Nuclear and Conventional Deterrence in Theory and Practice

Kathryn Cochran[a], Scott de Marchi[b], Max Gallop[c], John Nagl[d]

[a] Owner, Cochran Strategy, katie.m.cochran@gmail.com

[b] Professor, Duke University, scott.demarchi@gmail.com

[c] Senior Lecturer, Strathclyde University, max.gallop@gmail.com

[d] Professor of Warfighting Studies, U.S. Army War College, john.nagl.civ@armywarcollege.edu

This article reflects the opinions of the authors and not those of the United States Army or the United States Army War College. We received much-deserved criticism and advice from Ryan Baird, Gene Downum, David Epstein, Peter Feaver, Francis Gavin, Chris Gelpi, Edward Kaplan, Marc Trachtenberg, Georg Vanberg, and Dom Wright. Errors, of course, remain our own.

Abstract

Here, we focus on one of the most important yet understudied questions in international relations: Are nuclear weapons an effective deterrent that helps to prevent major powers conflict? This has been conventional wisdom and many scholars and policy makers have argued that the absence of global war after World War II has been the result of nuclear deterrence. We challenge this consensus on theoretical and empirical grounds and argue that deterrence strategies that rely on nuclear weapons and MAD are inherently dangerous and rest on assumptions that fail to account for the complexity of real-world crises.

Conversely, the value of United States conventional weapons has been underappreciated and serves in two vital roles -- deterring conflict and enabling flexible punishment strategies. We examine the quantitative literature on nuclear deterrence and find that existing research is inadequate if we want to understand the role of conventional and nuclear weapons in deterrence. We conclude with a series of case studies to illustrate important and underdiscussed facets of deterrence, and discuss the implications for the current Russian invasion of Ukraine.

"Elegance of method is indeed marvelously seductive, even when it is irrelevant or inappropriate to the major problems" --Bernard Brodie.

Introduction

During the Cold War, scholars in international relations focused on interstate conflict and in particular nuclear war. After the collapse of the Soviet Union in the late 1980s, the risk of a global conventional war or a strategic nuclear exchange seemed much less likely and attention shifted to other topics including questions of cooperation in the international system and political economy. Starting in the 1990s and continuing into the early 2000s, two shocks changed focus yet again. Methodological innovations from the credibility revolution[1] and the attacks on September 11, 2001 put counterinsurgency on the front-burner and led to a focus on intrastate conflict. Unsurprisingly, the Russian invasion of Ukraine in 2022 and continuing tensions over Taiwan with China have put large-scale conventional conflict and the possibility of nuclear weapon use back on the table as important topics for research.

Here, we focus on one of the most important yet understudied questions in international relations: has the absence of global war after World War II been the result of nuclear deterrence? Or, instead, has the deterrent value of nuclear weapons been overstated and the risks of strategies that depend on nuclear weapons understated? Understanding the role of nuclear weapons in preventing global war between major powers is enormously important if we wish to create effective deterrence strategies in the present day.

One prominent school of thought on this question emphasizes the deterrent value of nuclear weapons given the unthinkable consequences of mutually assured destruction (MAD).[2] Proponents believe that conflicts like the Russian-Ukraine war are unlikely to escalate to full-scale war between the major powers given the possibility of nuclear weapons use. This tradition, initiated by Schelling and carried forward by researchers

[1] The credibility revolution, which started in economics and spread throughout the social sciences, has led to skepticism about observational studies. Instead, more attention has been devoted to experimental methods. In the context of international relations over the past decade, this has meant that scholars often rely on natural experiments and focus on single-case studies (e.g., regression discontinuity designs to study intrastate conflict in Colombia).

[2] For an overview of the various forms of MAD, see Jervis (1989).

that include Jervis and Waltz, also discounts the usefulness of conventional weapons in deterrence, arguing that "levels of violence… blend into one another":

> "The felt need to be able to deny the Russians an advantage at any level of violence comes in part from seeing each level as separate, from failing to appreciate the multiple ways escalation can occur, and thus downgrading the importance of the chance that a conflict will intensify… Neither superpower can confidently move into an area of significant concern to the other without great risk of incurring very high costs – if not immediately, then as the result of a chain of actions that cannot be entirely foreseen or controlled" (Jervis, 1984, p. 148).

The main finding is that conventional deterrence does not work, states should trust to lotteries where full-scale nuclear conflict is "left to chance", and proliferation should be encouraged. If anything, Ukraine should have been safeguarded with its own nuclear weapons to prevent future conflict.[3] The greater the number of states that possess a secure second-strike, the less likely war will be at any level.

This approach to deterrence is widespread and continues to play a role in debates on strategy. Frank Gavin, writing in *The New Makers of Modern Strategy* (2023), notes that many policymakers and researchers support the use of nuclear weapons for deterrence, pointing to the crucial period in the Cold War when the Soviets were seen to have a conventional advantage: "Nuclear weapons helped solve a difficult strategic problem that confronted the United States during the Cold War -- how to deter and, if necessary, defend against attacks on Western Europe and East Asia, facing an adversary with superior conventional military capabilities, without making conventional military commitments whose costs may have been prohibitive as well as unpopular with the American public." Similarly, given the conventional superiority of the United States in the modern context, states including Iran, North Korea, and Russia would, under this logic, have an incentive to rely on nuclear weapons to deter aggression or invasion.

In contrast, we will argue in this paper that deterrence strategies that rely on nuclear weapons and MAD are inherently dangerous and rest on assumptions that fail to account for the complexity of real-world crises. Our analysis also suggests that the deterrent value of United States conventional weapons has been underappreciated and

[3] See, however, Budjeryn (2016) for the costs to Ukraine that would have been incurred by this strategy.

serves two vital roles. Conventional capabilities and theater level balance of power deter the initiation of conflict and allow for retaliation if conflict occurs (i.e., punishment strategies that do not run the risks of escalation that would be present with nuclear weapons use). Conventional weapons also "slow down" conflicts, allowing actors involved in a crisis to learn about each other and calibrate responses, thereby avoiding uncontrolled escalation.

This paper provides an overview of these two approaches, along with an investigation of the theoretical and empirical research from the quantitative literature on nuclear deterrence. The main finding of these sections is that existing research is inadequate if we want to understand the role of conventional and nuclear weapons in deterrence. We will then explore a series of case studies. Each of these serves to illustrate a key feature about nuclear deterrence which any effective policy must take into account. We conclude with a set of findings on deterrence that should inform policy, highlighting the application of these results to the current war in Ukraine.

Existing hypothesis: nuclear weapons deter conflict

Much of the literature on nuclear weapons owes an intellectual debt to Schelling, who identified the salient question of why the Soviet Union had not engaged in more aggression in the post-war period:

> "Even those who have doubted whether our massive-retaliation threat was a potent deterrent to <u>minor</u> aggression during the last several years but are perplexed that the Russians have not engaged in more mischief than they have, can note that the threat we voiced was backed by an additional implicit threat that we might be triggered by Soviet actions *in spite of ourselves* [emphasis added]" (Schelling, 1959, p. 4).

In Schelling's view, we have avoided large scale conflict in large measure because of the deterrent value of nuclear weapons. According to his logic, nuclear weapons and in particular a secure second-strike capability allow the United States to make threats that enhance deterrence and ensure peace. There are two components of Schelling's logic worth emphasizing. The first is that states depend on *probabilistic* threats to maintain deterrence and demonstrate resolve. Schelling argues that states make retaliatory threats to deter hostility and by design leave "something to chance" – i.e., if you attack

me, I may respond even if I do not want to.[4] States should not seek perfect control over their nuclear forces and instead depend on mechanisms like tripwires to deter aggressors.

The second is that the leaders of states are assumed to be rational and can calculate expected value for threats like the lotteries detailed above. Faced with the possibility (albeit small) of a massive response, aggressors avoid initiating conflict because the gains that might be achieved are offset by the possibility of large losses. In expected value terms, even a small risk of a massive retaliation offsets any gains from aggression.

More recent work has echoed this core logic and reached similar conclusions. Jervis questions the value of conventional weapons for states that hold a secure second-strike capability; in his mind, conventional and nuclear "blend into one another" and there is not much stand-alone value to conventional deterrence when nuclear deterrence is operative (Jervis, 1984, 1989). Waltz (Waltz, 1990; Waltz and Sagan, 2013), takes this logic even further and recommends that we encourage proliferation to ensure peace. If more states have nuclear weapons, then aggression will be less prevalent simply because more states will be able to make nuclear threats that offset potential gains. Like Schelling's original formulation, these arguments depend on the same two assumptions. The first is that the risk of a massive retaliation – which cannot be mitigated by state leaders – deters aggressors:

> "To ask why a country should carry out its deterrent threat if deterrence fails is to ask the wrong question. The question suggests that an aggressor may attack believing that the attacked country may not retaliate. This invokes the conventional logic that analysts find so hard to forsake. In a conventional world, a country can sensibly attack if it believes that success is possible. In a nuclear world, a would-be attack is deterred if it believes that the attacked may retaliate. Uncertainty of response, not certainty, is required for deterrence because, if retaliation occurs, one risks losing so much" (Waltz and Sagan, 2013, p. 23-24).

And, given the rationality of the state actors involved, the use of nuclear weapons will not occur by mistake:

[4] And even when it is no longer subgame perfect to do so.

> "Although some of the new nuclear states may be economically and technically backward, they will either have expert and highly trained scientists and engineers, or they will not be able to produce nuclear weapons. Even if they buy or steal the weapons, they will have to hire technicians to maintain and control them. We do not have to wonder whether they will take good care of their weapons. They have every incentive to do so. They will not want to risk retaliation because one or more of their warheads accidentally struck another country" (Waltz and Sagan, 2013, p. 21).

There are other researchers who have formalized the above arguments (e.g., Powell, 1987, 2015), but the core arguments on threats as lotteries and the presence of rational actors remain the same.

Alternative hypothesis: American conventional superiority deters global conflict

An alternative hypothesis is that the deterrence value of nuclear weapons has been overstated and that these weapons have made the world less safe in two distinct ways. The first is that accidents are possible and that the history of the Cold War between the Soviet Union and United States demonstrates how close the world came to an unacceptable outcome due to this risk. This is the primary focus of non-proliferation advocates. The more nuclear weapons exist in the world, the more likely that something will go wrong, with catastrophic consequences. The second is that states with nuclear weapons have license to engage in conventional conflicts and proxy wars that would be unthinkable without the security guarantee afforded by a secure second-strike capability.[5] The obvious example is today's conflict in Ukraine – Russia would never have attempted this type of aggression without their nuclear arsenal.[6] Both of these mechanisms argue for limiting the spread of nuclear weapons and inform skeptical views of their deterrent value. Sagan's arguments on organization theory and the inherent risks of accidental or uncontrolled escalation are particularly relevant here (Sagan, 1995; Waltz and Sagan, 2013).

[5] McNamara (1983) notes "The threat of such an action [a limited or strategic nuclear strike], therefore, has lost all credibility as a deterrent to Soviet conventional aggression".

[6] It is unlikely that the Russian leadership has ignored modern examples where the United States has been willing to engage in regime change to achieve policy objectives or defend *status quo* borders. And, throughout the conflict, Russia has relied on nuclear threats to deter United States / NATO forces from expanding the scope of their involvement in the war.

The absence of global war may well be in spite of the development of nuclear weapons; instead, the *conventional* weapons supremacy of the United States has been a higher-cost but preferable means for preserving peace. Conventional weapons that achieve a balance of power in theater likely deter initial aggressions. Conventional superiority has also allowed the United States to offer security guarantees to other nations, limiting proliferation of nuclear weapons and reducing the possibility of (accidental) use.[7]

Perhaps more important, conventional superiority affords the United States options to deter conflict or retaliate when conflict occurs without risking escalation to a nuclear exchange. Simply put, conventional superiority allows for less escalatory, more precise, and more credible threats of punishment. As Lieber and Press (2015) point out, conventional conflicts can escalate to the use of nuclear weapons given "the prospect of the noose" for leaders in states like North Korea.[8,9] But the chances of escalation can be mitigated by creating war plans with clear off-ramps for adversaries and limits on escalation. In contrast, in contexts where the United States has lacked conventional superiority (including the critical period from ~1961 to ~1983), the possibility of global war and use of nuclear weapons dramatically increased.

In the remainder of this paper, we will examine the logic of the mechanisms underlying both of these arguments. Has the world avoided global conflict because of the deterrent value of nuclear weapons and a reliance on mutually assured destruction? Or has a *Pax Americana* and the ability of the United States to influence theater level calculations, largely with conventional weapons, been responsible for relatively low levels of conflict? Our focus here is on quantitative studies of deterrence and attempts to provide systematic and generalizable evidence on the role of nuclear weapons.[10] We will cover both the formal theoretic literature and the empirical literature to see how reliable existing findings are on the deterrent value of nuclear weapons. Then, we will detail

[7] See Gavin (2015) for an overview of United States non-proliferation strategy.

[8] The tradeoff between punishment and ending crisis is a more general problem in the literature; see Krcmaric (2018).

[9] It is worth noting that Lieber and Press are less sanguine about the possibility of avoiding escalation than we are. But, regardless of one's priors on this question, the greater the margin of United States conventional superiority, the more control the United States will have in creating war plans designed to avoid escalation.

[10] The promise of the quantitative literature, and the reason we choose to focus on it here, is the claim that it can provide general findings that can guide deterrence policy today. This is in contrast to historiographical work that often focuses on single cases which likely have idiosyncratic features limiting modern application. The question then is whether or not this promise is borne out in the actual research produced by quantitative scholars…

four pivotal cases that allow us to engage with the assumptions underlying each argument. These cases consist of the Cuban Missile Crisis, US war plans to defend Europe against a conventional attack by the Soviet Union during the Cold War, US efforts to limit proliferation, and the current Ukraine conflict.

Problems with Formal Models of Deterrence

Threats that leave something to chance

On August 13, 1961, Berlin was divided by barbed wire that had been erected overnight, but the crisis that led to this event took place over the prior three years.[11] In November of 1958, Khrushchev demanded that the United States leave West Berlin. His ultimatum was predicated on the threat posed by NATO, especially the addition of West Germany to the alliance:

> "West Berlin is there to be used as an attack base against us. They are turning to blackmail. Five years ago—that was different. Then, we did not have the hydrogen bomb; now, the balance of forces is different. Then, we could not reach the USA. The USA built its policies upon the bases surrounding us. Today, America has moved closer to us—our missiles can hit them directly" (Khrushchev, November 10, 1958).[12]

It is worth noting that in the 1950s, American leadership perceived that the Soviets had a massive conventional advantage – possibly even a 10:1 ratio of active strength divisions (Bitzinger, 1989).[13] Thus, from the perspective of the United States, the only real defense if the Soviets took Berlin – or invaded Western Europe – was to acquiesce or to retaliate with nuclear weapons:

---

[11] This section on the Berlin crisis borrows heavily from the excellent treatment in Trachtenberg (1991) or Gavin (2012).

