Original Manuscript

Journal of Conflict Resolution
2023, Vol. 0(0) 1–26

DOI: 10.1177/00220027231185154
journals.sagepub.com/home/jcr
Sage

Deliberate Nuclear First Use in an Era of Asymmetry: A Game Theoretical Approach

Even Hellan Larsen[1]

Abstract

Most nuclear dyads are characterized by some degree of nuclear and conventional asymmetry. This paper argues that these asymmetries create an environment in which deliberate nuclear first use (DNFU) can be rational. This possibility has been discarded in the formal literature on nuclear escalation because of the common reliance on the assumption of mutually assured destruction (MAD). This paper develops a formal model that traces how and under what circumstances two types of DNFU are rational. First, nuclear imbalances and advancements in counterforce technologies create a damage limitation incentive for a strong actor. Second, conventional asymmetry creates an incentive for the coercive use of nuclear weapons by the weaker player. Moreover, this paper illustrates that these asymmetric conditions are a relevant characteristic in important and very different nuclear dyads: DPRK–US, Pakistan–India, and Russia–US. Thus, the model demonstrates the potential core drivers of DNFU in today's nuclear landscape.

Keywords

game theory, nuclear weapons, war, use of force, international security, asymmetric conflict

[1]Department of Political Science, University of Oslo, Oslo, Norway

Corresponding Author:
Even Hellan Larsen, Department of Political Science, University of Oslo, Moltke Moes vei 31,0851 Oslo, Norway. Boks 1097 Blindern, 0317 Oslo, Norway.
Email: e.h.larsen@stv.uio.no

Introduction

Russian nuclear threats during its war in Ukraine and the evolving nuclear arsenal of the Democratic People's Republic of Korea (DPRK) have revived the question of whether nuclear first use may be rational under certain conditions. For decades, theorists of nuclear strategy (Powell 1987; 1988; 1990; Schelling 1960) have thought almost entirely about a world of mutually assured destruction (MAD): one in which the other side will retain the ability to retaliate with nuclear weapons no matter what, making war essentially "unwinnable" and forcing conflicts to become a competition in risk-taking rather than in military capabilities. However, an increasing body of evidence (Green and Long 2017; Green 2020; Lieber and Press 2006; 2017; 2020; Long and Green 2015) has found that mutual destruction was not assured even during the Cold War, and there is even less reason to think that it can be taken for granted in important contemporary cases such as between the United States and DPRK.

The singular focus on risk-taking is increasingly questionable today, given the conventional and nuclear asymmetry characterizing many nuclear dyads. At the nuclear level, missile defenses and precision strike render total preemption of nuclear retaliation a realistic prospect. At the same time, the potentially devastating consequences of a conventional defeat create an incentive for the conventionally inferior actor to threaten and eventually employ nuclear weapons early in a conventional conflict. Accordingly, we need theory to understand how these characteristics should change our understanding of nuclear escalation in a crisis between nuclear-armed states. The present paper's model provides an important first step in this direction by demonstrating how the asymmetries on the conventional and nuclear levels rationalize deliberate nuclear first use (DNFU).

Nearly all formal studies explore the risk of nuclear war under the assumption of MAD, an assumption that makes DNFU irrational (Nalebuff 1986; Powell 1987; 1988; 1990; 2015; Schelling 1960). Similar assumptions have been adopted in policy analysis, where leaders, such as Putin and Kim Jong Un, presumably must be irrational to go nuclear (Baker and Harris 2017; Bowden 2017; Gabuev 2022; Shapiro 2022). Two exceptions in the formal literature are Powell's (2003) analysis of how missile defense influences stability between a small and large nuclear power and Zagere's (1987) analysis of how variations in the US–Soviet nuclear balance influence stability. It is important to note that neither of these formal models integrates both conventional and nuclear asymmetry. Therefore, the formal literature cannot help us understand the risk of DNFU under these conditions.

The existing literature's reliance on MAD has some important consequences for the analysis and conclusions. First, because DNFU never occurs under MAD, the literature has focused on two types of nuclear escalation risks:

(1) The first type is a purely preemptive first strike. This risk is driven by the conviction that the adversary is about to launch a nuclear strike and that the only thing worse than starting a nuclear war is to suffer a first strike from the

enemy (Powell 1989a; Schelling 1960, 207–54; 1966, 221–48). Still, under the assumption of MAD, such preemptive nuclear employment is only rational if a nuclear war *cannot* be avoided.

(2) The second type of nuclear risk relies on brinkmanship (Nalebuff 1986; Powell 1987; 1988). Here, the adversaries compete in nuclear risk tolerance, where the possibility of a nuclear war is based on *autonomous* risk (Schelling 1960, 199–201; 1966, 99–105).[1]

Autonomous risk factors (such as accidents or irrationality) or the belief that the enemy is about to launch a first strike can be a sufficient cause for nuclear first use. However, although military asymmetries can reinforce the willingness to run these risks, they are *not necessary* under the assumptions of conventional *and* nuclear asymmetry.

Although scholars have studied how asymmetry at the nuclear or conventional level influences nuclear strategy and escalation risks,[2] no paper has incorporated these two levels of asymmetry into the same framework. Furthermore, although the formal literature explores different types of nuclear strike options, all players have identical alternatives and logics that drive them in each of the models (Powell 1989a; 1989b; Wagner 1991). Thus, the literature has failed to explore how states in asymmetrical environments have different incentives for using nuclear weapons. Beyond nuclear scholarship, there is a large body of formal literature on conflict dynamics and crisis bargaining (Fearon 1995; Powell 2002). However, these studies have not separated the conventional and nuclear levels, hence overlooking the special characteristics driving the incentives for using nuclear weapons.

The current paper adds to the literature by developing a formal model exploring how asymmetry at the conventional and nuclear levels rationalizes two distinct types of DNFU. The first is a *damage limitation* incentive that arises because advancements in counterforce capabilities and imbalances in nuclear forces create a first-strike incentive for the stronger side. The second is an incentive for *asymmetric escalation*, that is, to escalate a conventional confrontation into a nuclear one (Narang 2014, 19), which is a situation driven by conventional asymmetry.

Like all formal models, this model is not a representation of any specific crisis—it is a stylized representation, detaching itself from historical complexity to try to bring the essence of the situation into greater focus (Powell 1990, 2). This formal approach demands clarity about the assumptions and consistency of reasoning, both of which are especially useful when studying a contrafactual phenomenon, such as DNFU, against a nuclear adversary. Formal modeling helps trace under exactly what types of conditions DNFU becomes rational, along with how the different parameters relate to each other in a more generic, precise, and rigorous way than the existing theoretical work.

The current paper begins by describing the asymmetries driving the potential for DNFU. Second, the formal model is described and analyzed. Third, I briefly discuss how three nuclear dyads—DPRK–US, Pakistan–India, and Russia–US—are all characterized by some degree of asymmetry on the two levels and, thus, pose a risk of DNFU. Finally, I offer conclusions and avenues for future research.

Deliberate Nuclear First Use

I understand DNFU as being the first intentional detonation of a nuclear weapon for military or coercive purposes. The term covers a wide range of potential actions, ranging from purely demonstrative detonations to massive nuclear attacks. Thus, nuclear employment does not necessarily lead to massive destruction. The scale of destruction depends on the nature of the first strike and subsequent escalation dynamics. Although DNFU is partially driven by a preemptive motive, it is distinct from previous deliberations of nuclear preemption because it can be rational, even when a non-nuclear outcome (referring to a peaceful or conventional war outcome) can be reached with certainty. Moreover, the concept diverges from brinkmanship because the decision to launch a first strike is always in the control of at least one of the adversaries.[3]

Nuclear asymmetry and damage limitation

By the term nuclear asymmetry, I refer to a situation in which *only* the stronger actor has two characteristics. First is a *secure second-strike capability*, which is the capacity for a state's nuclear arsenal to absorb a counterforce strike and still be able to strike back and cause unacceptable damage (Long and Green 2015, 38). Second, a *damage limitation capability* is the ability to reduce the damage the opposing force can inflict with nuclear retaliation.

This ability to limit damage is a spectrum that must be evaluated as the balance between an actor's nuclear survivability measures and the damage limitation capabilities of its adversary (Feaver 1992, 179). A state can pursue three strategies for ensuring nuclear survivability: (1) Hardening refers to deploying nuclear forces in reinforced structures that are difficult to destroy. (2) Concealment refers to efforts to stop adversaries from identifying and locating one's nuclear forces. (3) Finally, redundancy is used to bolster nuclear survivability (Lieber and Press 2017, 10, 16–18). Damage limitation capabilities include counterforce capabilities and strategic missile defenses, as well as nonkinetic measures such as civil defense and cyber-attacks. Missile defenses can limit damage by intercepting incoming missiles.[4] Counterforce strikes seek to immobilize or destroy the adversary's nuclear weapons, delivery systems, and communication systems.[5] According to Lieber and Press (2017, 18–32), a revolution in missile accuracy has made hardening less productive and diminished the survivability of fixed nuclear targets. At the same time, concealment is being eroded by the revolution in remote sensing technology (Lieber and Press 2017, 32–46; Long and Green 2015).[6] Moreover, technologically advanced and wealthy countries are more capable of taking advantage of and counteracting these technological developments (Lieber and Press 2017, 10). This reinforces the asymmetrical relationship between more sophisticated and technologically advanced nuclear powers and smaller nuclear powers.

