW Norton & Company. 1

Nalebuff, B. (1988) Minimal nuclear deterrence, *Journal of Conflict Resolution*, **32**, 411–425. 4

O'Neill, B. (2019) A game theory analysis of the stability–instability paradox, *Game Theory and Nuclear Stability in Northeast Asia*, p. 68. 7.5

Paine, J. and Tyson, S. A. (2020) 11 uses and abuses of formal models in political science, *The SAGE Handbook of Political Science*. 6

Perrow, C. (2011) *Normal accidents*, Princeton university press. 1, 2

Posen, B. R. (2014) *Inadvertent escalation*, Cornell University Press. 1, 2, 3

Powell, R. (1989) Nuclear deterrence and the strategy of limited retaliation, *The American Political Science Review*, pp. 503–519. 4

Powell, R. (2003) Nuclear deterrence theory, nuclear proliferation, and national missile defense, *International Security*, **27**, 86–118. 4

Powell, R. (2015) Nuclear brinkmanship, limited war, and military power, *International Organization*, **69**, 589–626. 1, 3, 2, 4, 12, 13, 7.5, 7.5, 32, 7.6, 35

Quackenbush, S. L. (2011) Deterrence theory: where do we stand?, *Review of International Studies*, **37**, 741–762. 4

Ramsay, K. W. (2017) Information, uncertainty, and war, *Annual Review of Political Science*, **20**, 505–527. 3

Rauchhaus, R. (2009) Evaluating the nuclear peace hypothesis: A quantitative approach, *Journal of Conflict Resolution*, **53**, 258–277. 7.5

Reich, N. (2022) Signaling strength with handicaps, *Journal of Conflict Resolution*, pp. Pre–Print. 1

Sagan, S. D. (1985) Nuclear alerts and crisis management, *International Security*, **9**, 99–139. 4, 7.4, 9

Sagan, S. D. (1994) The perils of proliferation: Organization theory, deterrence theory, and the spread of nuclear weapons, *International Security*, **18**, 66–107. 1, 2, 3, 7.4

Sagan, S. D. (2020) *The limits of safety*, Princeton University Press. 1, 2

Sagan, S. D. and Waltz, K. N. (1995) The spread of nuclear weapons: A debate. 9

Sagan, S. D. and Waltz, K. N. (2003) *The spread of nuclear weapons: a debate renewed*, WW Norton & Company. 2

Schelling, T. C. (1966) *Arms and influence*, Yale University Press. 1, 4, 7.4

Schelling, T. C. (1980) *The strategy of conflict*, Harvard university press. 1, 4

Schram, P. (2021a) Hassling: how states prevent a preventive war, *American journal of political science*, **65**, 294–308. 4

Schram, P. (2021b) When capabilities backfire: How improved hassling capabilities produce worse outcomes, *Accepted for publication, Journal of Politics*. 4

Schultz, K. A. (2010) The enforcement problem in coercive bargaining: Interstate conflict over rebel support in civil wars, *International Organization*, **64**, 281–312. 4

Shannon, M., Morey, D. and Boehmke, F. J. (2010) The influence of international organizations on militarized dispute initiation and duration, *International Studies Quarterly*, **54**, 1123–1141. 6

Slantchev, B. L. (2004) How initiators end their wars: The duration of warfare and the terms of peace, *American Journal of Political Science*, **48**, 813–829. 2, 3

Slantchev, B. L. (2005) Military coercion in interstate crises, *American Political Science Review*, **99**, 533–547. 1, 7.1

Slantchev, B. L. (2011) *Military threats: the costs of coercion and the price of peace*, Cambridge University Press. 1, 4

Smith, A. (1998) Fighting battles, winning wars, *Journal of Conflict Resolution*, **42**, 301–320. 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. 1, 7.4, 7.5

Spaniel, W. (2019) *Bargaining over the Bomb: The Successes and Failures of Nuclear Negotiations*, Cambridge University Press. 4

Tarar, A. (2016) A strategic logic of the military fait accompli, *International Studies Quarterly*, **60**, 742–752. 4

Waltz, K. N. (1981) The spread of nuclear weapons: More may be better: Introduction, *The Adelphi Papers*. 1, 7.2, 25, 26

Werner, S. (2000) Deterring intervention: The stakes of war and third-party involvement, *American Journal of Political Science*, pp. 720–732. 4

Zagare, F. C. and Kilgour, D. M. (1993) Asymmetric deterrence, *International Studies Quarterly*, **37**, 1–27. 4

Zagare, F. C. and Kilgour, D. M. (1998) Deterrence Theory and the Spiral Model Revisited, *Journal of Theoretical Politics*, **10**, 59–87. 4