[12] Minutes from the Discussion between the Delegation of the People's Republic of Poland and the Government of the USSR/ https://digitalarchive.wilsoncenter.org/document/minutes-discussion-between-delegation-prl-peoples-republic-poland-and-government-ussr-25. Khrushchev's logic is clear. When the United States held a massive nuclear advantage and secure second-strike capacity, Soviet options were limited. But, with the development of near-peer Soviet capabilities, they were free to use their conventional leverage.

[13] It is worth noting that this assessment was wrong and that many of the Soviet divisions were understrength or existed only on paper. Simultaneously, the Soviets were overestimating the United States and NATO. See, Vernon (1981) and Bitzinger (1989).

> "Eisenhower himself was one of the last true believers [in massive retaliation]. To him, even as late as 1960, a major U.S.-Soviet limited war was an absurdity. His 'basic philosophy,' he said during the Berlin crisis in 1959, was that America had 'to be willing to push its whole stack of chips into the pot' when such became necessary" (Trachtenberg, 1989, p. 330).

Schelling, along with other theorists, realized the inherent problem in this approach to security. Using West Germany as an example, if the Soviets acted first and seized control with conventional forces, would the United States at that point follow through on the threat of nuclear retaliation? Communication on exactly this point between Kennedy and de Gaulle in early June of 1961 is revealing.[14] de Gaulle raised the famous question "would we [the United States] be ready to trade New York for Paris?" His point was that this was unclear to the French government and thus was also likely to be a matter of some considerable ambiguity to the Soviets.

To demonstrate resolve, de Gaulle's advice to Kennedy was to create a red line on the status of Berlin and make clear that any abridgement of the *status quo* by the Soviets would mean "general war." But this approach to security made very little sense to theorists like Schelling. Once Berlin – or even the entirety of West Germany – was in Soviet hands, what did the United States or NATO have to gain by escalating? Imagine if the United States were to use nuclear weapons to neutralize Soviet conventional superiority. Even in the best-case outcome where this would not have prompted the Soviets to escalate to a strategic exchange, exercises at the time "made clear that Germany would be devastated" with millions dead (Nichols et al., 2012). War could not be won given mutual vulnerability and the Soviets might rightly conclude that the Americans would not choose to end the world over a *fait accompli*. In more technical terms, it would not be a subgame perfect equilibrium to retaliate with nuclear weapons after Berlin was in Soviet hands.

Thus was born the idea of a threat that leaves something to chance. Schelling, rightly, was interested in "deterrence *ex ante*, not revenge *ex post*" (Schelling, 1960, p. 187). And to create a threat that was plausible, Schelling argued that a state could not immediately spring to the (often completely disproportionate) threat of ending the world with a full-scale strategic exchange. Threats of this magnitude are both costly (especially in terms

[14] https://history.state.gov/historicaldocuments/frus1961-63v14/d30

of foreign relations and perhaps even domestically) and, in many contexts, implausible. For Schelling, the answer was to create a lottery that calibrates the threat:

> "If a hitchhiker pulls a gun on the driver of a car and the driver threats to kill them both unless the hitchhiker throws his gun out the window, making his threat by pressing the accelerator to the floor and creating a manifest risk of fatal accident, there is some chance that the accident will occur before the hitchhiker has a chance to comprehend the threat and comply. In this case, the risk of accidental fulfillment is an integral part of the threat" (Schelling, 1960, p. 183).

As the above example illustrates, the driver does not immediately drive over a cliff; given the costs for both actors, that would be an implausible threat. Rather, the driver creates a situation where a bad end at the bottom of a cliff is a *possibility* but not a *certainty*, which imposes (in expectation) costs the hitchhiker must consider. Best of all, if the driver conveys this strategy to the hitchhiker before the gun is pulled, a rational hitchhiker might choose to avoid conflict altogether.

This idea of threats that impose risks, and the closely related concept of demonstrating resolve, has permeated the theoretical literature in international relations. For example, Powell's (1987) early work on deterrence depends on the same type of lottery. State leaders have strategy sets that allow for exploiting an "opportunity" (or crisis) by escalating, quitting the crisis, or launching a full-scale nuclear attack. Crucially, however, at each stage where leaders can escalate, there is a move by nature where "an autonomous risk" occurs that may send the crisis "out of control". Notably, the outcome of this action is exactly the same as if either leader pursues a full-scale nuclear attack. In other words, leaders demonstrate resolve by repeatedly pulling the trigger in a game that is nearly identical to Russian roulette. The only differences are that in Powell's version of Russian roulette, the outcome of a "massive nuclear attack" is treated as an undefined constant and both players suffer (i.e., die) if the lottery comes out poorly.

More recent theoretical work continues in this tradition. Debs (2022) attempts to show how nuclear states make threats by "burning bridges" to demonstrate resolve.[15] In his

[15] In this treatment, burning bridges can be a certainty or a probabilistic threat. Oddly, though, these threats are not treated as strategic choices and are more of a *deus ex machina* in the model. It is extremely difficult to imagine any real-world application of the model as a result.

model, aggressors "trigger" an interstate crisis and demand a more favorable outcome than the *status quo* peace provides. The possibility of a nuclear war *should* incline states to back down from such crises (much less initiate them) and this represents a theoretical puzzle for Debs, who asks how the aggressor can force an adversary to comply instead. To allow for the possibility of compellent threats, Debs relies on two mechanisms advanced by Schelling. The first is to assume that aggressors can somehow burn all bridges (removing their own ability to back down from a crisis) and present their opponent with a choice between complying or running the risk of nuclear war.[16] The second is to assume that states may have a taste for (nuclear) war by adding a "hostility" parameter to their utility. With the addition of "irrational" preferences, states can then demand more favorable settlements because their adversaries recognize that the consequences of conflict do not deter them.[17] Unlike Powell's model, Debs doesn't present us with a game of Russian roulette. Rather, with the strong assumptions about the ability of state leaders to burn bridges and engage in "madman" diplomacy, he is modeling an even starker game. Leaders that initiate conflict demand concessions; if these are not provided, they are committed to nuclear war. By assumption, they somehow limit their ability to back down if their adversary refuses their demands – even if they might want to. How such a mechanism would work in practice or whether such a threat would be believable in the first place is ignored.[18]

---

[16] There is a substantial alteration, however, of Schelling's logic on this point as Debs notes in the introduction to his paper. Schelling confines burning bridges to states engaged in defense. In his mind, this is far more plausible than states burning bridges offensively:

> "To deter an enemy's advance it may be enough to burn the escape bridges behind me… To compel an enemy's retreat, though, by some threat of engagement, I have to commit to move. This requires setting fire to the grass behind me as I face the enemy, with the wind blowing toward the enemy… The assurances that accompany a compellent action – move back a mile and I won't shoot (otherwise I shall) and I won't then try again for a second mile – are harder to demonstrate in advance, unless it be through a long past record of abiding by one's own verbal assurances (Schelling, 1966).

[17] State leaders in this model can also "extract concessions by appearing irrational"; Debs provides the standard example of Nixon and his "madman" diplomacy to illustrate this mechanism. This is, however, a misreading of the historical record. While Nixon (and Kissinger) believed in using nuclear threats as signals, most notably with Operation Giant Lance, they "did not treat this nuclear alert as a competition in risk taking. Instead they attempted to make this an immaculate, risk-free alert operation, ordering the US military to not take any provocative actions or threatening moves against the Soviets and ruling out some specific operations, such as increases in peripheral reconnaissance, which they feared might lead to an incident" (Sagan and Suri, 2003). Ironically, Nixon's perpetual concerns about secrecy further undermined the plausibility of his nuclear threats and as a result, these were not generally seen as credible by the Soviets (see Burr and Kimball, 2015, for a broader overview).

[18] Burning bridges or creating threats that leave something to chance present substantial problems of mechanism design and it is not obvious how leaders would implement these strategies. If one burns all bridges to compel an adversary, how does one then "unburn" these bridges once the adversary concedes? If a state leader wanted to create a doomsday device outside of rational control, how would this loss of control

Finally, Larsen (2023) goes further and explores the conditions under which "deliberate nuclear first use can be rational". In his model, a weaker state faces an adversary that possesses stronger conventional and nuclear forces. For Larsen, this creates the possibility of a "use it or lose it" scenario for the weaker state given the stronger state can engage in counterforce strikes or force regime change. To generate the counter-intuitive result that weaker states like North Korea will engage in nuclear first use, Larsen depends on two assumptions. First, a lottery is embedded into the game: launching a nuclear attack has a fixed probability of rewarding the weaker state with maximum utility (i.e., the state prevails) or instead a retaliatory strike (i.e., the state suffers a fixed loss to utility).[19] The odd choice here is that this lottery allows for the possibility that a stronger state that is attacked with nuclear weapons might not retaliate in kind because of the Schelling-style threat of "pain not yet inflicted". Second, the payoffs for the weaker state contemplating nuclear first use are chosen such that there is little difference between a conventional defeat and a nuclear conflict. Together, these choices, like other models in this literature, make nuclear conflict "rational" by mitigating the costs of nuclear warfare and providing incentives to engage in nuclear lotteries.

The main issue, however, with these approaches – from Schelling to more modern theorists – is that they rest on the assumption that nuclear war is not some terrifying event. While nuclear war is assumed to be worse than other outcomes (i.e., prevailing > conceding > nuclear conflict), the differences between these outcomes are treated as a fixed constant, which masks the quite different nature of nuclear warfare. Even in his most recent work on nuclear brinksmanship, Powell (2015) assumes that "the additional cost of an all-out nuclear war" is a fixed and limited value.

This is, of course, absurd.[20] A nuclear war between states with a secure second-strike capability does not, as theorists assume, present actors with an outcome that is somewhat less valuable than a concession. An all-out nuclear war is possibly *world ending*. And in expectation, any small probability of ending the world has an expected

be implemented? And how would this leader convince adversaries that no "secret" backdoor control was possible?

19 In the payoff functions, suffering a nuclear first strike has a utility of 0; other outcomes (including suffering a retaliatory nuclear strike) are assumed to be more beneficial than this and are placed on the interval [0..1]. This formulation, obviously, allows for the main finding of the paper, which is that deliberate first use of nuclear weapons can be "rational".

20 The reason that this assumption is relied on is because it allows theorists to derive equilibrium outcomes that differ from "avoid all conflict".

payoff so negative that it is difficult to conceptualize.[21] Worse, if every threat utilized by a state depends on "leaving something to chance" or irrationally ignoring the costs of nuclear war (which leads to playing mixed strategies that place positive weight on beginning a nuclear war), this means that leaders are essentially playing Russian roulette over and over again to demonstrate resolve without a proper appreciation for what this means over multiple crises. At some point, the game will end very badly for everyone.

Counter-arguments to this position tend to focus on the fact that nuclear weapons a) have actual deterrent value b) have not been employed after the Second World War and c) the use of (tactical) nuclear weapons could potentially remain limited. These counter-arguments miss the main point, however. To be blunt, deterrence strategies that depend on Vasili Arkhipov casting the deciding vote to avoid using a 5-kiloton nuclear weapon and possibly triggering a more general nuclear exchange are foolhardy.[22] Similarly, if the United States had used (tactical) nuclear weapons to deter a Soviet invasion of West Germany, the strategic bet would have been that the chance of escalation by the Soviets to a full-scale strategic exchange is exactly zero.[23] Nearly every analyst would agree that the Soviets might have escalated; expected value calculations must consider these outcomes because even if they are seen as unlikely, the consequences are grave.

---

[21] Typically, theorists rely on Von Neumann and Morgenstern payoffs, where infinite values are excluded by assumption. In the case of nuclear war, the distinction between a payoff of negative infinity or a value that is very, very bad is not salient. Models that choose to mitigate by design the difference between nuclear and conventional war or simply leave this as a free parameter ignore the reality of a full-scale strategic exchange: life on the planet might cease. Whatever value one assigns to that potential outcome, it is negative enough to swamp other considerations in these models.

[22] Ironically, to the extent that norms against nuclear weapon use exist, it was relatively low-ranking Soviet officers, often against orders, that were pivotal in avoiding nuclear war. In addition to the Cuban Missile Crisis example, Colonel Stanislav Petrov's actions during Able Archer helped avoid nuclear war (https://www.smithsonianmag.com/history/the-1983-military-drill-that-nearly-sparked-nuclear-war-with-the-soviets-180979980/).

[23] Implicitly, supporters of nuclear deterrence are calculating what they see as the most likely outcome and ignoring less likely but very destructive outcomes. E.g., if the United States had used nuclear weapons in a limited capacity to defend the Fulda Gap, the strategic bet is that this would not have prompted a full-scale strategic exchange. The issue, of course is that:

$$E(\text{outcome} \mid \text{limited use}) = (1-p)\,(\text{limited exchange}) + p(\text{full-scale strategic exchange}).$$

One has to consider the fact that p, even if small, is greater than zero and given the magnitude of the consequence, is consequential for decision-making.

A fair assessment is that nuclear weapons provided deterrent value, but at a substantial cost *in expectation*. In other words, we got lucky and at multiple points during the Cold War there was a non-zero chance of a disastrous outcome. It is also worth noting that the United States and Soviets had substantial communication during the Cold War. In future contexts where nuclear weapons are relied on for deterrent value, we cannot count on the same levels of communication. The Chinese leadership, for example, does not pick up the phone.[24]

Not so rational actors

Misunderstanding the expected value of threats that involve nuclear war is alarming, but the applicability of formal models to real-world deterrence worsens when one considers the rational actor assumption relied on by theorists. All work in the formal theory tradition (as well as a substantial amount in more qualitative policy approaches) depends on this assumption and the best case that can be made for it is that state leaders may be quite sophisticated. As noted by Keohane (1999), state leaders are typically well-informed and their decision-making is aided by professional staffs of subject matter experts. That said, it is worth asking if a "one-size fits all" approach to behavior is a useful simplification or instead a dangerous assumption that masks variance across state leaders.[25]

Eisenhower, as noted above, had a very different attitude toward the use of nuclear weapons than Kennedy. This points to the fact that considerable behavioral variance exists even within a single case (i.e., the United States during the early part of the Cold War). If one broadens to other, less homogeneous cases, variance increases even more. In our case study on the Cuban Missile Crisis (see below), Castro's behavior is a worrisome exemplar that we should not trust the rational actor assumption, given that

[24] https://www.cbsnews.com/news/china-spy-balloon-beijing-us-crisis-phone-hot-line-dangerous-communications-gap/

[25] One can, of course, diversify the types of leaders that exist by adding parameters for their "taste" for conflict (per Debs) or assuming that leaders might be counterforce hawks or doves (per Larsen). To the extent that these parameters / types drive results in deductive models, one has to ask whether they have much verisimilitude. And, to the extent we wish to pin down values to arrive at more precise statements on policy, we would have to engage in what is essentially a psychological investigation of world leaders (which is compounded by the fact that they might misrepresent … everything).

his preferred outcome was a strategic nuclear exchange that would have annihilated his own nation.[26,27]

Schelling, like most economists, relied on this assumption because it allowed for the neat development of theory (Desch, 2019). But assuming that all state leaders will approach threats (and the lotteries these involve) in the same way does not fit the historical record. Real-world behavioral variance, even for elite actors, is prevalent in complex decision-making environments (de Marchi and Laver, 2023) and assuming away this variance places the main results of the theoretical literature on nuclear deterrence on a shaky foundation. In the case of nuclear weapons, we must also consider the organizations (and their idiosyncratic cultures) which are responsible for their use (Kaplan, 2015).