The damage limitation motive is not a prospect of gain but a desire to anticipate the loss from a nuclear first strike by the enemy (G. Snyder 1961, 104; Wagner 1991, 743). Under asymmetric conditions, a strong actor can be tempted to deliberately launch a counterforce strike to limit the damage in a nuclear war—even if a peaceful settlement is achievable. This incentive increases as nuclear asymmetry—understood as the presumed degree to which one can limit damage in a nuclear war—grows relative to the cost of backing down from the crisis. However, such DNFU is only rational if the adversary has a credible nuclear threat; if not, there is no nuclear attack that needs preemption. This demonstrates the necessity of incorporating the weaker side's asymmetric escalation motive into the same model.

Conventional asymmetry and asymmetric escalation

In the current paper, conventional asymmetry refers not only to the degree of conventional inferiority, but also to the degree to which a conventional defeat seems disastrous. A conventional defeat can cause great human and material costs, but maybe more important is that political leaders often suffer and are removed from power, prosecuted, or killed (Lieber and Press 2020, 102).

To threaten with limited nuclear strikes against conventional aggression is appealing for a state that finds its conventional deterrence insufficient.[7] What Narang calls an asymmetric escalation posture is "explicitly design to deter conventional attack by enabling a state to respond with rapid, asymmetric escalation to first employment of nuclear weapons against military and/or civilian targets" (2014, 19). The purpose is to deter conventional war in the first place, but if war breaks out, one relies on employing nuclear weapons to compel the adversary to back down.

For the weaker side, there are two rationales for deliberately using nuclear weapons. First, there is the "use-it-or-lose-it" pressure. This is when an actor fears that if it does not use its nuclear force immediately, it will be disarmed by an opponent. However, if the weak side has no means to limit the damage in a nuclear war, this will rarely be a sufficient cause to employ nuclear weapons first if the option of a peaceful settlement exists. Moreover, this dynamic has been thoroughly examined in the literature.[8] Therefore, the current paper is primarily concerned with the second type of escalatory behavior, which is similar to what is often referred to as a gamble for resurrection (Downs and Rocke 1994). This escalation pathway starts with an asymmetric escalation threat and continues with the deliberate employment of nuclear weapons if the initial coercion fails. Although nuclear escalation in a conventional war is an extreme gamble, it may be the only option available for a conventionally inferior adversary to try to produce an even slightly more acceptable outcome (Lieber and Press 2020, 102–3).[9] Thus, the attractiveness and credibility of this strategy rests on conventional asymmetry and, more precisely, on the degree to which gambling on a nuclear coercive strike is preferable to a conventional defeat (Schelling 1966, 103).

The weak actor's incentive to employ nuclear weapons in this manner can be instrumentally driven and aimed at improving one's chances of winning, avoiding losing, or ending the conflict more quickly (Smoke 1977, 24; Wagner 1991, 743). Alternatively, it can be coercively motivated, meaning that one tries to reduce the enemy's incentive to attack or retaliate by bolstering or reestablishing one's deterrence (Bruusgaard 2020, 7). Here, one intends to influence the opponent's strategic cost–benefit analysis by signaling a willingness to fight a nuclear war. One does this via an implicit or explicit threat of more damage if the adversary does not back down. It is not the destruction of the initial attack that is important; rather, it is the signal of one's willingness to escalate further, if necessary (Jervis 1979, 628; Schelling 1966, 105).

A model of Deliberate Nuclear First Use

This formal model is a stylization of DNFU in an *ongoing crisis*, which is characterized by some degree of asymmetry on both the nuclear *and* conventional levels, here favoring the strong state (S). The possibility of DNFU is explored in the context of a crisis because it is a more likely scenario than a "bolt from the blue" (Gerson 2010, 35). A crisis is a period of heightened tension and risk of military confrontation—such as disputes over territorial claims, removal/dismantlement of threating weapons, or demands to change a state's behavior. The model does not directly address the selection effects into crisis (Fearon 2002), although it explores whether the actors are willing to successively deepen the crisis instead of yielding.

Model Setup

The model consists of the minimum number of moves that capture the essence of the escalation dynamics. To model S's damage limitation incentive, S must have at least one move prior to the weak state's (W's) last move. Moreover, because W's nuclear weapons are used for coercive leverage during a crisis, W must have at least one possibility before S's last move to signal its willingness for asymmetric escalation. Accordingly, the model must consist of at least three moves. Decisions are made sequentially, as follows:

- Stage 1: W starts with three options. It can back down (Q), in which case S prevails (P) and W loses (L). Alternatively, W can launch an asymmetric escalation (AE) strike. Finally, W can deepen/continue (C) the crisis through political commitments and/or visible war preparations, such as troop mobilization, visible preparations of military capabilities, demonstration strikes, or tests. These actions do not inflict severe costs on the adversary; they are primarily meant to signal a willingness to fight.
- Stage 2: If W continues, S can submit to W's challenge (Q), resulting in S loosing (L) and W prevailing (P). Second, S can launch a counterforce (CF) strike.

Alternatively, it can deepen the crisis further by initiating a conventional military attack (C).
- Stage 3: If S initiates a conventional conflict, W has the last move. By virtue of its conventional inferiority, W will quickly lose a conventional confrontation. Moreover, because modern military operations often start with "intense campaigns against command-and-control facilities, leadership sites, and communications hubs" (Lieber and Press 2020, 102), any such operation will resemble the beginning of an invasion or decapitation campaign. Moreover, as central decision and communication nodes are targeted, leaders will feel the need to respond quickly before they lose their capability to do so by either gambling on AE or fighting a conventional war. A peaceful settlement in the midst of what looks like a massive military campaign seems unlikely. To model this, I simply assume that, by using conventional forces, S leaves W with the options of launching an AE strike or continuing (C) the conventional war.

The outcome of nuclear war is treated as a costly nuclear lottery. The expected value for the state initiating a nuclear war is labeled F (first strike), and the victim's valuation for suffering such an attack is labeled D (disaster). D is "just" a representation of the value that the players place on being the victim of the nuclear first strike. F is derived as a probability distribution over two outcomes. First, P corresponds to prevailing, without suffering any nuclear retaliation. The undesirable outcome, R, represents the value of suffering from the enemy's (nuclear) retaliation strike after one's first strike. The probability distribution between these outcomes is defined by the parameter $p \in (0, 1)$. Because it is difficult to win a nuclear war in military terms (Sechser and Fuhrmann 2013, 177–78), prevailing in a nuclear war is traditionally considered to be about winning by persuasion, a determination to stay in the fight, and the threat of pain not yet inflicted (Schelling 1966, 6–18). Accordingly, p represents the probability that the victim of a first strike will back down immediately, and $1 - p$ represents the probability that the victim of the first strike will retaliate. F is derived as follows: $F_i = p_i P + (1 - p_i)R$.[10]

The two players have different incentives for retaliating. S can limit the damage associated with the ongoing nuclear war. This can make continuing the nuclear war preferable to submission. W cannot limit damage, but it can increase its coercive leverage by retaliating in anticipation that S will eventually back down from the nuclear war. However, W's nuclear inferiority suggests that such an escalation spiral cannot go on for long—or maybe at all—before its nuclear assets are exhausted. Nonetheless, after suffering a first strike, W is already in its worst possible outcome, D. Accordingly, W may be in a position where it has nothing left to lose, thus being tempted to employ its residual nukes, regardless of the possibility of counterretaliation.

I assume that the superior actor wins a conventional war with certainty. This is an unrealistic assumption, but it brings attention to how a conventional defeat can rationalize DNFU.[11] By prevailing in a conventional war, S receives the payoff P. W's valuation of the conventional defeat outcome is denoted as CD. However, getting one's way by fighting a war is not the same as prevailing if the adversary submits (Fearon 1995). Thus, the value of a conventional war is discounted by the constant factor $\delta\,(0<\delta<1)$. This factor represents the costs associated with suffering conventional attacks (Wagner 1991, 732–34). By fighting a conventional war, W receives δCD, while S collects a winner's payoff, δP. These payoffs decline as the discount factor decreases; thus, a low δ signifies a high cost of fighting.

Likewise, if W escalates to the nuclear level at the last stage, its payoff is discounted by the same factor, δ. It does not matter whether this is just the cost of S's conventional campaign or the deteriorating effect a conventional operation has on W's nuclear first-strike capability (Posen 1991).

Without loss of generality, payoffs are treated as symmetric across players. P and D are the best and worst outcomes for both players and are normalized to 1 and 0, respectively. Accordingly, D is defined as a zero-point payoff, and every payoff is between 0 and 1. W has the following preference ordering: $P>L>CD>R>D$. S's preference ordering is as follows: $P>L>R>D$. The notations are summarized in Table 1, and the game is illustrated in Figure 1.

Table 1. Notations and definitions.

Notation	Definition
W	The conventional and nuclear inferior state
S	The conventional and nuclear superior state
AE	Asymmetric escalation
CF	Counterforce strike
C	Continuing/deepening the crisis
Q	Backing down from the crisis
P	Utility of prevailing (P = 1)
D	Utility of being victim to a nuclear first strike (D = 0)
R	Utility of being subject to a nuclear retaliation after one's own first strike
L	Utility of backing down
CD	W's utility for the conventional defeat outcome
F	Expected utility for initiating a nuclear war $F_i = p_iP + (1 - p_i)R$
p_i	Probability that the adversary backs down after suffering a nuclear first strike
$1 - p_i$	Probability that the adversary retaliates after suffering a nuclear first strike
δ	Factor by which payoffs are discounted after suffering conventional military attacks
h	Probability that W perceives a conventional defeat as an existential threat
h'	Updated probability that W perceives a conventional defeat as an existential threat
b	Probability that S is a counterforce hawk
ε_i	Difference between the valuation for backing down/conventional defeat and suffering nuclear retaliation

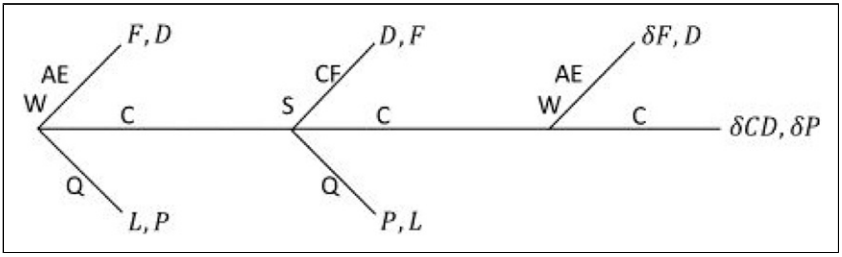

Figure 1. Complete information game.

Conventional and Nuclear Asymmetry: Different types

Based on the degree of conventional asymmetry, W either perceives a conventional defeat as *existential* or *tolerable*. Substantively, we can think of this as the degree to which a conventional defeat is believed to lead to regime or state collapse or not. The *existential* type's utility for a conventional defeat is $CD_e = R + \varepsilon_e$, where ε_e represents the incremental difference between the valuation of suffering a retaliation strike compared with a conventional defeat.[12] The *tolerable* type perceives a conventional defeat as being more acceptable. Its utility for a conventional defeat is $CD_t = R + \varepsilon_t$, where ε_t represents the larger difference in the valuation between the two outcomes.[13] Here, $\varepsilon_t > \varepsilon_e$ implies that $\frac{\varepsilon_t}{1-R} > \frac{\varepsilon_e}{1-R}$. This suggests that the *existential* type escalates to the nuclear level more often than the *tolerable* one because the difference in valuation between a conventional defeat and suffering nuclear retaliation is smaller. Because of the devastating consequences of any conventional defeat for the weak (Lieber and Press 2020, 102), I assume that the peaceful settlement offered at the first stage is preferable to any conventional defeat; thus, $L > CD_e > CD_t$. Accordingly, for W to prefer its nuclear first-strike option to backing down at its first decision node, p must be even higher.[14]

The uncertainty regarding S's type is based on the level of nuclear asymmetry and, more precisely, on the degree to which S can reduce the cost of suffering retaliation, R, such that a first strike is preferable to yielding. If this is the case, S plays CF if it is sufficiently convinced that the adversary will go nuclear in the final decision node. The type of player S with a more effective damage limitation capability is called a counterforce *hawk* and is characterized by a high valuation of $R_h = L - \varepsilon_h$.[15] Hence, ε_h represents the incremental difference between the valuation for backing down and suffering nuclear retaliation after one's own counterforce attack. This can create a commitment problem for S, on which it cannot credibly commit to not using its counterforce option, which, under extreme circumstances, can induce the opponent to immediately go nuclear (Powell 2006). The other type is a counterforce *dove*. It has a moderate damage limitation capability, thus expecting a higher retaliation cost, $R_d = L - \varepsilon_d$.[16] ε_d represents the *dove's* relatively high differentiation between the two potential outcomes. Because $\varepsilon_d > \varepsilon_h$, the *hawk* will use its counterforce options instead of backing down under more favorable circumstances. That is, because $\frac{\varepsilon_d}{1-L+\varepsilon_d} > \frac{\varepsilon_h}{1-L+\varepsilon_h}$, a *hawk* will go nuclear for lower p-values and for relatively more favorable backing down options compared with a *dove*. In addition, I assume that both types of S prefer the cost associated with fighting a conventional war to the possibility of retaliation after its counterforce attack; that is, $\delta > R_h > R_d$. However, nuclear war can still be preferred to conventional war but only if $p_s > \frac{\delta - R}{1 - R}$.

Four Scenarios with Complete Information

These four types generate four games with complete information. Complete information implies that everyone knows everything about the game and players. Thus, each player knows the adversary's preference regarding losing and going nuclear. I assume a constant and commonly known p-value such that the different types of each player will behave distinctively. Furthermore, the current paper focuses on the scenario in which W always prefers to back down instead of using its nuclear option at the outset, that is, $p_w < \frac{L - R}{1 - R}$. As described above, this scenario is less explored in the literature, and more importantly, it seems unlikely that the weak state, which cannot limit damage, would deliberately initiate a nuclear war when a peaceful settlement is attainable, unless the costs of backing down were *extremely* high.

In these conditions, each complete information game has one unique subgame perfect equilibrium (SPE) (see Table 2). In the *extreme double asymmetry* scenario, W backs down immediately because of S's credible counterforce threat. The hawk's nuclear credibility rests on its damage limitation capability relative to the cost of backing down in the crisis and the credibility of W's nuclear threat. The existential nature associated with a conventional defeat means that W's AE threat is credible. However, this does not influence the outcome because S's counterforce threat deters any initial aggression.

When conventional defeat is tolerable for W and S is a counterforce hawk, W backs down immediately. The counterforce option is not credible here because it is common knowledge that W never goes nuclear. Thus, it is S's conventional option that deters aggression from W in the *counterforce dominance* scenario.

If conventional defeat is perceived as existential for W and S is a counterforce dove, the stronger side is coerced into compliance by the credible threat of nuclear escalation in a conventional war. Actual nuclear employment is not needed in the *conventional*

Table 2. The four crisis scenarios and their SPEs.

	Counterforce hawk	Counterforce dove
Existential defeat	*Extreme double asymmetry* SPE: $\{(Q, AE), CF\}$	*Conventional disaster* SPE: $\{(C, AE), Q\}$
Tolerable defeat	*Counterforce dominance* SPE: $\{(Q, C), C\}$	*Moderate double asymmetry:* SPE: $\{(Q, C), C\}$

disaster scenario. Here, S's reluctance to use its own nuclear option preemptively forces it to submit to avoid suffering a nuclear first strike.

Finally, in the *moderate double asymmetry* scenario, no one prefers to go nuclear over loosing. Here, the nuclear level is inconsequential, and W is deterred to submit immediately because of S's dominance at the conventional level.

These four scenarios show that war never occurs with complete information (see Appendix 1 for more elaborate discussions of these SPEs). This finding is consistent with the results from more general crisis bargaining models (Fearon 1995; Powell 2002). However, when we introduce uncertainty about the type of player one faces, DNFU becomes a possibility.

Facing a Counterforce Hawk or a Dove?

In a model of incomplete information, uncertainty is modeled with a move by nature (N) at the beginning of the game. A dynamic game with incomplete information is solved with a perfect Bayesian equilibrium (PBE) (see Appendix 2 for PBE proofs). In Figure 2, (S) knows its own type, but W is uncertain whether it faces a counterforce *hawk* or *dove*. In the top game, S is a *hawk*, and in the bottom, S is a *dove*. The only difference between the games is S's first-strike payoffs. Nature draws the top game with the probability b and the bottom game with probability $1 - b$. W updates its belief via the Bayes rule wherever possible.[17] I assume a constant p-value such that the types behave in a distinct manner. Table 3 summarizes this game's PBEs.

If we assume that W is known to perceive a conventional defeat as *existential* $p_w > \frac{\varepsilon_e}{1-R}$, such that W goes nuclear in a conventional war, DNFU is possible. This implies that conventional war is effectively deterred by W's credible AE threat.