The core issue, then, is to know what type of leader and organization one is facing and to predict their behavior – assuming economic rationality sidesteps the core difficulty any model of deterrence must deal with:

> "The question of what deters Soviet aggression is an extremely difficult one. To answer it, we must put ourselves in the minds of several individuals who would make the decision to initiate war" (McNamara, 1983, p. 72).

As the current crisis in Ukraine illustrates, we have substantial intelligence – due to both the lengthy tenure of Putin and our even lengthier history with Russia – on Russian decision-making. This, no doubt, has helped the Biden administration calibrate our responses, deter (tactical) nuclear weapon use by the Russians, and prevent further escalation.

But consider how much less we know about other state leaders that possess nuclear weapons -- including the behavior of the leadership of North Korea and China if either

[26] One argument could be that Castro had more of a commitment to ideology than preserving his nation. To the extend we allow utility formulations where state leaders have a taste for nuclear war, this undermines the formal theoretic tradition. Instead, to the extent that one believes that "odd" utility functions of this type exist, one is engaged more in a behavioral study of an individual leader than a strategic study of deterrence more broadly.

[27] For a more modern case on evolving behavior by the Russian leadership today, see Kendall-Taylor, et al. (2023).

state became involved in a crisis with the United States. If we approach every crisis in exactly the same fashion and assume that the same deterrent threats will work equally well in all contexts, and we do not have good intelligence *ex ante* on the type of actor we are facing, the consequences could be disastrous.

Norms

There is, however, one additional argument that Schelling and more recent scholars depend on, which is that institutions (i.e., norms) also reinforce the above logic concerning the limited use of nuclear weapons. A feature of nuclear weapons that enhances deterrence and reduces risk of an accident is that they are *different* by tradition than conventional weapons:

> "In fact, what we are dealing with in the analysis of limited war is tradition. We are dealing with precedent, convention, and the force of suggestion. We are dealing with the theory of unwritten law – with conventions whose sanction in the aggregate is the need for mutual forbearance to avoid mutual destruction, and whose sanction in each individual case is the risk that to break a rule may collapse it and that to collapse it may lead to a jointly less favorable limit or to none at all, and may further weaken the yet unbroken rules by providing evidence that their 'authority' cannot be taken for granted. What makes atomic weapons different is a power tradition that they are different" (Schelling, 1960, p. 260).

This reliance on norms is a curious addition, given the assumption by Schelling and other theorists that actors are purely rational when it comes to threats. Unfortunately, norms against the use of nuclear weapons – including by the United States during the Cold War and Russia more recently – have been weak. As we show in our case studies (see below), when push came to shove, the United States was willing to use limited nuclear weapons and based warfighting plans for the defense of Europe on their use.[28]

[28] See, for example, NSAM 109, which was distributed during the Kennedy administration:

> "If, despite Allied use of substantial non-nuclear forces, the Soviets continue to encroach upon our vital interests [in Europe], then the allies should use nuclear weapons… The allies only partially control the timing and scale of nuclear weapon use. Such use might be initiated by the Soviets, at any time after the opening of small-scale hostilities. Allied initiation of limited nuclear action may elicit an reply in kind; it may also prompt unrestrained pre-emptive attack" (see, https://www.sscnet.ucla.edu/polisci/faculty/trachtenberg/decl/20oct61(a)(NSA).pdf and

Similarly, Sciutto (2024) has pointed out that the Russians were also prepared to use limited nuclear weapons had the war in Ukraine gone less well in the fall of 2022 and spring of 2023.[29] While states seem to believe that nuclear weapons are distinct from conventional weapons, the threshold to use nuclear weapons is likely much lower than expected (Davis Gibbons and Lieber, 2019). Given the empirical evidence from the post-WW2 era, norms are simply too flimsy to place much confidence in as an explanation for why we have not seen the use of nuclear weapons.

We are thus left with the uncomfortable conclusion that formal theoretic investigations of nuclear weapons and deterrence have not added much to our knowledge. Nor do these models shed light on our main question: is the best way to create deterrence (and limit escalation in crises) the threat of nuclear weapon use or instead American conventional superiority? And for more specific policy questions involving escalation – what should the United States response be to tactical nuclear weapon use by Russia in Ukraine? – the answers provided by formal theorists are dangerous.

Takeaway lessons

There are three main lessons from theoretical investigations of nuclear deterrence. The first is that theorists are asking the right question – how do we deter global war and nuclear weapon use *ex ante*? Strategies which depend on retaliation are much less credible and even if they are employed, it simply means that deterrence has failed and we run the risk that the world ends in nuclear Armageddon.

The second is that any deterrence strategy consisting of threats that leave something to chance presents us with an extremely unappealing outcome, especially if it is used repeatedly. Halprin, writing in the *Washington Post*, attacked the logic of Schelling and other formal theorists on this issue in stark terms:

> "What NATO now relies on to deter a Soviet conventional invasion of Western Europe is a doomsday machine linked to a roulette wheel. In other words, the

https://www.history.navy.mil/research/library/online-reading-room/title-list-alphabetically/m/more-bang-buck.html).

[29] Leaked Russian documents from earlier this year also illustrate this point. See, https://www.ft.com/content/f18e6e1f-5c3d-4554-aee5-50a730b306b7.

> alliance says to the Soviets: There are so many nuclear weapons lying about in Europe that if you start a war one of them may go off. And, if one weapon explodes, many more are likely to be used. The result would be the destruction of the world" (Halprin, June 28, 1987).[30]

While it may well have been the case that employing tactical nuclear weapons down to the platoon level to deter a Soviet invasion of Europe through the Fulda Gap had the desired effect, the risks of such a strategy are intolerable and other approaches to deterrence should be sought.[31]

The third is that the rational actor assumption, while convenient for mathematical models, is not descriptive of real-world actors. Pairing this assumption with the claim that nuclear war is only somewhat less bad than conventional war is more troubling still. Andrew Marshall, the head of the Office of Net Assessment, made the case for a deeper and more contextual investigation of the motives of state actors:

> "To get inside the mind of another leadership you really have to start from scratch. It is not just a question of how their history and culture make them different from us, but the ways in which the structure of their regime affects the way they view the outside world and respond to developments … Valuable insights can be gained by observing what the other side does, but the real gains come from asking why they have done it. It may not seem reasonable to you, but it probably does to him. Rationality, after all, can come in many different forms" (Andrew Marshall, quoted in Barrass, 2009).

It is worth pointing out that rationality assumptions in this literature were not delivered on stone tablets from some deity. The idea, for example, that state leaders assess lotteries that may end in nuclear Armageddon in the fashion suggested by theorists is a modeling choice. There is no mathematical basis for the idea that nuclear war is simply some fixed constant worse than conventional; in fact, the alternative choice (i.e., that a nuclear conflict has some shockingly huge negative value) is if anything more plausibly "rational". At root, crucial assumptions in this literature are difficult to support; arguing

[30] See, https://www.washingtonpost.com/archive/opinions/1987/06/28/natos-impotent-deterrent/63319c27-36db-4a56-9f07-e05f26c0cb88/.

[31] On the deployment of the "Davy Crockett" tactical nuclear system in the 1960s, see https://armyhistory.org/the-m28m29-davy-crockett-nuclear-weapon-system/.

that these models represent "rational actors" has deflected scrutiny and hindered more refined work on the behavior of state leaders. We adopt Marshall's approach and argue below that one advantage of conventional weapons for deterrence is that they are slow in comparison to nuclear weapons.[32] In cases that involve conflict between nuclear powers, slow is good because this provides an opportunity for actors to learn about each other during a crisis.

We now turn to an investigation of the empirical literature on deterrence. Unlike the theoretical literature, which attempts to produce results on deterrence via deduction, this literature has instead explored evidence from real-world cases where (nuclear) states were engaged in a crisis.

Problems with Empirical Models of Deterrence

In addition to the theoretical tradition in international relations, an empirical literature examines nuclear deterrence by looking at historical cases where interstate crises occurred. Like the theoretical literature most (but not all) researchers doing empirical work have supported the idea that nuclear weapons have deterrent value. In addition, researchers have hypothesized more generally that states that possess nuclear weapons have advantages in interstate bargaining and thus should achieve better outcomes when they are involved in crises.

Empirical researchers assume that these historical cases are independent and identically distributed (IID). Accordingly, they use these data to conduct large-*n* statistical analyses to discern the systematic advantages nuclear states possess when they become involved in crises. To the extent we can trust empirical results showing that nuclear weapons provide advantages in interstate bargaining, that would lead us to conclude that while the existing theoretical models have flaws, alternatives theories might exist that support reliance on nuclear weapons for deterrence. If, however, there are also flaws in the empirical work, that would serve as additional evidence that the entire approach in the quantitative literature to nuclear deterrence is wrong.

[32] See Kaplan (2015) for an overview of the "compressed timeline" of nuclear weapons use.

There are two main datasets used to study nuclear deterrence.[33] The first, and most prominent, is the International Crisis Behavior (ICB) dataset (Brecher, Wilkenfeld, Beardsley, James, and Quinn, 2023) which details 496 crises from 1918 – 2019. These crises are focused on interstate "military-security" events; i.e., threats that might lead to military hostilities (Brecher and Wilkenfeld, 2022). Most empirical research in this area depends on ICB and it represents the "gold standard" of data work in international relations. The dataset has been through 15 versions and both narratives and sources are well-documented and available online; this level of transparency helps enormously with replication.

A second Militarized Compellent Threats (MCT) dataset was created by Sechser and Fuhrmann for their work on nuclear bargaining (2013, 2017). The goal of this dataset is to "distinguish between deterrent and compellent threats" and includes 210 interstate crises from 1918 to 2001. These data are, unfortunately, difficult to work with. Complete narratives and sources were due to be released in 2013, but the authors have yet to do so. Worse, cases are poorly (i.e., not uniquely) identified in the dataset, making it difficult to do independent validations of the international crises.

Here, we will evaluate this literature to determine whether we can have confidence in the findings.[34] In particular, we will focus on the match between the historical record

[33] There is also a separate empirical literature focusing on militarized interstate disputes (MIDs) and using data from https://correlatesofwar.org/data-sets/mids/. While MIDs have also been used to examine the relationship between nuclear weapons and conflict, issues also exist with these data. See Bell and Miller (2015), Reiter, Stam and Horowitz (2016a, 2016b), and Reiter (2020) for an excellent overview of the main problems.

Note also that MIDs has quite different properties than the datasets considered here (see Maoz, 1988). Despite the apparent domain restriction to *militarized* disputes, the post-WW2 MIDs data has 1726 observations (ending in 2014). In comparison, ICB has 387 (ending in 2019) and MCT has 104 (ending in 2001). Obviously, given the definition of "crisis" should be a larger domain than "militarized dispute", this is a matter of some concern. In large part, it seems that MIDs codes much less salient cases as militarized disputes. E.g., MIDs dispute number two involves conflict between Canada and the United States over the Alaska border. This does not rise to the level of a genuine dispute and MIDs coding as a militarized conflict between the United States and the United Kingdom (given that at that point in time, Britain still had oversight over Canadian foreign policy), is something of a reach. Worse, the Canadian government perceived that Britain had betrayed them in this dispute, which was one of the features that led the Canadian government to seek full independence in the 1930s. In interest of parsimony, we will focus in the main text on the ICB and MCT data.

[34] Asal and Beardsley (2007) use the ICB dataset to examine the relationship between the number of nuclear states involved in a crisis and the level of violence. Their 2009 article extends these tests in the same vein by exploring if opponents of nuclear states end crises sooner and if nuclear states prevail and face shorter crises when salience is high. Kroenig (2013) uses a subset of twenty crises from ICB to examine if the state with "nuclear superiority" is more likely to win in a crisis. Sechser and Furhmann (2013, 2017) use MCT to

and the codings from these datasets as well as the choices made in the statistical models used by the more prominent researchers in the field.

Historiography is hard

A running joke in seminars on quantitative approaches to international relations is that our results would be cleaner if only there was more conflict. International crises, especially those that involve nuclear states, are relatively uncommon, and this complicates applied statistical approaches to the study of deterrence. That said, this scarcity provides one benefit, which is that it allows us to examine individual cases and determine whether codings for variables of interest are supported by the historical record. This is a necessary validation step, because results from statistical models depend on the quality of the data they are based on.

Fortunately, there are several examples of exactly this type of work conducted by scholars well-trained in historiography. Trachtenberg (1991) and Gavin (2012, 2020) examine the Berlin crisis, Plokhy (2021) the Cuban missile crisis, and Miles (2020) and Fraise and Egeland (2023) the Able Archer crisis. All of this work is admirable and conveys the main lesson: we do not have monolithic explanations of even these high-profile historical events, archival work continues and can alter our prior understandings, and many of the variables that quantitative scholars care most about (e.g., who is compelling and who is deterring?) are especially difficult to pin down.

Gavin (2020) raises the important question for our work here, which is how reliable are the codings in the datasets used by quantitative international relations scholars? His main conclusion is that for complex historical cases, "any effort" was going to be inadequate (a point we will return to in the next section). We recognize that this sort of complexity exists in many of the crises, but we ask a simpler question: for highly visible crises and for variables in which consensus exists and are more easily measurable, are the choices made in the ICB and MCT correct?

One salient example concerns the 1956 Suez Crisis in which France, Israel, and the United Kingdom attacked Egypt with the goal of overthrowing Nasser's government. In the ICB data, the narrative is that the crisis was "triggered by an Israeli invasion of the

examine whether or not nuclear brinkmanship succeeds and if nuclear states are more successful in "coercive diplomacy".