Beyond this, we can distinguish between two scenarios. First, if $p_w > \frac{L-R}{1-R}$, W prefers a nuclear first strike over yielding. With this assumption, a DNFU can be initiated by W

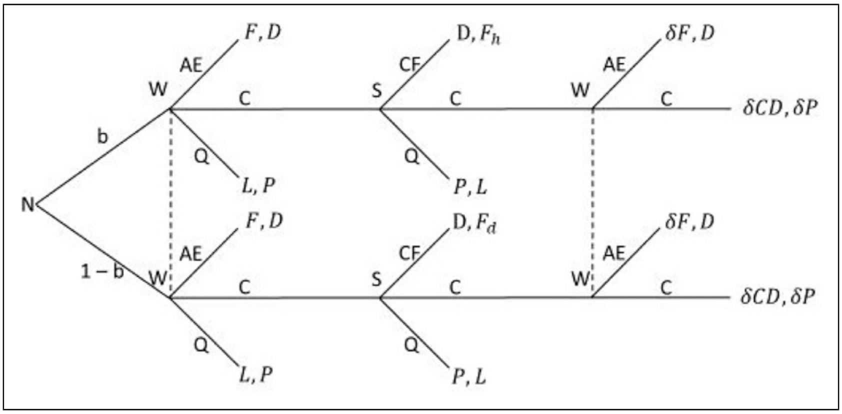

Figure 2. One-sided incomplete information with uncertainty about S's type.

Table 3. PBEs with uncertainty about S's type.

	Strategies	Beliefs
	PBE 1.1: *W is existential and* $p_w > \frac{\varepsilon_e}{1-R}$; *and* $\frac{\varepsilon_d}{1-L+\varepsilon_d} > p_s > \frac{\varepsilon_h}{1-L+\varepsilon_h}$	
W_1	Continues if $L \leq (1-b)$, quits otherwise[20]	b
S_1	Hawk launches CF strike, dove quits	–
W_2	Launch an AE strike	Any b' sustains this PBE
	PBE 1.2: *W is tolerable and* $p_w < \frac{\varepsilon_t}{1-R}$; $\frac{\varepsilon_d}{1-L+\varepsilon_d} > p_s > \frac{\varepsilon_h}{1-L+\varepsilon_h}$; *and* $p_s < \frac{\delta-R}{1-R}$	
W_1	Quits	b
S_1	Hawk and dove continue	–
W_2	Continues	$b = b$'

from the outset, resembling a "use-it-or-lose-it" scenario or by a *hawk* after W's continuation. The logic governing the "use-it-or-lose-it" scenario is that W is sufficiently convinced that S will respond with a nuclear first strike if the crisis continues. Thus, W prefers to gamble on a nuclear first strike from the outset instead of attempting coercion through deepening the crisis using more limited means. On the other hand, the counterforce scenario is possible if W's continuation of the crisis is *not* deterred by the credibility of S's counterforce option. In this contingency, S has a *separating* strategy, where the counterforce *hawk* goes nuclear and the *dove* backs down. Accordingly, if S ultimately turns out to be a *hawk*, it will respond with a CF attack.

In the remainder of the current paper, I focus on a scenario in which W prefers backing down to starting a nuclear war. Because conventional war is deterred by W's credible AE threat, DNFU is only possible through a counterforce attack by the stronger side under these assumptions. The logic governing behavior in *PBE 1.1* is that, based on the relative importance of the issue at hand and the *uncertainty* about the effectiveness of the strong actor's damage limitation capabilities—and protected by its asymmetric nuclear threat—the weak actor can decide to deepen the crisis and hope that its stronger adversary will prefer to submit instead of starting a nuclear war. Hence, W deepens the crisis if its utility for backing down is below its prior belief about S being a *dove*; if not, it backs down immediately. This implies that a sufficiently effective damage limitation capability can neutralize the weaker side's escalation incentive (Kroenig 2018; 2013). Ultimately, S has a *separating* strategy, which means that, in response to initial crisis continuation by W, if S turns out to be a *hawk*, it goes nuclear, and as a *dove,* it backs down.[18] This illustrates a possible pathway to DNFU, which was impossible without asymmetry on *both* levels. Because, without asymmetry on the conventional level, W would not have a credible nuclear option, so S would prefer to start a conventional war instead.[19] Moreover, S would only use its nuclear option if asymmetry on the nuclear level enabled it to reduce the enemy's retaliation capacity to an acceptable level. This illustrates why both levels of asymmetries must be considered to understand the risks of DNFU.

If W sees a conventional defeat as *tolerable* and $p_w < \frac{\varepsilon_t}{1-R}$, the game becomes predictable. Because it is known that the *tolerable* type never goes nuclear, S has a *pooling* strategy in which both the *hawk* and *dove* use conventional force if $p_s < \frac{\delta - R}{1-R}$. This is a rather uncontroversial assumption, implying that S prefers winning a conventional war over starting a nuclear war. Thus, in *PBE 1.2*, the weaker party will back down immediately because it knows that S, regardless of its type, will start a conventional war if it continues the crisis, which is an outcome which is always less preferable than submission. Hence, if everyone knows that W's nuclear threat is not credible, S prevails.

Existential or tolerable conventional defeat?

Figure 3 illustrates a situation in which W knows its own type but S is uncertain about whether it faces an *existential* or a *tolerable* type (see Appendix 2 for PBE proofs). S updates its belief via Bayes' rule upon observing the crisis deepening. In the top game, a conventional defeat is perceived as *existential*, and in the bottom, it is perceived as *tolerable.* Nature draws an *existential* type with probability h and *tolerable type* with probability $1 - h$. The only differentiation between the games is W's value for a conventional defeat. This game's PBEs are summarized in Table 4.

If $p_s > \frac{\varepsilon_h}{1-L+\varepsilon_h}$, such that S is a known counterforce *hawk*, the crisis ends immediately, regardless of the credibility of W's AE option. The logic underlying the *PBE 1.3* is that, under any contingency, backing down is the worst option, in expectation, for a hawk. Even though the credibility of the AE threat influences the hawk's equilibrium behavior, it does only so by increasing the attractiveness of the counterforce option. The conventional war option is preferred at low credibility levels, and at higher levels of credibility, the hawk will—because it has the capacity to lower the expected cost of a nuclear retaliation strike—prefer the CF option over submission. This implies that W does not have any incentive to stay in the crisis because its only motivation to continue

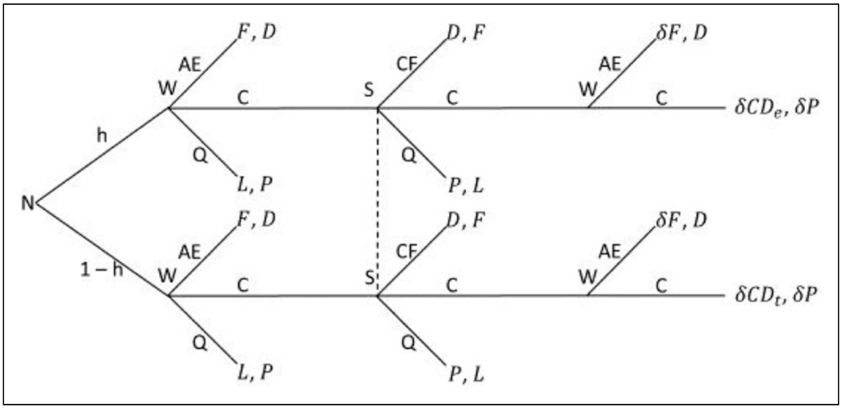

Figure 3. One-sided incomplete information game with uncertainty about W's type.

Table 4. PBEs with uncertainty about W's type.

	Strategies	Beliefs
	PBE 1.3: S is a counterforce hawk and $p_s > \frac{\varepsilon_h}{1-L+\varepsilon_h}$; and $\frac{CD_t - R}{1-R} > p_w > \frac{CD_e - R}{1-R}$	
W_1	Existential and tolerable quite	–
S_1	Continues if $(1 - h') > \frac{F_h}{\delta}$, launches CF otherwise	Any h' sustains this PBE
W_2	Existential launches AE strike, tolerable continues	–
	PBE 1.4: S is a counterforce dove and $p_s < \frac{\varepsilon_d}{1-L+\varepsilon_d}$; and $\frac{CD_t - R}{1-R} > p_w > \frac{CD_e - R}{1-R}$	
W_1	Existential escalates, and tolerable continues with probability $\sigma_w = \frac{hL}{(L-\delta)(h-1)}$, and quits with probability $1 - \sigma_w$	–
S_1	Continues with probability $\sigma_s = \frac{P-L}{P-\delta CD_t}$, and quits with probability $1 - \sigma_s$	$h' = \frac{\delta - L}{\delta}$
W_2	Existential launches AE strike, tolerable continues	–
	PBE 1.5: S is a counterforce dove and $p_s < \frac{\varepsilon_d}{1-L+\varepsilon_d}$; $\frac{CD_t - R}{1-R} > p_w > \frac{CD_e - R}{1-R}$; and $(1 - h) < \frac{L}{\delta}$	
W_1	Existential and tolerable continue	–
S_1	Backs down	$h = h$'
W_2	Existential launches AE strike, tolerable continues	–

is to coerce the adversary into submission; this is a possibility that is discarded by the adversary's damage limitation capability. Thus, W has a *pooling* strategy in which both types immediately back down because backing down is always preferable to suffering a counterforce attack or conventional war.

This suggests that if S is known to never back down, W has no incentive to remain in the crisis, regardless of the credibility of its nuclear option. As such, a near-splendid first-strike capability can deter the adversary into submission, even against an adversary with a credible AE threat. This scenario demands an extremely effective damage limitation capability. Therefore, it is most relevant in dyads consisting of a new and less-sophisticated nuclear state—where survivability measures are underdeveloped and/or delivery systems are not reliable or scarce—facing a nuclear superpower. The DPRK–US dyad is probably the most representative nuclear dyad for this type of reasoning, given the enormous nuclear asymmetry. Moreover, this can be a relevant equilibrium whenever a new nuclear state emerges and it faces a sophisticated nuclear adversary. One impending example could be the Iran–US dyad.

If S is a nuclear dove and $p_s < \frac{\varepsilon_d}{1-L+\varepsilon_d}$, S prefers backing down instead of launching a counterforce strike. This assumption produces the *semi-separating PBE 1.4*, which includes the potential for deliberate AE in a conventional war. The possibility for coercive success for W creates an incentive for the *tolerable* type to imitate an *existential* type by deepening the crisis, even if it is not actually willing to employ nuclear weapons in a conventional conflict. This bluffing motivation is driven by the dove's predisposition to back down if it is sufficiently convinced that a conventional defeat is perceived as existential by W. Consequently, W has a *semi-separating* strategy at its first node, where the *existential* type plays a *pure* strategy of always playing C, and the *tolerable* type plays a *behavioral* strategy, where it sometimes backs down and sometimes continues.

This entails that, although S's belief that W is *existential* increases upon observing continuation, S remains uncertain about what type it is facing. Consequently, the *dove* can decide to start a conventional war instead of yielding, if the cost for backing down relative to the credibility of the AE threat is sufficiently high, even after the weak actor's attempt to signal its determination to go nuclear. This potential conventional escalation can trigger a nuclear response, but only if W turns out to perceive a conventional defeat as *existential*. As conventional asymmetry increases, the incentive for AE grows. Accordingly, conventional asymmetry can also deter a conventional war in the first place. Still, in a high-stakes crisis for both adversaries, uncertainty about the credibility of the AE threat can tempt the strong side to start a conventional war, despite the risk of AE by the weaker side. Note that this type of DNFU is possible only because of W's bluffing incentive. If the *tolerable* type had no incentive to bluff, S would know that it was facing an *existential* enemy upon observing continuation and, therefore, would always back down.

Finally, we have a *pooling PBE 1.5*, in which both types of W deepen the crisis and the *dove* always backs down. Thus, DNFU or conventional war will never materialize in this equilibrium. The reason for this is that W only adopts this *pooling* strategy if it *knows* that the dove will back down. It can anticipate this response with certainty if the commonly known credibility of its AE threat is below the *dove's* valuation for backing down relative to the cost of fighting a conventional war, that is, $(1-h)<\frac{L}{\delta}$. If this is not true, W knows that S always plays C, meaning that its incentive to remain in the crisis would disappear. We can think of this equilibrium as representing a crisis that is of less importance for S or that it is very likely that W perceives a conventional defeat as *existential*. One example of this can be NATO's reluctance to directly confront Russia in Ukraine, which is a decision influenced by Russia's AE threat.

Empirical Relevance

It is beyond the scope of the current paper to investigate specific crisis behaviors in detail. However, I will briefly explore whether the core assumptions that have gone into this model—that is, the asymmetric conditions—are a central feature in three nuclear dyads. Below, I argue that this double asymmetry is a representative characteristic in three important and very different nuclear dyads: DPRK–US, Pakistan–India, and Russia–US. Accordingly, the model should tell us something about the potential core drivers of DNFU in these dyads.

DPRK—United States

The DPRK–US dyad is extremely unbalanced when it comes to both the nuclear and conventional levels. Relatively early in any conventional conflict on the Korean

Peninsula, the conventional military balance would favor the US and its allies in the region (Hackett and Fitzpatrick 2018). Moreover, Pyongyang knows that the chances of regime survival would be effectively zero in the case of a conventional defeat (Lieber and Press 2020, 107; Panda 2020, 287–88). With regime survival as its strategic core objective, Pyongyang has an enormous incentive to escalate any conventional war to the nuclear level, even if the chance of successful nuclear coercion is small (Lieber and Press 2020, 115–16; Panda 2020, 81–83; 85). This suggests that Pyongyang sees conventional defeat as an *existential* threat.

As of 2021, the DPRK is estimated to possess enough fissile material to produce 40–50 nuclear weapons (Kristensen and Korda 2021c, 23). The DPRK's delivery systems are mostly land-based transport-erector-launchers (TELs). With a series of high-profile missile tests starting in February 2017, Pyongyang has gradually demonstrated a potential capability to reach the entire continental US with its nuclear-capable inter-continental ballistic missiles (ICBMs) (Panda 2020, chaps. 7–8).[21] Subsequently, they began to advance their short- and medium-range systems. They again broke their moratorium on ICBM testing in March 2022.[22] These developments suggest that Pyongyang is striving for a more redundant, mobile, flexible, and usable force, thus increasing its capacity to engage in nuclear coercion.

The US maintains a nuclear stockpile of approximately 3800, where 1800 warheads are deployed and distributed on different land, sea, and air platforms (Kristensen and Korda 2021a). Its nuclear survivability is unquestionable against the DPRK. By relying on its sophisticated counterforce capability in a first strike and its missile defense systems to intercept the residual DPRK forces,[23] the US has a considerable capability to limit damage against the small North Korean nuclear arsenal.[24] Thus, this dyad is also characterized by significant nuclear asymmetry.

During the 2017 crisis, both sides attempted to persuade the adversary about its own willingness to go first. On the American side, the Trump administration was not subtle in emphasizing its counterforce capabilities. For example, Secretary of Defense Mattis reminded the DPRK that the US "possess the most precise, rehearsed, and robust defensive and offensive capabilities on Earth" (BBC News 2017). Such statements were possibly intended to signal that the US resembled a *hawk* willing to preempt in this scenario, thereby persuading Pyongyang to halt its provocations and back down.

As demonstrated by *PBE 1.4*, the DPRK had an incentive to inflate the perception of the *existential* nature of a conventional defeat to compel the US to back down. Pyongyang deepened the crisis by demonstrating its capability to inflict damage, here combined with clear statements trying to bolster their nuclear credibility, for example, by providing statements and pictures visualizing their readiness to hit military bases in Guam (Jackson 2019, 139; Panda 2020, 185).

Although the US could be subject to DNFU pressure in a crisis (Press 2022), American war plans for the DPRK in 2017 focused largely on conventional first strikes and only plans of nuclear escalation if the DPRK had struck first (Panda 2020, 295–96). The military option that received the most attention was the so-called American bloody nose strategy (Riley-Smith 2017), which refers to a limited American attack on certain nuclear and ballistic targets in the DPRK that is meant to demonstrate a willingness to escalate further (Dempsey 2018). The logic was that the DPRK would be deterred from going nuclear and be frightened into denuclearizing. However, given the extreme conventional asymmetry, the DPRK cannot afford to wait if it believes that the US is coming after its nukes or coming at them with massive force (Jackson 2019, 164–65, 197, 207). Therefore, most experts agree that the most probable nuclear employment scenario resembles the AE pathway from *PBE 1.4*. Here, Pyongyang, in the early stage of a conventional conflict, would launch a deliberate limited nuclear strike to convey their willingness to inflict more pain by holding American cites at risk with their ICBMs (Jackson 2019, 164; Narang and Panda 2020; Panda 2020, 292; Press 2022, 11).

Pakistan—India

Dyads with less extreme asymmetries on the two levels are also influenced by these deliberate employment pressures. The India–Pakistan dyad has traditionally been characterized by numerical nuclear parity and conventional imbalance (Narang 2010; Tasleem 2018).

India is believed to hold two to three times the conventional military power compared with Pakistan (Kapur 2009). This significant advantage in manpower and materials would probably be decisive in a conventional confrontation. Thus, Pakistan is facing a conventional superior and nuclear-armed adversary, which, according to Narang (2014, 91), poses "an existential land threat to the state, particularly since the 'core' of Pakistan, Punjab, is extremely vulnerable to Indian land conventional power." Therefore, Pakistan has developed a force structure and command-and-control architecture appropriate for AE (Narang 2014, 83–90); in addition, its nuclear strategy threatens with the early use of nuclear strikes in a tactical environment, countering or stalemating a conventional war with India (Bernstein 2014, 107–8; Narang 2014, 78). Moreover, Pakistan has made it explicitly clear that its nuclear doctrine is to employ nuclear weapons first to confront Indian conventional aggression (Jacob 2018, 205–6; Narang 2010, 58; 2014, 79).