Sinai" on October 29, 1956. The Soviets threatened France, Israel, the United Kingdom, and the United States (possibly, with strategic nuclear weapons), and the United States was concerned about threats to "its allies, its deterrent credibility, and its international influence."[35]

But is this what really happened? The short answer is "no", not least because the United States from the start was against the military use of force by its allies and was still seeking to cultivate ties to the Arab states during the 1950s. As Warner (1991) relates from archival work on the United States, Eisenhower's position, revealed to the NSC during the crisis, was "epitomized" by the statement "How could we possibly support Britain and France if in doing so we lose the whole Arab world?" Accordingly, the United States threatened the British economy and demanded that its allies remove their forces from the region. On October 30, the United States also sponsored United Nations Security Council resolutions to that effect (which were vetoed by the British and French but *supported by the Soviets*). In response to these actions, the German Chancellor Konrad Adenauer attacked the United States for "chumminess with the Russians."[36]

The actual situation, then, is quite a bit more complex than the ICB coding of this case, which frames this as a conflict between the United States and its allies on one side and the USSR and Egypt on the other.[37] While it is true that the USSR threatened United States allies, this case does not reduce to a simple conflict between superpowers during the Cold War. And this mistake in understanding the (sometimes muddled) nature of the alliances compounds other choices made in describing this case. For example, there is a clear difference between the military outcome of the conflict and the political outcome that followed. ICB, reasonably enough, focuses more on the political outcome and codes Egypt and the Soviets as victors, with France and the United Kingdom as the defeated. But the United States is also coded as a victor and Israel's outcome is a "partial achievement of basic goal(s)". We thus have very strange dyadic results, with both the US and USSR winning (despite being placed by ICB in opposed dyads).

[35] See, http://www.icb.umd.edu/dataviewer/?crisno=152. For the official history, see https://history.state.gov/milestones/1953-1960/suez.

[36] For an overview of the difficulties presented by this case, see Dietl (2008).

[37] In many of these analyses, crises are split into opposed dyads – i.e., two states that are adversaries in the crisis. This means that for any given conflict, there may be multiple observations in the data if the number of involved actors is greater than two. Accordingly, in this case, the (opposed) dyads included are US-USSR, UK-USSR, UK-Egypt, France-USSR, France-Egypt, USSR-Israel, Egypt-Israel. Curiously, both the US and the USSR, for their dyad, are coded as "winning" the conflict.

In the MCT data, given that the focus is on trying to determine which state initiates the crisis with a "militarized compellent threat", the coding differs from ICB in important ways. Here, we have observations for both France and the United Kingdom threatening Egypt, and separately, the Soviets threatening both France and the United Kingdom. Oddly, though, the United States is not an actor at all in the conflict, nor is Israel.

It is important to stress that we are not cherry-picking cases -- similar mistakes exist in many other highly salient cases that have extensive historical records. For the Yom Kippur War of 1973, the ICB coding of the outcome of the conflict is difficult to justify (no states are coded as losing), a closely related Israeli mobilization in April is treated as a separate crisis (but not similar actions by the Egyptians[38]), and many of the participants in combat during the war are excluded from the data. In the MCT data, there are no observations at all for this crisis, despite the fact that this is one of the qualitative case studies in Sechser and Fuhrmann's 2017 book. For the Soviet invasion of Afghanistan, ICB codes the Soviets as "winning" the crisis, Afghanistan as "losing", and both the United States and Pakistan as "stalemated". The MCT again has no identifiable record of the conflict.

And so on.[39] Given that there are so few cases in the population, mistakes are consequential. As noted above, Kroenig (2013) bases his model on just twenty nuclear crises between 1945 and 2001 and *four* of these – Angola, Afghanistan, Able Archer, and Nicaragua – account for the main results of his statistical model. And, if one corrects the coding error for just one case – Afghanistan – the results lose statistical significance. A strong argument can be made that if the choice is between statistical analyses based on a small sample of miscoded data or a purely qualitative examination of the historical record for key cases, the latter probably conveys more useful information.

Last, as noted by Trachtenberg, one should define the population of cases that are relevant based on the particular research question that is investigated.[40] In part, datasets

[38] The Egyptian army in crucial period of 1972 to 1973 mobilized twenty-two times (see, https://www.hudson.org/foreign-policy/hidden-calculation-behind-yom-kippur-war-michael-doran).
[39] There are, of course, further oddities but at this point we risk boring the reader. One additional example, though, is particularly worrisome. In the entirety of the MCT data, the United States only issued a compellent threat against the Soviets twice (and only issued a total of 17 threats in the post-WW2 period in total); the Soviets never issued a threat against the United States (and only issued a total of 9 threats). Perhaps worse: of the threats issued by the United States, ten of the seventeen occur after the collapse of the Soviet Union began in 1988. Apparently, there was no Cold War.
[40] Personal communication, 2024.

like ICB, MCT, and MIDs suffer from the fact that inclusion criteria are loose and very different types of crises are included.[41] The huge variety of crisis types also makes it more likely that coding mistakes will occur given the inevitable difficulty in applying variable definitions across dissimilar cases. Statistical models, however, depend on the assumption that cases are IID; this implies that cases are of roughly equal salience and generated by the same underlying process. Lumping together border disputes between allies, fishing disputes, mobilizations based on (faulty) intelligence, and crises that either carry a risk of full-scale war or actually progress to full-scale war is a poor design choice. It is also suboptimal to disaggregate single conflicts into "separate" cases – e.g., multiple events exist for several of the major wars included in the data.[42] Between these design choices and the substantial errors contained in individual cases, it is difficult to trust the results of the statistical models based on these datasets.

Statistical issues

In addition to issues involving the coding of historical crises, a number of problems exist with the applied statistics employed by empirical researchers. Unlike the errors above – which require better historical work – these issues could in principle be corrected by rerunning existing models with more appropriate assumptions. Four mistakes, in particular, adversely affect the quality of the statistical results in this literature.

The first issue concerns the measurement of the concepts (variables) included in the statistical models. It is important to emphasize that measurement is often a difficult theoretical issue; we are, in essence, committing to a particular conceptual understanding of (possibly quite complex) real-world phenomena. In the preceding section, we have seen that establishing what an alliance is or what constitutes a victory can be challenging in individual cases. But the problems run deeper as illustrated by the issues involved in measuring relative nuclear capabilities of actors involved in a crisis – a central variable in much of this literature. As Gavin (2020) points out,

[41] To take one prominent example, was the 1983 Able Archer incident an interstate crisis given that the United States and NATO were, initially, unaware that the Soviets were concerned?

[42] For example, in the ICB, there are eleven crises in ICB for the United States involvement in the broader Vietnam conflict, three for the Korean War, and two for the Yom Kippur War. In the MCT data, there appears to be three crises for Vietnam and both the Korean War and the Yom Kippur War are missing. There is also evidence of Western bias, in the sense that Vietnam has eighteen total crises but Afghanistan has one (in the ICB). As we will argue below, weighting cases (both by the number of actors involved and by splitting a single case into multiple crises) has unfortunate statistical properties. It also makes it very difficult to compare results across studies given different datasets use different weights for each crisis.

> "Obviously, there is a profound difference between numerical superiority, no matter how large, and a nuclear balance that enables a country to launch a first-strike that allows it to absorb an acceptable level of retaliatory damage" (Gavin, 2020, p. 60).

Nuclear superiority, then, is a complex measurement of the quality of nuclear forces, which includes a large number of distinct factors. Counting the raw number of missiles (as Kroenig does) or warheads (as Sechser and Furhmann do) is not a necessary simplification; rather, it does damage to the underlying reality of the nuclear forces of states involved in a crisis. Kroenig (2013) notes that missile counts (and ratios between actors involved in a crisis) are not an ideal measure, but defends the choice due to the fact that better measures are "not available" for all cases. Similarly, Sechser and Furhmann (2017) relate that counts (and ratios) of warheads are "crude".

Our own reading of the data is that better measures are in fact available from multiple sources – and it is not as if that many states possess nuclear weapons. The type of delivery vehicle (e.g., which states have ballistic missile submarines), ability to destroy hardened targets, command and control considerations, ability to solve the missile fratricide problem, etc. all exist in multiple studies.[43] As Lieber and Press (2017) point out, relatively recent improvements in missile accuracy and remote sensing give the United States advantages that allow for the possibility of successful counterforce strikes that were not achievable just two decades ago – and this despite the fact that Russia still maintains a lead in a naïve count of warheads. This is not to say that it is easy to establish the overall effectiveness of a state's nuclear arsenal – quite the contrary! But it is better to rely on multiple measures and devote resources to this problem than to rely on a simple and misleading measure. The quality of our inferences depends on the quality of our measures.

The second issue concerns the case universe used in many of these papers. As noted above, even with (over) expansive definitions of what constitutes a crisis, we do not have all that many cases. What should be apparent is that to produce applied statistical models, we need as large a sample size as possible. Accordingly, many of the researchers in this field include cases from ~1918 to the early 2000s in their data, arguing that "*it is appropriate to include cases occurring before 1945, since the prenuclear era provides*

[43] For an overview, see https://apps.dtic.mil/sti/tr/pdf/ADA111287.pdf as one example.

valuable information about the outcomes of compellent threats made by nonnuclear states" (Sechser and Furhmann, 2013).[44]

This is false. If one is considering the effects of nuclear weapons on deterrence and the outcome of international crises, it is best to think of the possession of nuclear weapons as a treatment variable. Some states have nuclear weapons, other states do not, and we would like – as much as possible – to control for all other factors that distinguish these states apart from their nuclear arsenals. Clearly, this is difficult to accomplish with observational data; we can conduct no experiments here. But just as clearly, states before World War II are unlikely to be "essentially the same" as states in the post-war period; we also have abundant evidence that different historical regimes are distinct when it comes to the data generating process of interstate crises and war (Jenke and Gelpi, 2017). Accordingly, one must limit the sample to post-WW2 cases.

The third issue follows from these concerns about the population of cases: states are likely to both select into possessing nuclear weapons and into international crises. Acquiring nuclear weapons is a difficult and expensive process that carries with it potentially severe diplomatic, economic, and military consequences. When states make the decision to develop these weapons, it is often because of the risk of being targeted in future crises or the desire to initiate crises. Further, the decision to initiate a crisis is also a strategic action based on the expected outcome, which is influenced by the conventional and nuclear balance of the involved states. Examining the effect of nuclear arsenals on conflict initiation (or other, related variables) is thus very difficult given that we do not have as-if random assignment of nuclear weapons throughout the international system and crises are not independent of earlier conflicts. While there have been well-purposed attempts to address these questions, existing research does not address the larger issues involved in modeling selection nor confounders that account for possession of nuclear weapons in the first place (Brandt and Schneider, 2007).[45]

The last issue concerns how particular cases are weighted in these datasets. Given that the main concern in much of this literature is to examine the advantage (or not) states accrue from possessing nuclear weapons, datasets are constructed to be dyadic. I.e., for the Suez War of 1956 in the ICB data, we have a row for each dyad of interstate actors that participated in the conflict as adversaries, for a total of eight observations for the

[44] Notably, Asal and Beardsley (2007, 2009) truncate the data to 1945 – 2002, which is the correct approach.

[45] For example, Gartzke and Jo (2009) attempt to jointly model the decision to proliferate and initiate crises; Kroenig (2013) tries to account for whether states "initiate crises they expect to win".

case.[46] Other cases can be much larger or much smaller; e.g., the first crisis for the Korean War has 61 observations, but Able Archer has only a single observation. While this may seem innocuous, the effect this has on our statistical models is that different historical cases have different weights, determined not by salience but by how many actors are involved in the conflict.[47] The results of these models are thus suspect given that different conflicts are weighted very differently.

Takeaway lessons

In the theoretical literature, the trajectory of the field was established early on by researchers and policy makers working out the implications of nuclear deterrence throughout the Cold War period. For the empirical literature, quantitative scholars were operating with the benefit of hindsight. The advent of quantitative methodology as a subfield within political science was in 1983 (Franzese, 2015) and most of the large-*n* statistical work in international relations has taken place in the decades that followed. For these researchers, it was an established fact that nuclear weapons were used for deterrence and that post-WW2, no great powers war had occurred.[48]

Accordingly, much of the effort in the empirical literature has been to determine how much of an advantage the possession of nuclear weapons confers to a state. Examining deterrence directly, given the fact that nuclear weapons were only used once, is impossible with applied statistical models. But one can examine the outcomes of interstate crises, looking to see if nuclear states possess an advantage in bargaining. Given the priors of the field, it is unsurprising that much of the empirical literature has supported a positive role for nuclear weapons. Asal and Beardsley (2007, 2009), for example, echoing earlier work (Huth, Gelpi, and Bennett, 1993), find that the possession of nuclear weapons "has an important and pacific impact… that lends support for an optimistic view of the stabilizing effect of nuclear weapons". Kroenig (2013) finds that a nuclear advantage allows "states to push harder" and win crises.

[46] This means, minimally, we include a row for: USSR-France, USSR-Israel, USSR-UK, Egypt-France, Egypt-Israel, Egypt-UK. And as noted above, the United States should also be included, though it is not as obvious from the historical record how.

[47] Concretely, this means that for the post-WW2 period, 21% of all observations concern the Korean War and the first Gulf War.

[48] This, of course, depends on how one defines "great powers" war. By some measures, great powers war has been on the decline since the Napoleonic wars of the early 19th century (in frequency, though obviously not severity). See Levy and Thompson (2011) for an overview.

More recent studies, however, have presented a more muddled view of the benefits (if any) of nuclear weapons in crisis bargaining. This is in large part due to the statistical issues noted above and the difficulty of finding causal mechanisms in this substantive area (Gartzke and Kroenig, 2016). Gartzke and Jo (2009), for example, provide evidence that "nuclear weapons matter less for war and peace than is generally presumed". Sechser and Furhman (2013, 2017) find that "compellent threats from nuclear states have not been more successful than threats from nonnuclear states". And, Bell and Miller (2015) find that nuclear states "are not significantly less likely to fight wars" but are, in fact, "more prone to low-level conflict".

We are thus left without firm conclusions about the deterrent value of nuclear weapons or the effects these weapons have for states engaged in interstate bargaining. Worse, given the issues we have identified above, the results in the empirical international relations literature on nuclear weapons are not reliable. We believe it is fair to say that most empirical researchers likely believe that nuclear weapons have had some deterrent value, as well as utility in interstate bargaining. But reliable results are not present in this literature to support that view.

In our view, given both the mistakes in historiography and in statistical modeling, a number of actions need to be taken to increase confidence in future empirical research on deterrence (Note: for details on how fragile existing results are, see our **Technical Appendix**).