On the nuclear level, experts estimate that India possesses 150 nuclear warheads (Kristensen and Korda 2020) and Pakistan 165 nuclear warheads (Kristensen and Korda 2021d). The conventional wisdom has been that India retains a no-first-use (NFU) policy and a centralized nuclear force that is meant for retaliation (Narang 2014, 94–120). However, Clary and Narang (2019, 7–8) argue that India is gradually carving out its NFU pledge by making statements about preemption while acquiring the capabilities appropriate for a counterforce option against Pakistan, including more accurate and responsive delivery systems, surveillance platforms, and ballistic missile defenses. According to the authors, this potentially provides "India with a limited ability to disarm Pakistan of strategic nuclear weapons."[25]

This implies that India is moving in a more hawkish direction and that we are entering a period characterized by some degree of double asymmetry. If we assume that Pakistan perceives a conventional defeat as *existential*, *PBE 1.1* shows that Pakistan can still behave aggressively in a crisis but only if the probability that India is a *hawk* is lower than its valuation for backing down from that crisis. This leads to the intuitive conclusion that India's ability to deter the initial escalation increases with the probability of it being a *hawk*. Hence, it is also noteworthy that Indian escalation seems to be more deliberate in recent crises compared with its crisis management approach in previous crises (Pegahi 2019, 154). Moreover, some analysts have argued that India's new escalator willingness and nuclear signaling in the 2019 and 2016 crises "reflect India's sustained shift toward a preemptive counterforce posture" (Mansoor Ashraf 2019).

This also suggests that Pakistan will only deepen a crisis in which it has relatively high stakes because there is still a possibility that India will submit. What forces Pakistan to be more restrictive is the risk of receiving a nuclear counterforce strike in return. This illustrates India's motivation to acquire a counterforce option; they hope to get out of the paralysis caused by Pakistan's convincing AE posture. As illustrated by *PBE 1.4* and *1.5*, Pakistan has previously felt bolstered to act aggressively in crises by carrying out subconventional or terrorist attacks against India and seeking long-standing revisionist aims without fearing retaliation (Narang 2010, 64). In isolation, a more hawkish India might imply fewer severe crisis in the South Asian theater. However, as demonstrated in *PBE 1.1*, this also introduces an escalation pathway in which India is the nuclear initiator.

Russia—United States

Russia possesses a large and diverse stockpile of substrategic nuclear weapons, which constitute a flexible and appropriate force for fighting a limited nuclear war (Kristensen and Korda 2021b). The deterioration of Russian conventional forces in the first decades after the Cold War made Russia more reliant on this substrategic nuclear force to deter regional conventional conflict (Bruusgaard 2020). Adamsky writes, "On the regional level, the nuclear arsenal's mission became to deter, and if deterrence were to fail, to terminate large-scale conventional aggression through limited nuclear use on the theater of military operations" (2014, 95). Bruusgaard (2020, 18–27) shows how Russian reliance on the early use of nuclear weapons in conventional conflicts has declined as its conventional capabilities gradually improved from 2010 onwards. Still, Western analysts and policymakers are worried that Russia, because of the dynamics demonstrated in *PBE 1.4*, might employ its substrategic nuclear weapons to manage escalation in a large-scale conventional conflict with the West (Department of Defense 2018; Kofman, Fink, and Edmonds 2020).

Despite the redundant nature of Russia's nuclear forces, Russia has a long-standing concern about its force survivability when facing the combination of American counterforce capabilities and missile defenses (Fink and Oliker 2020, 40). According to Lieber and Press (2006), a US counterforce surprise attack could have a reasonable chance of success in peacetime. However, the prospect of success decreases during a crisis. Thus, although the two superpowers are more or less numerically balanced at the nuclear level, American damage limitation capabilities can create a counterforce temptation in a high-stake crisis, as predicted by *PBE 1.1*. However, this is only the case if the US's cost for backing down from a crisis was exceptionally high and if it was sufficiently convinced that Russia was about to escalate a conventional conflict to the nuclear level. Accordingly, even the nuclear superpower dyad is characterized by some degree of double asymmetry, making DNFU an actual concern. However, given the comprehensive nature of both parties' nuclear and conventional capabilities, such nuclear escalation would probably only be considered in extreme circumstances and only after several rungs of conventional escalation have been excused.

Russia's poor performance in conventional warfare in Ukraine and increasing costs associated with a Russian conventional defeat points to a greater risk of nuclear escalation. Furthermore, Russia arguably draws on the *semi-separating* strategy from *PBE 1.4*. Although the Russian partial mobilization and annexation in September/October 2022 have probably been driven by several military and political objectives (Kirby 2022), we can also interpret them as costly signals to increase the potential ramifications of a conventional defeat—and accordingly an attempt to bolster their AE threat aimed at the US and NATO (Bloomberg 2022; Kremlin 2022). Accordingly, a DNFU is conceivable if Russia (by design) finds itself in a situation in which a conventional defeat is perceived as *existential*. However, this is only possible if the West decides to continue or potentially increase its support for Ukraine because of the high stakes in the crisis relative to its belief about the credibility of Russia's AE threat.

Conclusion

The current paper demonstrates how DNFU can be rational in an asymmetric environment. This is a possibility that has been ignored in the formal literature because of its reliance on the MAD condition.

Two levels of asymmetry make DNFU possible. First, asymmetry in nuclear arsenals and advancements in counterforce capabilities create a temptation for damage limitation for the superior actor. Second, conventional asymmetry generates an incentive for the weaker actor to use its nuclear weapons to coerce the stronger adversary into submission and eventually employ its nuclear weapons to bring a conventional conflict into a stalemate.

Through formal modeling, the current paper traces how and under what circumstances these two types of DNFU are rational. Because all actors prefer at least one alternative outcome to any nuclear war, DNFU is only possible with uncertainty about one's adversary. First, a counterforce strike is possible if the weak side escalates the crisis because it believes that nuclear coercion can succeed and if the stronger adversary turns out to be willing to use its damage limitation option instead of backing down. Second, nuclear escalation by the weak side is enabled by its incentive to signal its determination to go nuclear in a conventional war and the strong side's uncertainty about the credibility of this AE threat. If the stronger actor is not sufficiently convinced of this threat, it can decide to start a conventional war, despite the risk of AE by the weak side.

Because today's nuclear landscape is increasingly characterized by different degrees of asymmetry on both the conventional and nuclear levels, which makes DNFU an ever more pressing concern, these pathways can no longer be ignored. Therefore, we need more research focusing on deliberate nuclear wars. Formal studies should aim to develop more realistic and precise models, building on assumptions representative of today's nuclear landscape. Potential avenues for new research are to expand the levels of uncertainty, give the player more fine-grained options or incorporate brinkmanship dynamics into a nuclear and/or conventional asymmetric model. Moreover, these models should be evaluated by analyzing specific crisis behaviors in greater detail. It is only by understanding the potential avenues of nuclear war that we can make an informed choice of how to avoid it.

The stylized representation in the current paper brings the core drives for DNFU into sharper focus. By tracing how the variations in the different parameters enable DNFU, we can learn about how nuclear war can signify a rational choice in a crisis, despite its horrendous consequences. A better understanding of these pathways will help inform the development and implementation of those policies that are meant to reduce the risk of nuclear war.

This model suggests that more transparency about the preferences and capabilities reduces the risk of DNFU. Moreover, we can infer that we should reduce the degree to which a conventional defeat is perceived as an existential threat. This reduces both DNFU incentives simultaneously. First, this implies that the weak are more willing to sustain a conventional defeat compared with risking a nuclear coercive strike, and the secondary consequence is that this reduces the counterforce incentive at the same time. Still, to reduce the general risk of nuclear war, a policy approach must consider a broader spectrum of escalation pathways, including accidental and inadvertent use, in which DNFU is not the only—or necessarily the most likely—pathway to nuclear war.

Acknowledgments

The author is grateful for the constructive and helpful feedback from Scott Gates, Målfrid Braut-Hegghammer, Kristin Ven Bruusgaard, James Cameron, Henrik Hiim, Jon Hovi and Stephen L. Quackenbush. The author is also grateful for the very constructive advice from the anonymous reviewers. The manuscript benefited from being workshopped in the Oslo Nuclear Project Reading Group at the University of Oslo, at the Nuclear Security Workshop at the Hertie School in Berlin, and at the ISA 2022 Annual Convention.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Even Hellan Larsen https://orcid.org/0000-0003-1443-0020

Supplemental Material

Supplemental material for this article is available online.