The first is to recode the data, both to correct cases that do not match the historical record and to create better measures of empirical phenomena (e.g., the balance of nuclear forces for states involved in a crisis). Until we have a better match between the historical record and the datasets we use for statistical modeling, we cannot proceed. We should, though, accept the hard reality that in many of these cases, it is difficult or impossible to create a single, veridical account of "what happened". To deal with this, we should code complex cases in multiple ways, reflecting differences that exist in the historical literature. We can then run ensembles of statistical models, sampling from different codings to generate "average" results (de Marchi, et al., 2024). If, for example, a finding only holds up for a particular interpretation of a case, then we should place less confidence in that result.

The second, as noted above, is to limit the case universe to the post-WW2 period. We should also define a case universe that is more consistent and plausibly IID. Very different types of crises might add to our sample size, but this comes at a considerable cost for the quality of our findings. Fishing disputes between NATO members, to take one example, do not inform us about nuclear deterrence.

The last is to avoid weighting the importance of cases by the number of actors involved. There are several possible strategies one could employ to address this issue. We favor creating a Monte Carlo, where multiple samples are drawn where each crisis is limited to contributing one observation (Cutler, et al., 2016).

Case Studies

We thus are left in something of a pickle. We are unable to trust the main results from either the theoretical or the empirical literature on nuclear deterrence. And we are no closer to understanding, based on existing research, what the role of conventional weapons is in deterrence or the answer to more concrete questions raised by specific crises. For example, if Russia uses a (tactical) nuclear weapon in the Ukraine conflict, what should the response be by the United States and NATO?

To generate intuition about these questions and to provide some foundation for our findings, we cover four cases. Unlike large-*n* empirical studies, we are not attempting to discover regularities that might be present in large datasets. Instead, we use each of these cases to hone in on an important aspect of the real world that should be accounted for by deterrence theories. In other words, these are not "tests" of our findings, but rather features of the real-world that should bear on any plausible deterrence strategy. For example, showing that cases exist where a state leader values ideological commitments highly enough to ignore the costs of nuclear war is enough to demonstrate that deterrence strategies should account for this possibility.

Case Study: Cuban Missile Crisis and the Rational Actor Assumption

Given the theoretical literature on nuclear deterrence and the assumption that actors prefer to avoid nuclear war, it is worth asking if counter-examples exist in real-world cases. Castro's beliefs during the Cuban Missile Crisis serve as an alarming study on whether or not we can trust the rational actor assumption in all contexts. With the attention this case has received in the literature, it is surprising that one critical aspect of it has not been emphasized. Castro did not possess "rational" preferences over

outcomes. In fact, in the middle of the crisis he sent his infamous "Armageddon" telegram to Khrushchev, just after the United States imposed a blockade around Cuba:

> "If… the imperialists invade Cuba with the goal of occupying it, the danger that that aggressive policy poses for humanity is so great that following that event the Soviet Union must never allow the circumstances in which the imperialists could launch the first nuclear strike against it. I tell you this because I believe the imperialists' aggressiveness is extremely dangerous and if they actually carry out the brutal act of invading Cuba in violation of international law and morality, that would be the moment to eliminate such danger forever through an act of clear legitimate defense, however harsh and terrible the solution would be, for there is no other" (Castro to Khrushchev, Oct. 26, 1962, https://nsarchive2.gwu.edu/nsa/cuba_mis_cri/docs.htm).

Castro knew Cuba would not survive a nuclear war between the United States and the Soviets, but was so enraged when Khrushchev "blinked" and deescalated the crisis that he "smashed a fist into a mirror, wounding himself" (Plokhy, 2021). Che Guevara, in an interview after the crisis, also revealed that "if the missiles had been under the Cubans' control, rather than the Russians', they would have been fired".[49]

Fortunately, Khrushchev, perhaps because he had lived through the devastation of two world wars, had a different behavioral model than Castro:[50]

> "In a letter, Fidel Castro proposed that we ourselves should be the first to start an atomic war. Do you know what that would mean? That probably cannot even be expressed at all. We were completely aghast. Castro clearly has no idea about what thermonuclear war is. After all, if a war started, it would primarily be Cuba that would vanish from the face of the Earth. At the same time, it is clear that with a first strike one cannot today knock the opponent out of the fight. There can always be a counter-strike, which can be devastating. There are, after all, missiles in the earth, which intelligence does not know about; there are missiles on submarines, which cannot be knocked out of the fight right away, and so on. What would we gain if we ourselves started a war? After all, millions of people

[49] Quotes are from https://www.newyorker.com/news/news-desk/castros-defining-crisis; also see https://www.archives.gov/publications/prologue/2012/fall/cuban-missiles.html.

[50] For Khrushchev's mindset on nuclear war, see Chapter 6 in Desch (2019).

would die, in our country too. Can we even contemplate a thing like that? Could we allow ourselves to threaten the world of socialism which was hard won by the working class? Only a person who has no idea what nuclear war means, or who has been so blinded, for instance, like Castro, by revolutionary passion, can talk like that" (Khrushchev, Oct. 30, 1962, https://digitalarchive.wilsoncenter.org/document/minutes-conversation-between-delegations-cpcz-and-cpsu-kremlin-excerpt).

The strong assumption in almost all formal models is that actors involved in a crisis prefer to avoid nuclear war and rationally calculate the risks associated with nuclear threats. This does not hold in the Cuban case: Castro was either unaware of or unconcerned with the costs involved in a strategic exchange between the Soviet Union and United States.

Recent models by Debs and Larsen (see above) have taken steps to model irrational actors, but their proposed mechanisms do not describe Castro accurately. In Debs, the core mechanism is that a country with a taste for war can bargain more aggressively prior to conflict. In Larsen, a country that does not distinguish between conventional regime change and war may preemptively launch a limited nuclear strike in the hope that the stronger country backs down. Castro, in contrast, wanted to "eliminate such danger [i.e., the United States] forever" via a full-scale strategic strike by the Soviet Union. In this case, Castro was not using nuclear weapons for coercion. He actually wanted a full-scale strategic exchange due to his ideological commitments.

Accordingly, deterrence policy should not depend on the rational actor assumption without considerable evidence on the behavior of the actual actors involved in a conflict. Instead, we need to understand and account for the "anti-imperialist" ideology of Castro, the fear of a resurgent Germany held by the Soviet leadership, etc.

Case Study: US Defense of Europe and Norms

In the aftermath of the Second World War, the United States almost immediately began to draw down its conventional military forces despite the emerging threat of the Soviet Union. The North Atlantic Treaty Organization was founded in 1949 as a bulwark against the Soviet state, but a shattered Western Europe had little to offer against what was perceived to be a fearsome Soviet juggernaut. The answer, in both Europe and the Pacific, was the first use of nuclear weapons by the United States in response to a Soviet

(conventional or nuclear) attack. Accordingly, the United States deployed nuclear weapons to the United Kingdom in 1954; the arsenal peaked in the 1970s with some seven thousand nuclear warheads in Europe stationed in Germany, Italy, France, Belgium, Turkey, the Netherlands, and Greece. There remain perhaps 100 B61 gravity bombs in Europe today.[51]

The nuclear-first strategy was no secret and was in fact quite openly discussed. American Secretary of State John Foster Dulles told NATO's governing body, the North Atlantic Council, in 1954 that "if an all-out Soviet attack occurred, whether atomic or otherwise, the NATO response would be a defense employing atomic weapons."[52] Winston Churchill was an unabashed fan of this strategy, telling his Parliament in March of 1955 that "but for American nuclear superiority, Europe would already have been reduced to satellite status and the Iron Curtain would have reached the Atlantic and the Channel."[53]

While the initial strategy relied upon the use of large nuclear (and, after 1952, even more powerful hydrogen) bombs delivered first by airplanes and later by ballistic missiles, as years went by America developed an extraordinary range of nuclear weapons and relied on them all the way down to the tactical level of war. The most infamous is probably the M28/M29 Davy Crockett, deployed in Europe from 1961 to 1971; with an initial range of about one- and-a-quarter miles and a blast radius of some 500 feet, it was mounted on a Jeep or M113 Armored Personnel Carrier and fired by a crew of 3 or 5. Some 2100 were produced; the weapon was discontinued, according to Brigadier General Alvin E. Cowan, because of "great fear that some sergeant would start a nuclear war."[54]

Of greater tactical importance than the Davy Crockett launchers were atomic artillery shells, several thousand of which were put into service in the 1950s and 1960s and only withdrawn in 1991 after the collapse of the Soviet Union. Designed to be fired by

---

[51] Jonathan Masters and Will Merrow, "Nuclear Weapons in Europe: Mapping US and Russian Deployments" Council on Foreign Relations, March 30, 2023, available at https://www.cfr.org/in-brief/nuclear-weapons-europe-mapping-us-and-russian-deployments

[52] Statement to the North Atlantic Council, April 24, 1954, cited by Eric Edelman in Brands (2023, p. 677). This remains true today.

[53] Remarks to Parliament March 1, 1955, cited in *ibid*., 671.

[54] *Proceedings of the Tactical Nuclear Weapons Symposium*, September 3-5, 1969, p. 173. This report also reveals the existence of a computer program named SATAN, for "Simulation for the Assessment of Tactical Nuclear Weapons" on page 127.

conventional howitzers, the rounds were also slated to be employed by NATO members including West Germany, the Netherlands, Italy, Belgium, and Canada. Their use in the event of a Soviet invasion was a foregone conclusion. Robert Farley has noted that "Conventional warfighting plans by the major NATO powers often amounted, almost literally, to efforts to reach the English Channel just ahead of the tanks of the Red Army. NATO expected to liberally use tactical nuclear weapons to slow the Soviet advance, an action which would inevitably invite Soviet response (the Soviets also prepared for this dynamic)."[55]

The idea that norms against the use of nuclear weapons were widespread flies in the face of the Cold War historical record—not just in global war, but also in the smaller crises that the possession of nuclear weapons by both sides made possible. Eisenhower reportedly threatened the use of nuclear weapons to end the Korean War after his predecessor had refused General Douglas MacArthur's request to use them against a then nuclear-weapons-free People's Republic of China. Dulles proposed (and Eisenhower rejected) plans to employ nuclear weapons during the Quemoy-Matsu crisis of 1958, and their use was also seriously considered during both the French and American wars in Vietnam. In his memoir *In Retrospect*, Secretary of Defense Robert McNamara writes that "The president and I were shocked by the almost cavalier way in which the Chiefs and their associates, on this and other occasions, referred to, and accepted the risk of, possible use of nuclear weapons."[56]

We thus conclude, given warfighting plans from both the United States and Russia, that norms against nuclear weapon use are not as robust as many policy-makers and academics describe. When push comes to shove, it is our expectation that states will employ nuclear weapons if any vital interests are threatened. And, determining which interests are vital to a state should be a highly salient area of research.

Case Study: US Efforts to Limit Proliferation and Conventional Security Guarantees

In 1954, West German Chancellor Adenauer ruled out production of nuclear weapons on German soil, but intentionally elided the possibility that Germany might acquire nuclear weapons produced elsewhere. In fact, in the ensuing decade, Adenauer made multiple

[55] Robert Farley, "NATO Had Only One Cold War Plan to Stop a Russian Invasion: Nuclear War", *The National Interest*, 3 October 2019, available at NATO Had Only 1 Cold War Plan To Stop a Russian Invasion: Nuclear War | The National Interest

[56] Jonathan Guyer, "The Secret History of American Tactical Nukes", *Vox*, October 21, 2022, available at Would Putin use a small nuke? Look to the US's troubled nuclear history. - Vox

arrangements with the intent to acquire such weapons, first with a secret agreement with France and Italy in 1957 where the former Axis powers would help finance France's nuclear program in exchange for a share of the weapons, and then later in a similar agreement with de Gaulle in the early 1960s (Schwarz, 1991).

West Germany's desire for nuclear weapons had two components. First, West Germany had doubts about the credibility of United States commitments to the continent:

> "French analysis of the nuclear-strategic world situation was very close to Adenauer's own [in 1957]. The faster the Americans developed their intercontinental ballistic missiles, however, and the more vulnerable they became to Soviet weapons, the more it was to be feared that they would sooner or later pull out of Europe militarily. Provision had to be made for that day" (Schwarz, 1991, p. 318).

It is notable that they were concerned both with American vulnerability to Soviet weapons, which suggest doubts about the logic of MAD in extended deterrence situations, and that the United States would pull its conventional forces out of Europe. This later point speaks to the important role that United States conventional weapons and the theater balance of power played in Germany's conception of security. If conventional security guarantees of the United States were withdrawn, Germany believed that possession of nuclear weapons would mean that "an attack on Germany or France would mean immediate and total war for the whole of the alliance" (Schwarz, 1991, p 321). Germany would thus achieve deterrence, but at the terrible price of increasing the likelihood of nuclear Armageddon.

Second, there was the concern that if they did not have a say in the alliance's nuclear planning, decisions in wartime would be made at Germany's expense. For example, "In 1955, a NATO exercise called Carte Blanche went through the likely stages of an all-out war between east and west. And in the shortest time, starting with the British commanders in the north, Germany -- West and East -- was hit by several hundred nuclear explosions fired by the west in a desperate attempt to stop Soviet tanks from reaching the Rhine" (Bange, 2009, p. 364). When German political elites learned of this exercise, their desire to have a say in NATO's strategic planning, or have an independent deterrent, gained momentum. United States counterproliferation efforts were needed to address Germany's twin concerns of the United States caring too little about the

continent to risk war over it, and enough about the continent to sacrifice Germany in the event of a general war.

To promote non-proliferation, the United States attempted to channel Germany's desire for proliferation through NATO, with the proposal of the Multilateral Nuclear Force in 1963 featuring joint control of NATO's nuclear deterrent, but this proposal eventually faltered (Bange, 2009). The United States also attempted to wield its influence to convince West Germany (and Italy) to sign on to the Comprehensive Test Ban Treaty and later the Non-Proliferation Treaty (Gerzhoy, 2015). This was done with a combination of carrots (the Multilateral Nuclear Force and an agreement where West Germany would have the final say over the use of nuclear weapons on German territory) and sticks (threats to withdraw its conventional guarantees to the Federal Republic). Ultimately, it was United States threats to withdraw its conventional forces that prompted Germany to accede to United States demands and sign the Non-Proliferation Treaty – "rather than voluntarily forswearing acquisition of a nuclear capability, the FRG was compelled to do so by explicit U.S. threats of military abandonment" (Gerzhoy, 2015, p. 94).

This is by no means the only case that demonstrates the importance of United States conventional guarantees. Three months after the armistice halted the Korean War and froze the border between the warring factions at the 38th parallel, the United States signed a Mutual Defense Treaty with South Korea, with both parties promising "mutual aid" and assistance in the event of an attack. This treaty laid the groundwork for an alliance which saw the deployment of United States conventional forces to the Korean peninsula. But by 1970, a rising conventional threat from North Korea coupled with decreasing United States support in theater prompted South Korea to consider developing its own nuclear program (Jang, 2016). South Korea's desire for nuclear weapons was not driven by shifts at the strategic level, but rather by the possibility the United States would withdraw its conventional forces and alter the theater balance of power. It was not until United States halted troop withdrawals and the Reagan administration reaffirmed commitment to the protection of the peninsula that South Korea halted its nuclear weapons program.

As a last example, Israel is an illustrative case where United States non-proliferation policy failed. In the 1950s, Israel began considering nuclear weapons development due to the existential threat posed by its Arab neighbors and its belief that it could not rely on other states for defense. The United States was unable to compel Israel to forgo this development in large part because it was unable to credibly commit to defending Israel

in the event of an attack. Ultimately, acknowledgement of this limitation is what pushed the United States toward acceptance of the Israeli policy of nuclear ambiguity. Although Israel planned to develop nuclear weapons as early as 1955, the Suez Crisis played a central role in Israeli going forward with a nuclear program (with the assistance of France, which felt similarly wary about the security guarantees of the United States). It was during this crisis that Israeli learned that "when it came to security, Israel was on its own" (Joshi, 2000).

In each of these cases, United States conventional superiority and a willingness and ability to wield that superiority to back security guarantees was an important component of the United States policy on non-proliferation. In contrast, the nuclear arsenal of the United States and a reliance on MAD were not seen as effective guarantees by other states. It is worth noting that states that ultimately pursued nuclear weapons (i.e., both Israel and France) did in fact see deterrent value to possessing these weapons. That said, the risks involved were dire compared to conventional deterrents. Nuclear weapons for conventionally weak states leave few options when war begins and the risk of accidental use is always present (and increases as more states attain nuclear arsenals).

Case Study: Ukraine War and US Conventional Retaliatory Capability

If the problem America faced during the Cold War was a Russia that was too strong to be deterred with only conventional weapons, the problem flipped in the 1990s. The dissolution of the Soviet Union gutted the Russian military, leaving the United States with by far the world's most capable military.

That situation did not, of course, mean the end of conflict. A post-Soviet Russia was soon led by a former KGB agent who was convinced that the greatest geopolitical tragedy of the twentieth century was not the Second World War, in which some twenty-five million Soviet citizens died, but the collapse of the Soviet Union forty-five years later. Putin, determined to restore the Soviet empire, invaded Ukraine covertly in 2012 and then openly in 2022.

Ukraine had surrendered its nuclear weapons, then the third largest stockpile in the world, in 1994 according to the terms of the Budapest Memorandum, which promised that the United States and Russia would defend Ukraine against any possible aggressor. Ukraine's weapons were demobilized or returned to Russia in return for United States

technical and financial assistance and Russian oil and gas. While Ukraine did not have the ability to maintain or employ nuclear weapons after the dissolution of the Soviet Union, the Budapest Memorandum was no guarantee of sovereignty when Putin turned his eyes on Ukraine. The United States, so openly reliant on nuclear weapons to defend Europe during the Cold War, now found that Russian nuclear weapons limited its ability (or willingness) to defend Ukraine, an aspirant to NATO membership, against open aggression.

While the United States was confident that Russia was preparing to invade Ukraine in early 2022, most observers projected an easy Russian victory, with some suggesting a likely Ukrainian insurgency against Russian occupation.[57] The United States was hesitant to provide Ukraine with significant armaments as a result. Russian military performance did not live up to expectations while Ukraine, perhaps assisted through US and allied intelligence, fought hard against the initial invasion, driving Russian forces back from Kiev.[58]

American aid for Ukraine evolved over time, with the provision of increasingly lethal weapons as the Biden administration learned to discern the intent of a Russia that quickly found itself on its heels. Notably, Russia has resorted to threats about the use of nuclear weapons to stave off additional Western support. These threats reached a crescendo in October of 2022 as Ukraine appeared poised to achieve significant battlefield successes in eastern Ukraine, with President Joseph Biden reportedly warning a group of donors that "we have a direct threat of the use of a nuclear weapon if in fact things continue down the path they've been going."[59]

Putin was attempting to overcome what appeared to be inferiority in conventional capabilities with the threatened use of tactical nuclear weapons to change the situation on the battlefield, just as the United States had throughout the Cold War. The threats were not only judged credible by the United States, but also by the person who had the

---

[57] John Nagl, "Will Ukraine be Afghanistan All Over Again for Russia?" *Foreign Policy*, February 22, 2022, available at A Ukrainian Insurgency After Russian Invasion Could Spell Disaster for Putin (foreignpolicy.com)

[58] Paul Sonne, Isabelle Khurshudyan, Serhiy Morgunov, and Kostiantyn Khudov, "Battle for Kyiv: Ukrainian Valor, Russian Blunders Combined to Save the Capital", *Washington Post*, August 24, 2022, available at Battle for Kyiv: How Ukrainian forces defended and saved their capital - Washington Post.

[59] David Sanger, "Biden's Armageddon Moment: When Nuclear Detonation Seemed Possible in Ukraine", *The New York Times*, March 9, 2024, available at Biden's Armageddon Moment: When Nuclear Detonation Seemed Possible in Ukraine - The New York Times (nytimes.com)

best reason in the world to doubt their veracity: Ukrainian President Vladimir Zelensky. In March of 2024 when David Ignatius asked Zelensky about President Biden's reluctance to provide Ukraine with more lethal weapons, Zelensky answered: "I think he's cautious about nuclear attack from Russia … He's [Putin] crazy. There is nobody in the world who can tell you 100 percent what he will do. That's why Biden is cautious."[60]

The important point from our perspective is that the United States strategy in Ukraine has not led to an escalation of the crisis. No nuclear weapons have been used despite the aid given to Ukraine and conventional warfare has not "blended" into nuclear. Pivotal, however, in this success is the fact that the United States has a long history with Russia, Putin has been in power for almost twenty-five years, and the current stage of the conventional conflict in Ukraine has unfolded slowly over two years. This has given the actors involved in the crisis time to assess how to best respond.[61] In a nuclear war, there is no time for learning. And, in a crisis with states that are less informed about each other (e.g., a potential conflict between the United States and China), the risk of accidents or escalation would increase.[62]

Findings

As we have noted above, researchers in this area deserve credit for identifying the crucial question that any theory of deterrence must account for: How do we deter *ex ante*? Retaliating *ex post* with (strategic) nuclear weapons is not an effective strategy. Unfortunately, as we have seen, quantitative studies – both empirical and theoretical – have not fulfilled their promise of providing systematic evidence on the utility of nuclear weapons for deterrence.

While possession of a secure second-strike capability is essential, nearly every theoretical result on the role of nuclear weapons is problematic. Mechanisms that remove decision-making from state leaders to create threats that leave something to chance carry the risk of unacceptable outcomes. Similarly, "madman" diplomacy and demonstrating resolve

[60] David Ignatius, "Zelensky: "We are trying to find some way not to retreat" *Washington Post*, 29 March 2024, available at https://www.washingtonpost.com/opinions/2024/03/29/ignatius-zelensky-interview-ukraine-aid-russia/

[61] For evidence on the Biden administration learning Russia's red lines, see https://www.brookings.edu/articles/arming-ukraine-without-crossing-russias-red-lines/.

[62] For the state of communication between the United States and China, see https://apnews.com/article/politics-united-states-government-lloyd-austin-china-9d1b7c9aa40b22d0bda497ba29be8d9b. For the difficulties in such communication, see https://www.rand.org/pubs/commentary/2022/07/another-hotline-with-china-isnt-the-answer.html.

by escalating the risks of nuclear conflict have in practice been both unsuccessful and dangerous. And, contexts in which states are at a conventional disadvantage and rely on nuclear weapons for deterrence increase the likelihood of uncontrolled escalation.

Even as strong an advocate of the use of American military power abroad as Henry Kissinger was concerned about an overreliance on nuclear weapons, albeit early in his career, writing that "at a minimum, the conventional capability of the free world should be of such a size that a nuclear defense becomes the last and not the only recourse. The best situation is one in which the conventional forces of the free world can be overcome only by nuclear weapons."

The authors of this paper, in a phrase many of them thought that they would never write, agree with Kissinger.

A fair reading of the post-WW2 history, indicates that while nuclear weapons provided deterrence, they also in expectation carried unacceptable risks -- global war was avoided not because of cold, consistent strategy but in large measure due to luck. While direct confrontations between the United States and USSR were avoided, this period was hardly peaceful. Multiple conflicts occurred with heavy involvement of the major powers, casualties were extremely high in both Korea and Vietnam, and escalation to nuclear weapon use was often on the table. The idea that nuclear weapons secured the peace in the post-WW2 era ignores both the nuclear risks and the conventional conflicts that were present.

If one broadens to the effects of nuclear weapons on interstate bargaining, the picture remains quite murky. Despite the fact that relatively few cases exist of interstate crises that involve states with nuclear weapons, empirical work often runs afoul of the historical record and has not relied on best practices when it comes to statistical modeling. It is fair to say that no reliable evidence exists on the effects of nuclear arsenals on interstate bargaining.

Our approach in this paper has been to illustrate the flaws that undermine results from both empirical and theoretical work on deterrence. We do not present our own game theoretic model of deterrence here, nor do we engage in large-*n* empirical work to arrive at general findings. Instead, by examining the logic of deterrence and using salient case

studies, we are able to arrive at a set of preliminary findings that should bear on deterrence strategies today. Importantly, deterrence involving nuclear states is an odd context in the sense that strategies cannot work most of the time. Mistakes that lead to nuclear weapons use have catastrophic consequences for all actors involved in a crisis and deterrence strategies must, as a result, be especially concerned with fault tolerance.

Our investigation of the existing literature and the case studies above presents us with *five* principles that any policy on deterrence must account for.

1. Deterrence based on MAD and creating a non-zero probability of escalating (intentionally or non-intentionally) to nuclear war carries unacceptable risks. Encouraging nuclear proliferation also carries unacceptable risks. Conventional superiority in-theater, while more expensive, is a far less risky deterrence strategy that makes the initiation of conflicts much less likely.
2. Nuclear conflicts are likely to be "fast"; conventional conflicts are "slow". To prevent uncontrolled escalation, states should rely on conventional weapons to slow down conflicts long enough to learn about the actors involved. This need is especially acute given that all actors may not be rational or share our beliefs or preferences. A corollary is that instead of assuming that all states (and state leaders) are the same, much more research needs to be conducted to uncover differences that exist across states. Not all states will respond the same way to any given threat and not all states will have the same red lines that would trigger escalation.
3. United States conventional superiority has aided our efforts to curb proliferation and deter conflict. This is due in large part to the conventional security guarantees we have been able to offer.
4. We should expect that if a nuclear state suffers substantial conventional losses or has vital interests threatened, then the use of nuclear weapons will be considered. Norms are not, on their own, as effective as we would like.
5. We should also expect that states that possess a secure second-strike capability will engage in conventional aggression that they would not otherwise attempt. United States conventional superiority provides options in conflicts of this type without creating direct threats to the homeland or the sovereignty of the aggressor's government. This works to limit escalation to nuclear weapon use.

To illustrate the above principles, we consider the difficult case posed by the current Ukraine conflict. Our motivating questions are: Why have the Russians not already used

a (tactical) nuclear weapon in Ukraine? And, what should the United States response be if the Russians use a (tactical) nuclear weapon?

First, it is worth noting that while nuclear weapons have reduced the chances of large-scale war between states that possess a secure second-strike capability, Russia's nuclear stockpile has enabled its initial invasion of Ukraine and minimized the risk of a catastrophic loss for Russia. Possession of nuclear weapons allows aggressors to upset the established balance of power and the established norms of the international system. As a consequence, the United States sparingly doles out military capabilities to Ukraine and limited Ukraine from using these capabilities on strikes against Russian soil.

Second, this conflict has been slow compared to the likely timeline of a nuclear conflict. The current stage has taken place over two years and as noted above, this has provided the Biden administration with time to learn about Russian motives and calibrate aid to the Ukrainians without stepping over Russian red lines.

Third, intelligence has revealed that the use of nuclear weapons is not off the table for the Russians. If Russian forces had collapsed in the fall of 2022, or if the spring offensive of 2023 had been more successful for Ukraine, the use of nuclear weapons was entirely possible.[63] The United States and NATO must plan for this possibility and craft a deterrence strategy that does not end in full-scale nuclear war. Crucially, deterrence strategies should slow escalation and provide options short of Armageddon.

Given these factors, we believe the answer to the first question on why the Russians have not used nuclear weapons is attributable to a successful United States deterrence strategy that has not presented an acute threat to the Russian government or homeland. American *conventional* superiority has allowed for substantial aid to Ukraine without resorting to blunter strategies (e.g., the direct involvement of United States combat troops). This has helped the United States to deter further Russian aggression and punish Russia for attacking Ukraine, while at the same time avoiding more direct threats to Russian sovereignty that might result in escalation to the use of nuclear weapons.

[63] See https://www.csis.org/analysis/russian-nuclear-calibration-war-ukraine for an excellent overview of the timeline and Russia's shifting stances on nuclear weapons use.

The answer to the second question on a United States response to tactical nuclear weapon use by Russia is more nuanced. United States conventional superiority means that if Russia did choose to use tactical nuclear weapons, America could respond with devastating effect without crossing the nuclear threshold—for example, by sinking what remains of the Russian Black Sea fleet from the air with conventional weapons. Most important, this punishment would offset Russian gains from using a nuclear weapon, without threating the Russian government or homeland. If the United States lacked such a clear conventional advantage, any punishment strategy would be much riskier, involving American boots on the ground and thus more direct threats to the Russian homeland.[64] These actions would then prompt the Russians to escalate based on everything we know about their warfighting doctrine.

The Possibility of Using Tactical Nuclear Weapons to Bolster Weak Conventional Forces

At this point, it should be clear that we have very little faith in the use of *strategic* nuclear weapons as a substitute for conventional forces. For us, deterring the initial act of aggression with conventional forces prevents crises from taking the first step toward uncontrollable escalation. Yet we have seen that for both the United States during the Cold War and the Russian state today, the threat of *tactical* nuclear weapons use has been employed as an alternative deterrent. The logic is that tactical weapons can blunt a conventional attack and may not escalate to a catastrophic full-scale strategic exchange.