Notes

1. For a critique of this approach, see (Zagare 1990).
2. See Narang (2010; 2014), Bruusgaard (2020), Lieber and Press (2020; 2017; 2006), Glaser and Fetter (2016), Long and Green (2015; 2017), and Green (2020).
3. For a conceptual discussion, see Powell (1985).
4. See Baklitskiy, Cameron, and Pifer (2021) and Sankaran and Fetter (2022) on missile defense.
5. For a critical view on counterforce, see Glaser and Fetter (2016) and Snyder et al. (2019).
6. These technological developments make non-nuclear counterforce more plausible (Futter and Zala 2021).
7. For a discussion on whether nuclear weapons can deter conventional aggression, see Lieber and Press (2020, 98–101).
8. See Feaver (1992, 165), Morgan et al. (2008, 43, 86), and Gerson (2010).
9. For a skeptical view on nuclear coercion, see Sechser and Fuhrmann (2013); for a more optimistic view, see Kroenig (2013).
10. This representation is based on Wagner (1991, 733).
11. By not modeling a chance to win a conventional war, we sacrifice some analytical rigor for the sake of parsimony. If the weak side had a winning chance in a conventional war, the attractiveness of this option would grow as this probability increased.
12. The existential type's threshold value: $\frac{(R+\varepsilon_e)-R}{1-R} = \frac{\varepsilon_e}{1-R}$.
13. The tolerable type's threshold value: $\frac{(R+\varepsilon_t)-R}{1-R} = \frac{\varepsilon_t}{1-R}$.
14. $\frac{L-R}{1-R} > \frac{\varepsilon_t}{1-R} > \frac{\varepsilon_e}{1-R}$.
15. The hawk's threshold value: $\frac{L-(L-\varepsilon_h)}{1-(L-\varepsilon_h)} = \frac{\varepsilon_h}{1-L+\varepsilon_h}$.
16. The dove's threshold value: $\frac{L-(L-\varepsilon_d)}{1-(L-\varepsilon_d)} = \frac{\varepsilon_d}{1-L+\varepsilon_d}$.
17. Bayes rule: $\Pr\{A|B\} = \frac{Pr\{B|A\}\,Pr\{A\}}{Pr\{B\}}$.
18. S has no incentive to bluff because it has one move and because W's decision proceeding S's move is indifferent to its belief about S's type. In Appendix 3, I demonstrate that S never bluffs in a four-move game.
19. If $p_s < \frac{\delta-R}{1-R}$.
20. If $p_w < \frac{L-R}{1-R}$.
21. Reports have questioned the reliability of these missiles, especially the re-entry vehicle (Kim and Solovyov 2017). However, US intelligence seemed relatively convinced the re-entry vehicle was good enough to hit the continental US (Panda 2017).
22. See CSIS (2017) for an overview of the DPRK's missile tests.

23. The ground-based midcourse defense (GMD) system is designed to defend the US homeland against a small number of ICBMs.
24. See Lieber and Press (2017) on US's counterforce capability against the DPRK, and see Sankaran and Fetter (2022) on US's missile defense capability against the DPRK.
25. These claims are contested; see Rej (2017).

References

Adamsky, Dmitry. 2014. "Nuclear Incoherence: Deterrence Theory and Non-Strategic Nuclear Weapons in Russia." *Journal of Strategic Studies* 37 (1): 91–134.

Baker, Peter, and Gardiner Harris. 2017. "*Deep Divisions Emerge in Trump Administration as North Korea Threatens War*." New York, NY: The New York Times. https://www.nytimes.com/2017/08/09/us/politics/north-korea-nuclear-threat-rex-tillerson.html

Baklitskiy, Andrey, James Cameron, and Steven Pifer. 2021. "*Missile Defense and The Offense-Defense Relationship*." Deep Cuts Working Paper.

BBC News. 2017. "*North Korea Warned by Mattis Not to Invite 'Destruction'*", BBC News. https://www.bbc.com/news/world-us-canada-40879485

Bernstein, Paul I. 2014. "The Emerging Nuclear Landscape." In *On Limited Nuclear War in the 21st Century*, edited by Jeffrey A Larsen and Kerry M Kartchner, 101–28. Stanford, CA: Stanford University Press.

Bloomberg. 2022. "*Transcript: Vladimir Putin's Televised Address on Ukraine*." Bloomberg.-Com, https://www.bloomberg.com/news/articles/2022-02-24/full-transcript-vladimir-putin-s-televised-address-to-russia-on-ukraine-feb-24

Bowden, Mark. 2017. "*How to Deal With North Korea*." Washington, DC: The Atlantic. https://www.theatlantic.com/magazine/archive/2017/07/the-worst-problem-on-earth/528717/

Bruusgaard, Kristin Ven. 2020. "Russian Nuclear Strategy and Conventional Inferiority." *Journal of Strategic Studies*, 4: 1–33. DOI: 10.1080/01402390.2020.1818070

Clary, Christopher, and Vipin Narang. 2019. "India's Counterforce Temptations: Strategic Dilemmas, Doctrine, and Capabilities." *International Security* 43 (3): 7–52. DOI: 10.1162/isec_a_00340

CSIS. 2017. "*North Korean Missile Launches & Nuclear Tests: 1984-Present*." Missile Threat. https://missilethreat.csis.org/north-korea-missile-launches-1984-present/

Dempsey, Michael P. 2018. "*What If Kim Jong Un Decides to Bloody America's Nose First?*" War on the Rocks. https://warontherocks.com/2018/02/kim-decides-bloody-americas-nose-first/

Department of Defense. 2018. "*Nuclear Posture Review*." Department of Defense. https://dod.defense.gov/News/SpecialReports/2018NuclearPostureReview.aspx

Downs, George W., and David M. Rocke. 1994. "Conflict, Agency, and Gambling for Resurrection: The Principal-Agent Problem Goes to War." *American Journal of Political Science* 38 (2): 362–80. DOI: 10.2307/2111408

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

Fearon, James. 2002. "Selection Effects and Deterrence." *International Interactions* 28 (1): 5–29.

Feaver, Peter D. 1992. "Command and Control in Emerging Nuclear Nations." *International Security* 17 (3): 160–87.

Fink, Anya Loukianova, and Olga Oliker. 2020. "Russia's Nuclear Weapons in a Multipolar World: Guarantors of Sovereignty, Great Power Status & More." *Daedalus* 149 (2): 37–55. DOI: 10.1162/daed_a_01788

Futter, Andrew, and Benjamin Zala. 2021. "Strategic Non-Nuclear Weapons and the Onset of a Third Nuclear Age." *European Journal of International Security* 6 (3): 257–77. DOI: 10.1017/eis.2021.2

Gabuev, Alexander. 2022. "Putin's Doomsday Scenario." The Atlantic. *November* 11, 2022. https://www.theatlantic.com/ideas/archive/2022/11/russia-putin-nuclear-escalation-ukraine-war/672082/

Gerson, Michael S. 2010. "No First Use: The next Step for US Nuclear Policy." *International Security* 35 (2): 7–47.

Glaser, Charles L., and Steve Fetter 2016. "Should the United States Reject MAD? Damage Limitation and U.S. Nuclear Strategy toward China." *International Security* 41 (1): 49–98. DOI: 10.1162/ISEC_a_00248

Green, Brendan R. 2020. *The Revolution That Failed: Nuclear Competition, Arms Control, and the Cold War*. Cambridge: Cambridge University Press.

Green, Brendan R., and Austin Long 2017. "The MAD Who Wasn't There: Soviet Reactions to the Late Cold War Nuclear Balance." *Security Studies* 26 (4): 606–41. DOI: 10.1080/09636412.2017.1331639

Hackett, James, and Mark Fitzpatrick. 2018. "*The Conventional Military Balance on the Korean Peninsula*." International Institute for Strategic Studies.

Jackson, Van. 2019. *On the Brink: Trump, Kim and the Threat of Nuclear War*. Cambridge, United Kingdom: Cambridge University Press.

Jacob, Happymon. 2018. "The India-Pakistan Nuclear Dyad: Strategic Stability." In *The End of Strategic Stability? Nuclear Weapons and the Challenge of Regional Rivals*, edited by Lawrence Rubin and Adam M. Stulberg, 66–88.

Jervis, Robert. 1979. "Why Nuclear Superiority Doesn't Matter." *Political Science Quarterly* 94 (4): 617–33. DOI: 10.2307/2149629

Kapur, S. Paul. 2009. *Dangerous Deterrent: Nuclear Weapons Proliferation and Conflict in South Asia*. Singapore: NUS Press.

Kim, Christine, and Dmitry Solovyov. 2017. "*Seoul Says North Korea Puts Washington in Range, but Needs to Prove Critical Technology*." Reuters, sec. Aerospace and Defense. https://www.reuters.com/article/us-northkorea-missiles-southkorea-idUSKBN1DV3F1

Kirby, Paul. 2022. "*What Russian Annexation Means for Ukraine's Regions*." BBC News, https://www.bbc.com/news/world-europe-63086767

Kofman, Michael, Anya Fink, and Jeffrey Edmonds. 2020. "*Russian Strategy for Escalation Management: Evolution of Key Concepts*." CNA.