It is thus reasonable to ask, if a state is conventionally weak in a theater, should it supplement its forces by deploying tactical nuclear weapons? Our answer is "no", for two main reasons. First, the same flawed logic of a threat that leaves something to chance applies with equal force to tactical nuclear weapons. If a state has perfect command and control over these weapons and waits until an attack occurs, then the only option that remains may be to "spoil" the prize after an aggressor has already won on the conventional battlefield. To achieve real deterrence with these weapons, they need to be pushed forward to the edge of the battlefield in the same way the United States did in the defense of Europe. This, of course, substantially increases the likelihood that accidental use will occur. If tactical nuclear weapons are used, even in small numbers, a possibility exists that the conflict will escalate dramatically, in no small part because states would be unable to commit to using these weapons *only* to reestablish conventional parity.

[64] The main issue with using American ground forces in Ukraine is that it would be impossible for the United States to commit to not going further and threatening the Russian government or homeland.

Second, imagine the outcome if the United States were forced to use tactical nuclear weapons in defense of Europe, South Korea, or Taiwan. What then? Analysts have always been concerned about counterforce usage precisely because they weaken the opponent, which constrains the response to further (possibly strategic) nuclear weapon use. We have no doubt that Russia, North Korea, or China, faced with a defensive first use of tactical nuclear weapons by the United States and its allies, would respond in an escalatory fashion. As we have noted above, the relatively fast pace of nuclear weapons compared to conventional forces limits learning during conflict and increases the possibility of a catastrophic outcome.

The United States is currently the only country in the world, due to its conventional superiority, that can uphold both the non-use norm and the non-proliferation norm while at the same time responding to nuclear first use by another state. Crucially, the United States can punish states for nuclear aggression without further escalating toward more nuclear weapon use. In contrast, for nuclear states that are conventionally weak, the only options available are to rely on nuclear weapons or walk away.

Walking away can, of course, be an unpalatable option. From the Russian perspective, if they had faced defeat on the battlefield in fall of 2022, the use of tactical nuclear weapons might be seen as preferable to retreat. Going forward, if the United States and NATO allies become more involved in the conflict such that it tips the outcome toward Ukraine, the use of tactical nuclear weapons might again seem compelling to Russia. In our view, however, this logic falls apart if one considers the most likely outcome. Even discounting the possibility that nuclear use might well encourage escalation, what would the Russians gain? Ukraine would be spoiled. The United States would surely retaliate, punishing Russia to offset any battlefield advantage. At that point, Russia would have its conventional forces in disarray and extreme difficulty defending the homeland. The point here is that if a state is conventionally weak and both sides in a conflict have a secure second-strike capability, the first use of tactical nuclear weapons does not lead to a favorable outcome for the weaker state. It is, in fact, better to walk away and admit that if a theater is not important enough to invest in to achieve conventional superiority, then it is likely not important enough to escalate to tactical nuclear weapons.

Conclusion

Our core argument is that the world has avoided using nuclear weapons more by good luck than by design. While nuclear weapons, in some contexts, have served as a deterrent, they have never done so without imposing unacceptable risks of escalation. Simply put, if we were to repeatedly "rewind" the historical tape from the end of World War II to the present day, in many of these trials nuclear weapons would have been used.

While we have been critical of quantitative approaches to the study of nuclear deterrence, we are supportive of the overall goals of theorists and empiricists working in this area. It is certainly the case that better historiography is needed to inform the work of quantitative scholars, but history does not generalize. In-depth studies of single cases – even important ones like the Cuban Missile crisis – contain too many idiosyncratic features to generate policy advice for the modern context. And, history is difficult to falsify; that requires new cases to test hypotheses against.

Clearly, more study on deterrence is needed and we close with three recommendations. First, the different communities studying deterrence are currently far too siloed. Practitioners, historians, and quantitative scholars must collaborate more closely than in the past to improve the quality of research. Understanding deterrence and in particular the role of conventional and nuclear forces should be one of the most salient questions in the social sciences.

Second, better data should be collected to inform deterrence research. While we will always struggle with the relative lack of cases and the difficulties inherent in doing historiography in this area, we can improve our understanding of international crises compared to the status quo. Our view is that quantitative research has a great deal to offer deterrence studies, but only if the data improve and we strive to account for differences that exist between states. This is not just a matter of correcting mistakes in historical work. Rather, we need more comprehensive data on each of the factors that bear on deterrence. For example, the behavior of state leaders is enormously important and currently, no datasets collect statements and actions by state leaders that bear on deterrence.[65]

[65] There is, of course, the possibility that state leaders engage in cheap talk. But in our view, it is surprising – especially if one examines statements over a longer time period – how often leaders reveal a great deal about their mental models of the world.

Last, we should strive to create testable propositions from our research on deterrence in *each* of the theaters where conflict is likely. We cannot, of course, do this directly – there are no nuclear conflicts to draw on for data. But we can take a page from the empirical international relations literature and create indirect tests of our hypotheses by studying phenomena with more plentiful data (e.g., interstate crises). If the study of deterrence remains a purely backward-looking endeavor, there will continue to be debate, but it will be difficult to choose among competing theories on deterrence. Theater specific research will also allow us to investigate whether general principles hold in deterrence, or instead if each separate context requires more specific investigation. Social scientists tend to seek general explanations of phenomena (and this is valuable), but in the case of deterrence variance between states may overwhelm general principles. Transforming the field requires sharp tests against new phenomena and we should take every opportunity to expose our ideas to crises that occur in the real-world (Gleditsch, 2022).

Citations

Asal, Victor, and Kyle Beardsley. "Proliferation and international crisis behavior." Journal of Peace Research 44.2 (2007): 139-155.

Beardsley, Kyle, and Victor Asal. "Winning with the Bomb." Journal of Conflict Resolution 53.2 (2009): 278-301.

Bange, Oliver. "NATO as a framework for nuclear nonproliferation: The West German case, 1954–2008." International Journal 64.2 (2009): 361-382.

Barrass, Gordon S. "The Great Cold War: a journey through the hall of mirrors." Stanford University Press, 2009.

Bell, Mark S., and Nicholas L. Miller. "Questioning the effect of nuclear weapons on conflict 1." Nonproliferation Policy and Nuclear Posture. Routledge, 2015. 295-312.

Bitzinger, Richard. "Assessing the Conventional Balance in Europe, 1945-1975." RAND Note, 1989.

Brands, Hal, ed. *The New Makers of Modern Strategy*. (Princeton, N.J.: Princeton University Press, 2023).

Brands, Hal, ed. *War in Ukraine: Conflict, Strategy, and the Return of a Fractured World* (Washington, D.C.: Johns Hopkins University Press, 2024).

Brandt, Patrick T., and Christina J. Schneider. "So the reviewer told you to use a selection model? Selection models and the study of international relations." Unpublished manuscript 4.23 (2007).

Brecher, Michael, and Jonathan Wilkenfeld. A study of crisis. University of Michigan Press, 2022.

Brecher, Michael, Jonathan Wilkenfeld, Kyle Beardsley, Patrick James and David Quinn (2023). International Crisis Behavior Data Codebook, Version 15. http://sites.duke.edu/icbdata/data-collections/

Budjeryn, Mariana. "Was Ukraine's nuclear disarmament a blunder?" World Affairs 179.2 (2016): 9-20.

Burr, William, and Jeffrey P. Kimball. Nixon's nuclear specter: The secret alert of 1969, madman diplomacy, and the Vietnam war. University Press of Kansas, 2015.

Cutler, J., De Marchi, S., Gallop, M., Hollenbach, F. M., Laver, M., & Orlowski, M. (2016). Cabinet formation and portfolio distribution in European multiparty systems. British Journal of Political Science, 46(1), 31-43.

Davis Gibbons, Rebecca, and Keir Lieber. "How durable is the nuclear weapons taboo?." Journal of Strategic Studies 42.1 (2019): 29-54.

Debs, Alexandre. "How Could States Use Nuclear Weapons? Four Models After the Bomb." Security Studies 31.3 (2022): 317-331.

de Marchi, Scott, Max Gallop, and John Nagl. "Historical Confidence Intervals and Nuclear Superiority". Working paper, 2024.

de Marchi, Scott, Christopher Gelpi, and Jeffrey D. Grynaviski. "Untangling neural nets." American Political Science Review 98.2 (2004): 371-378.

de Marchi, Scott, and Michael Laver. The governance cycle in parliamentary democracies. Cambridge University Press, 2023.

Desch, Michael. Cult of the irrelevant: The waning influence of social science on national security. Princeton University Press, 2019.

Dietl, Ralph. "Suez 1956: A European Intervention?" Journal of Contemporary History 43.2 (2008): 259-278.

Fraise, Thomas, and Kjølv Egeland. "Able Archer: How close of a call was it?." Bulletin of the Atomic Scientists 79.3 (2023): 155-160.

Franzese, Robert J. "An Introduction to Quantitative Research in Political Science." (2015).

Gartzke, Erik, and Dong-Joon Jo. "Bargaining, nuclear proliferation, and interstate disputes." Journal of Conflict Resolution 53.2 (2009): 209-233.

Gartzke, Erik, and Matthew Kroenig. "Nukes with numbers: Empirical research on the consequences of nuclear weapons for international conflict." Annual review of political science 19.1 (2016): 397-412.

Gavin, Francis J. Nuclear statecraft: history and strategy in America's atomic age. Cornell University Press, 2012.

Gavin, Francis J. "Strategies of inhibition: US grand strategy, the nuclear revolution, and nonproliferation." International Security 40.1 (2015): 9-46.

Gavin, Francis J. Nuclear weapons and American grand strategy. Brookings Institution Press, 2020.

Gerzhoy, Gene. "Alliance coercion and nuclear restraint: How the United States thwarted West Germany's nuclear ambitions." International Security 39.4 (2015): 91-129.

Gleditsch, Kristian Skrede. "One without the other? Prediction and policy in international studies." International Studies Quarterly 66.3 (2022): sqac036.

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction (Vol. 2, pp. 1-758). New York: springer.

Hewitt, J. Joseph. "Dyadic processes and international crises." Journal of Conflict Resolution 47.5 (2003): 669-692.

Huth, Paul, Christopher Gelpi, and D. Scott Bennett. "The escalation of great power militarized disputes: Testing rational deterrence theory and structural realism." American Political Science Review 87.3 (1993): 609-623.

Jang, Se Young. "The Evolution of US Extended Deterrence and South Korea's Nuclear Ambitions." Journal of Strategic Studies 39, no. 4 (2016): 502–20.

Jenke, Libby, and Christopher Gelpi. "Theme and variations: Historical contingencies in the causal model of interstate conflict." Journal of Conflict Resolution 61.10 (2017): 2262-2284.

Jervis, Robert. The illogic of American nuclear strategy. Cornell University Press, 1984.

Jervis, Robert. The meaning of the nuclear revolution: Statecraft and the prospect of Armageddon. Cornell University Press, 1989.

Joshi, Sharad. "Israel's nuclear policy: A cost-benefit analysis." Strategic Analysis 23, no. 12 (2000): 2089-2110.

Kaplan, Edward. To kill nations: American strategy in the air-atomic age and the rise of mutually assured destruction. Cornell University Press, 2015.

Krcmaric, Daniel. "Should I stay or should I go? Leaders, exile, and the dilemmas of international justice." American Journal of Political Science 62.2 (2018): 486-498.

Kendall-Taylor, A., Kofman, M., Lokker, N., and Hautala, H. (2023). Assessing the Evolving Russian Nuclear Threat. Center for a New American Security, 24.

Keohane, Robert. 1999. "Ideology and Professionalism in International Institutions: Insights from the Work of Douglass C. North." Alt, J., M., Levi and E. Ostrom [ed] (1999): 228–46.

Kroenig, Matthew. "Nuclear superiority and the balance of resolve: Explaining nuclear crisis outcomes." International Organization 67.1 (2013): 141-171.

Larsen, Even Hellan. "Deliberate Nuclear First Use in an Era of Asymmetry: A Game Theoretical Approach." Journal of Conflict Resolution (2023).

Levy, Jack S., and William R. Thompson. The arc of war: Origins, escalation, and transformation. University of Chicago Press, 2011.

Lieber, Keir A., and Daryl G. Press. "The new era of counterforce: Technological change and the future of nuclear deterrence." International Security 41.4 (2017): 9-49.

Lieber, Keir A., and Daryl G. Press. "Preventing Escalation during Conventional Wars." (2015).

Maoz, Zeev. "Conflict datasets: Definitions and measurement." International interactions 14.2 (1988): 165-171.

McNamara, Robert S. "The military role of nuclear weapons: Perceptions and misperceptions: Foreign Affairs, Fall 1983." Survival 25.6 (1983): 261-271.

Miles, Simon. "The War Scare That Wasn't: Able Archer 83 and the Myths of the Second Cold War." Journal of Cold War Studies 22.3 (2020): 86-118.

Nagl, John. "Will Ukraine be Afghanistan All Over Again for Russia?" *Foreign Policy*, February 22, 2022.

Nichols, Thomas M., Douglas T. Stuart, and Jeffrey D. McCausland. Tactical nuclear weapons and nato. Strategic Studies Institute, US Army War College, 2012.

Plokhy, Serhii. Nuclear Folly: A History of the Cuban Missile Crisis. WW Norton & Company, 2021.

Powell, Robert. "Crisis bargaining, escalation, and MAD." American Political Science Review 81.3 (1987): 717-735.

Powell, Robert. "Nuclear brinkmanship, limited war, and military power." International Organization 69.3 (2015): 589-626.

Reiter, Dan, Allan C. Stam and Michael C. Horowitz. 2016a. "A Revised Look at Interstate Wars." Journal of Conflict Resolution 60 (5): 956–976.

Reiter, Dan, Allan C. Stam and Michael C. Horowitz. 2016b. "A Deeper Look at Interstate War Data version 1.1." Research and Politics 3 (4): 1–3.

Reiter, Dan. "11 Measurement Replication in Qualitative and Quantitative Studies." The Production of Knowledge (2020): 284.

Rauchhaus, Robert. "Evaluating the nuclear peace hypothesis: A quantitative approach." Journal of Conflict Resolution 53.2 (2009): 258-277.

Sagan, Douglas. The limits of safety: Organizations, accidents, and nuclear weapons. Vol. 53. Princeton University Press, 1995.

Sagan, Scott D., and Jeremi Suri. "The madman nuclear alert: Secrecy, signaling, and safety in October 1969." International Security 27.4 (2003): 150-183.

Sagan, Douglas, and Kenneth Waltz. The spread of nuclear weapons: an enduring debate. WW Norton, 2013.

Schelling, Thomas. The Threat That Leaves Something To Chance. RAND Limited Document HD-A1631-1, 1959.

Schelling, Thomas C. The Strategy of Conflict. Harvard university press, 1960.

Schelling, Thomas C. "Arms and influence." Yale university press, 1966.

Schwarz, Hans-Peter. Konrad Adenauer: A German politician and statesman in a period of war, revolution, and reconstruction. Vol. 2. Berghahn Books, 1995.