Kremlin. 2022. "*Address by the President of the Russian Federation*." President of Russia. http://en.kremlin.ru/events/president/news/69390. (Accessed September 21, 2022)

Kristensen, Hans M., and Matt Korda. 2020. "Indian Nuclear Forces, 2020." *Bulletin of the Atomic Scientists* 76 (4): 217–25. DOI: 10.1080/00963402.2020.1778378

Kristensen, Hans M., and Matt Korda. 2021a. “United States Nuclear Weapons, 2021.” *Bulletin of the Atomic Scientists* 77 (1): 43–63. DOI: 10.1080/00963402.2020.1859865

Kristensen, Hans M., and Matt Korda. 2021b. “Russian Nuclear Weapons, 2021.” *Bulletin of the Atomic Scientists* 77 (2): 90–108. DOI: 10.1080/00963402.2021.1885869

Kristensen, Hans M., and Matt Korda. 2021c. “North Korean Nuclear Weapons, 2021.” *Bulletin of the Atomic Scientists* 77 (4): 222–36. DOI: 10.1080/00963402.2021.1940803

Kristensen, Hans M., and Matt Korda. 2021d. “Pakistani Nuclear Weapons, 2021.” *Bulletin of the Atomic Scientists* 77 (5): 265–78. DOI: 10.1080/00963402.2021.1964258

Kroenig, Matthew. 2013. “Nuclear Superiority and the Balance of Resolve: Explaining Nuclear Crisis Outcomes.” *International Organization* 67 (1): 141–71.

Kroenig, Matthew. 2018. *The Logic of American Nuclear Strategy: Why Strategic Superiority Matters*. New York, NY: Oxford University Press.

Lieber, Keir A., and Daryl G. Press. 2006. “The End of MAD? The Nuclear Dimension of U.S. Primacy.” *International Security* 30 (4): 7–44. DOI: 10.1162/isec.2006.30.4.7

Lieber, Keir A., and Daryl G. Press. 2017. “The New Era of Counterforce: Technological Change and the Future of Nuclear Deterrence.” *International Security* 41 (4): 9–49.

Lieber, Keir A., and Daryl G. Press. 2020. *The Myth of the Nuclear Revolution: Power Politics in the Atomic Age*. Ithaca and London: Cornell University Press.

Long, Austin, and Brendan R. Green. 2015. “Stalking the Secure Second Strike: Intelligence, Counterforce, and Nuclear Strategy.” *Journal of Strategic Studies* 38 (1–2): 38–73. DOI: 10.1080/01402390.2014.858150

Mansoor, Ahmed, and Maimuna Ashraf. 2019. “*India and Pakistan's Crisis Means a New Arms Race*.” Text. The National Interest. The Center for the National Interest. https://nationalinterest.org/feature/india-and-pakistans-crisis-means-new-arms-race-50467

Morgan, Forrest E., Karl P. Mueller, Evan S. Medeiros, Kevin L. Pollpeter, and Roger Cliff. 2008. *Dangerous Thresholds: Managing Escalation in the 21st Century*. 1st ed. Santa Monica, CA: RAND Corporation. www.jstor.org/stable/10.7249/mg614af

Nalebuff, Barry. 1986. “Brinkmanship and Nuclear Deterrence: The Neutrality of Escalation.” *Conflict Management and Peace Science* 9 (2): 19–30. DOI: 10.1177/073889428600900202

Narang, Vipin. 2010. “Posturing for Peace? Pakistan's Nuclear Postures and South Asian Stability.” *International Security* 34 (3): 38–78. DOI: 10.1162/isec.2010.34.3.38

Narang, Vipin. 2014. *Nuclear Strategy in the Modern Era: Regional Powers and International Conflict*. Princeton University Press.

Narang, Vipin, and Ankit Panda. 2020. “North Korea: Risks of Escalation.” *Survival* 62 (1): 47–54. DOI: 10.1080/00396338.2020.1715064

Panda, Ankit. 2017. “*US Intelligence: North Korea's ICBM Reentry Vehicles Are Likely Good Enough to Hit the Continental US*.” https://thediplomat.com/2017/08/us-intelligence-north-koreas-icbm-reentry-vehicles-are-likely-good-enough-to-hit-the-continental-us/

Panda, Ankit. 2020. *Kim Jong Un and the Bomb: Survival and Deterrence in North Korea*. London: Hurst & Company.

Pegahi, T. Negeen. 2019. "From Kargil to Pulwama: How Nuclear Crises Have Changed Over 20 Years." *The Washington Quarterly* 42 (2): 149–61. DOI: 10.1080/0163660X.2019.1626690

Posen, Barry R. 1991. *Inadvertent Escalation: Conventional War and Nuclear Risks*. Ithaca and London: Cornell University Press.

Powell, Robert. 1985. "The Theoretical Foundations of Strategic Nuclear Deterrence." *Political Science Quarterly* 100 (1): 75–96. DOI: 10.2307/2150861

Powell, Robert. 1987. "Crisis Bargaining, Escalation, and MAD." *American Political Science Review* 81 (3): 717–35. DOI: 10.2307/1962673

Powell, Robert. 1988. "Nuclear Brinkmanship with Two-Sided Incomplete Information." *The American Political Science Review* 82 (1): 155–78. DOI: 10.2307/1958063

Powell, Robert. 1989a. "Crisis Stability in the Nuclear Age." *American Political Science Review* 83 (1): 61–76. DOI: 10.2307/1956434

Powell, Robert. 1989b. "Nuclear Deterrence and the Strategy of Limited Retaliation." *The American Political Science Review* 83 (2): 503–19. DOI: 10.2307/1962402

Powell, Robert. 1990. *Nuclear Deterrence Theory: The Search for Credibility*. Cambridge: Cambridge University Press.

Powell, Robert. 2002. "Bargaining Theory and International Conflict." *Annual Review of Political Science* 5 (1): 1–30. DOI: 10.1146/annurev.polisci.5.092601.141138

Powell, Robert. 2003. "Nuclear Deterrence Theory, Nuclear Proliferation, and National Missile Defense." *International Security* 27 (4): 86–118.

Powell, Robert. 2006. "War as a Commitment Problem." *International Organization* 60 (1): 169–203. DOI: 10.1017/S0020818306060061

Powell, Robert. 2015. "Nuclear Brinkmanship, Limited War, and Military Power." *International Organization* 69 (3): 589–626. DOI: 10.1017/S0020818315000028

Press, Daryl G. 2022. "The Deliberate Employment of US Nuclear Weapons: Escalation Triggers on the Korean Peninsula." *Journal for Peace and Nuclear Disarmament* 0 (0): 1–14. DOI: 10.1080/25751654.2022.2062896

Rej, Abhijnan. 2017. "*India Is Not Changing Its Policy on No First Use of Nuclear Weapons*." War on the Rocks. https://warontherocks.com/2017/03/india-is-not-changing-its-policy-on-no-first-use-of-nuclear-weapons/

Riley-Smith, Ben. 2017. "*Exclusive: US Making Plans for 'bloody Nose' Military Attack on North Korea*." The Telegraph, https://www.telegraph.co.uk/news/2017/12/20/exclusive-us-making-plans-bloody-nose-military-attack-north/

Sankaran, Jaganath, and Steve Fetter. 2022. "Defending the United States: Revisiting National Missile Defense against North Korea." *International Security* 46 (3): 51–86. DOI: 10.1162/isec_a_00426

Schelling, Thomas C. 1960. *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.

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

Sechser, Todd S., and Matthew Fuhrmann 2013. "Crisis Bargaining and Nuclear Blackmail." *International Organization* 67 (1): 173–95.

Shapiro, Jeremy. 2022. "*We Are On a Path to Nuclear War*." War on the Rocks. https://warontherocks.com/2022/10/the-end-of-the-world-is-nigh/

Smoke, Richard. 1977. *War: Controlling Escalation*. Cambridge: Harvard University Press.

Snyder, Glenn. 1961. *Deterrence and Defense: Toward a Theory of National Security*. Princeton, NJ: Princeton University Press.

Snyder, Ryan, Benoît Pelopidas, Keir A. Lieber, and Daryl G. Press. 2019. "Correspondence: New Era or New Error? Technology and the Future of Deterrence." *International Security* 43 (3): 190–93. DOI: 10.1162/isec_c_00338

Tasleem, Sadia. 2018. "Pakistan's View of Strategic Stability: A Struggle between Theory and Practice." In *The End of Strategic Stability? Nuclear Weapons and the Challenge of Regional Rivalries*, edited by Rubin Lawrence and Adam N. Stulberg, Washington DC: Georgetown University Press.

Wagner, R. Harrison. 1991. "Nuclear Deterrence, Counterforce Strategies, and the Incentive to Strike First." *The American Political Science Review* 85 (3): 727–49. DOI: 10.2307/1963848

Zagare, Frank C. 1987. *The Dynamics of Deterrence*. Chicago: University of Chicago Press.

Zagare, Frank C. 1990. "Rationality and Deterrence." *World Politics* 42 (2): 238–60. DOI: 10.2307/2010465