Sciutto, Jim. The Return of Great Powers: Russia, China, and the Next World War. Dutton, 2024.

Sechser, Todd S., and Matthew Fuhrmann. "Crisis bargaining and nuclear blackmail." International organization 67.1 (2013): 173-195.

Sechser, Todd S., and Matthew Fuhrmann. Nuclear weapons and coercive diplomacy. Cambridge University Press, 2017.

Trachtenberg, Marc. "Strategic thought in America, 1952-1966." Political Science Quarterly 104.2 (1989): 301-334.

Trachtenberg, Marc. History and strategy. Vol. 1. Princeton University Press, 1991.

Vernon, Graham D., ed. Soviet Perceptions of War and Peace. National Defense University Press, Fort Lesley J. McNair, 1981.

Waltz, Kenneth N. "Nuclear myths and political realities." American Political Science Review 84.3 (1990): 730-745.

Warner, Geoffrey. "The United States and the Suez Crisis." International Affairs 67.2 (1991): 303-317.

Technical Appendix for *Nuclear and Conventional Deterrence in Theory and Practice*

Case Universe

One complicating factor is that almost every empirical article in this literature relies on a different sample for analysis. As noted in the main text earlier articles tended to use Militarized International Disputes (MIDs) data and later articles use the International Crisis Behavior (ICB) or Militarized Compellent Threats (MCT) datasets.

For the empirical papers discussed in the main text, the following samples were used:[66]

Huth, Gelpi, and Bennett (1993): MIDs (with hand-codings), 1832 – 1983, challenger-target dyads, N = 97.

Asal and Beardsley (2007): ICB, 1945 – 2002, non-dyadic, N = 334 (with 53 of these crises dropped due to missingness).

Asal and Beardsley (2009): ICB (Hewitt, 2003, version), 1945 – 2001, dyadic, N = 1218 (with 294 crises).

Kroenig (2013): ICB (with supplemented independent variables and limited to nuclear crises), 1945 – 2001, challenger-target dyads, N = 52 (with 20 crises).

Sechser and Fuhrmann (2013, 2017): MCT, 1918 – 2001, challenger-target dyads, N = 236 (with 210 crises but 6 dropped due to missingness).

Bell and Miller (2015): two datasets.

- Replicating Rauchhaus (2009): MIDs, 1885 – 2000, time series dyadic, N = 611310 (between 103 and 2660 total disputes depending on model with 155701 observations dropped due to missingness).
- Replicating Gartzke and Jo (2009): MIDs, 1945 – 2000, time series challenger-target dyads, N = 1078629 (with 1481 total disputes but 27810 observations dropped due to missingness).

[66] In this section, non-dyadic refers to datasets with one row per case; dyads contain a row for state A and state B as well as state B and state A (with the dependent variable coded according to the sequence in each row); challenger-target dyads are a subset of this where the first state is the initiator and the second state is the defender.

All of these articles are, in some sense, speaking to each other given that the hypotheses, outcome variables, and independent variables are similar. But, differences across samples are substantial and this complicates direct comparisons, especially if one considers missingness, rare event data, selection, etc.

It is also not at all obvious why the MIDs dataset has almost three times as many cases as ICB (see Hewitt, 2003, for an overview). MIDs depend on a dispute becoming militarized. ICB, on the other hand, defines a crisis much more broadly: "a threat to one or more basic values, along with an awareness of finite time for response to the value threat, and a heightened probability of involvement in military hostilities" (Brecher and Wilkenfeld, 2022). Our sympathies lie with ICB in the sense we would like to capture cases that are both militarized and non-militarized, especially since "most international crises erupt in a non-war setting" and then may or may not escalate further (Brecher and Wilkenfeld, 2022). But, the differences in the case universe remain troubling and further analysis is needed here.

Reliability of Existing Results

In what follows, we explore the reliability of several empirical models discussed in the main text of our paper. It is important to note that we are here relying on the same measures as employed by these authors and are not (except for the examination of Kroenig) correcting mistakes in historical work.

Clearly, standards for empirical work have changed in the discipline with the advent of the credibility revolution and widespread use of machine learning methods. Given the substantive area – interstate crises with the threat of war – it is difficult to imagine conducting (natural) experiments. We can, however, examine the reliability of these observational studies by thinking more deeply about nuclear weapons as a treatment variable, running regularized regressions to see if theoretically relevant variables actually contribute to models, and focusing on out-of-sample performance using *k*-fold cross validation.

All of the work we cover here is carefully done but iterative science demands we adopt better methods as they become available. Through no fault of the researchers involved in this literature, we have made substantial strides in statistical methods over the last two decades and it is worthwhile to re-examine prior findings with better methods as they become available.

Huth, Gelpi, and Bennett (1993)

This research explores great powers conflict and advances the following hypothesis: "*If the defender possesses a second-strike nuclear capability, the probability that the challenger will escalate a militarized dispute against the defender will decrease*" (p. 613). A straight-forward interpretation is that state possession of a nuclear arsenal is the treatment variable and we would like to investigate the effect this has on the likelihood of a state initiating a militarized dispute.

To take one famous example – the Franco-Prussian War of 1870 – Huth, Gelpi, and Bennett are investigating the counter-factual for this case. If France had nuclear weapons and a secure second-strike capability in 1870, does Bismark initiate the war?[67] This may seem like an odd framing, but a summary of their sample shows that we are, in fact, considering questions like this. Of the total case universe of 97 cases, 34 occur after 1944. Of these, 31 have nuclear arsenals and a secure second-strike. In other words, there is almost no variance on the treatment variable post-1944.

Typically, we would establish the reliability of the main findings in this paper by conducting an out-of-sample test (de Marchi, Gelpi, and Grynaviski, 2004). The most straight-forward way of doing this (given the fixed, historical sample) is to rerun their model with L1 regularization and lambda set to one standard error from the value that generates the minimum mean cross-validated error.[68] Doing so retains the nuclear variable. But overall model performance is poor, with an out-of-sample deviance ratio of .007. There is also no way to run the model using the more appropriate post-1944 sample given the lack of variance on the treatment variable.

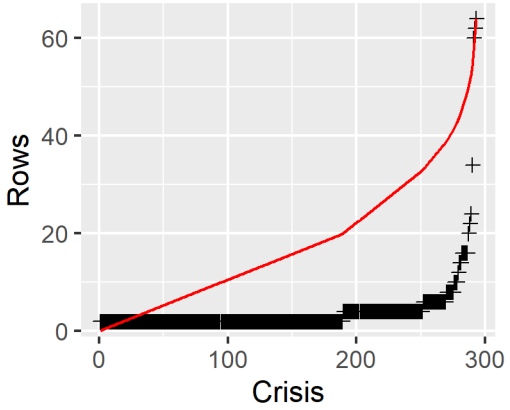
Given the lack of variance on the treatment, issues of selection, overall poor model fit, and possible confounders that coincide with the advent of nuclear arsenals (e.g., the creation of a bipolar system, economic interdependence, force projection capabilities, etc.), it is difficult to trust the main results of this research.

[67] France initiated the conflict, but Huth, Gelpi, and Bennett code Prussia as the challenger. There is a great deal of historical work on this question and coding Prussia as the challenging power is a reasonable (but not entirely straight-forward) choice.
[68] See Hastie et al. (2009) for an overview.

Asal and Beardsley (2009)

This analysis relies on undirected dyads from ICB with cases drawn from 1945 – 2001. The main hypothesis is that states with nuclear weapons "*are more likely to prevail in either gaining concessions or convincing an opponent to back down in their crises,*" especially when faced with states without nuclear weapons. As with the previous article, this work is carefully done but suffers from a number of flaws that are more apparent today than when the research was conducted.

A central difficulty (that is not unique to these authors) is that the dataset is highly imbalanced due to its dyadic construction. The analysis is conducted on a dataset with 1218 observations generated from 294 distinct crises. The vast majority of crises (189 of the 294) have exactly two observations (i.e., one dyad for state A versus state B and another for the reverse). Despite their overall proportion of crises, these cases only make up only 31% of the sample. Unfortunately, four cases with the largest number of involved actors -- three from the Korean War and one from the first Gulf War -- account for 220 observations and 18% of the sample. The figure below shows the number of rows accounted for by each of the 294 crises, with the red line representing the cumulative proportion of the data accounted for by the crises:



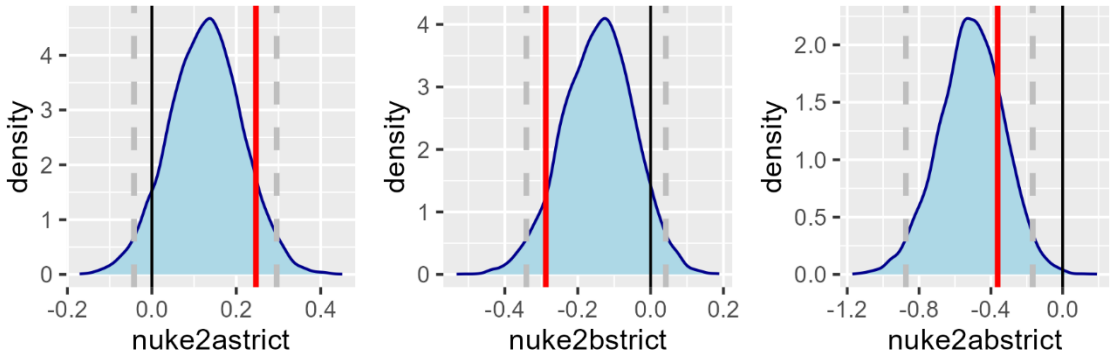
This imbalance is concerning for two main reasons. First, if our goal is to understand nuclear coercion, an imbalanced dataset will not place emphasis on the cases we care about most. The Cuban Missile crisis, for example, is a highly salient case. Nuclear weapon use was on the table and it is, perhaps, the closest the world has come to a strategic exchange. In comparison, the Gulf war was a relatively minor conflict where nuclear weapon use was never considered by the actors involved in the conflict. In Asal

and Beardsley's sample, the Cuban Missile Crisis has 1/16th the weight of the Gulf War; this indicates that we have non-IID cases with very different weights. Second, we have demonstrated that numerous mistakes exist in historical codings for these cases. To the extent that errors exist in cases with larger numbers of actors, this can disproportionately affect the sample.

To examine the effects of imbalanced data, we use a Monte Carlo approach. We create 10000 datasets, in each trial selecting two rows for each crisis without replacement, and then rerun Beardsley and Asal's core models using "victory" and the less restrictive "prevail" as dependent variables.[69] We examine the distribution of the coefficients for Asal and Beardsley's main independent variables of concern here (i.e., the possession of nuclear weapons by each state).

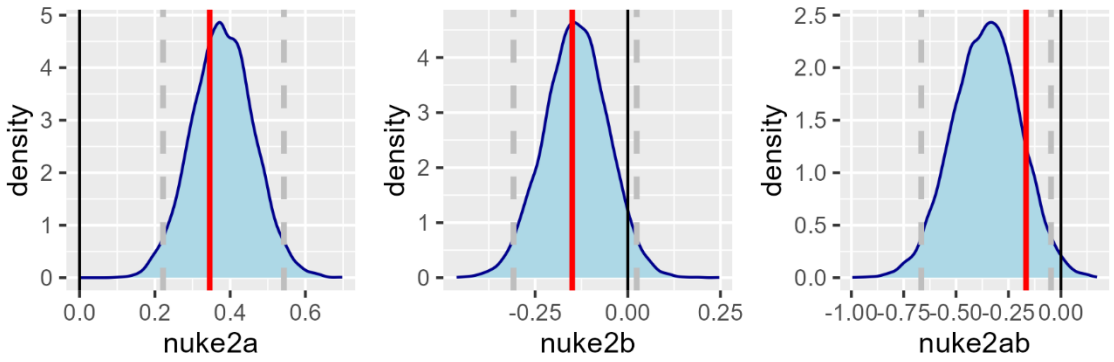
When dealing with the stricter dependent variable (i.e., victory = 1, all other outcomes = 0), our results using balanced samples are on average associated with smaller effects for the possession of nuclear weapons. In particular, Asal and Beardsley's primary result -- that states with nuclear weapons are much more likely to succeed in crisis bargaining -- loses statistical significance at the 10% level in-sample.

The distribution of these variables is presented as a blue density function, with the red line indicating the coefficient estimated in Beardsley and Asal's original models, and the dashed lines representing 95% credible intervals from our Monte Carlo:



[69] In their paper, these are respectively models 2 and 1 from Table 2. The approach we take to examining the results from their paper hold for each of the models they present, but we focus on the first two here to conserve space.

Asal and Beardsley's models are relatively more robust when dealing with the more permissive prevail dependent variable (i.e., set to one if either a victory or compromise is achieved). Here the offensive effect of nuclear weapons is relatively unchanged, and the defensive effect is similar (though it loses conventional statistical significance at the 95% level). The imbalanced data appears, in this simulation, to be somewhat underestimating the interaction effect.



What is apparent in both models is that none of these effects of nuclear weapons have a particular large magnitude. It is thus reasonable to ask how much the possession of nuclear weapons by either state (along with the interaction) adds to overall model fit. Using the original, imbalanced sample, we compare the original model estimated by Asal and Beardsley and a model without the nuclear independent variables. Using the victory dependent variable, this produces an in-sample AUC of .6606 versus .6431, indicating that the inclusion of nuclear covariates has a marginal contribution to model fit. Similarly, using the prevail dependent variable, the in-sample AUC is .6315 versus .6213.

Rerunning these models using the original samples with L1 regularization and lambda set to one standard error from the value that generates the minimum mean cross-validated error also shows that the empirical specifications are very fragile. Out-of-sample deviance ratios for either dependent variable hover around .01 and (depending on the random number seed used) no covariates are retained.[70]

Kroenig (2013)

[70] In models where there exist non-zero coefficients, it is worth noting that nuclear weapons of state A have a small effect.

Kroenig's analysis begins by taking the same set of ICB dyads used in Asal and Beardsley's work, but subsets these data to examine cases where *both* states have nuclear weapons. The main hypothesis focuses on the effect of nuclear superiority, measured either as a ratio of warheads or an indicator that state A has more warheads than state B (which, as we have noted in the main text, are flawed measures of nuclear capabilities). Based on his analysis, Kroenig belives that "*states that enjoy nuclear superiority will be more likely to win nuclear crises*".

A major issue here is that this data relies on only 52 observations derived from 20 crises; miscoding a single case may thus have a disproportionate effect on the results. To show this, we focus on the impact of changing one pivotal case: the Soviet invasion of Afghanistan. We switch the coding of Afghanistan from a Soviet victory (in Kroenig's data) to a US victory (which is a much more reasonable coding given the history of the conflict).

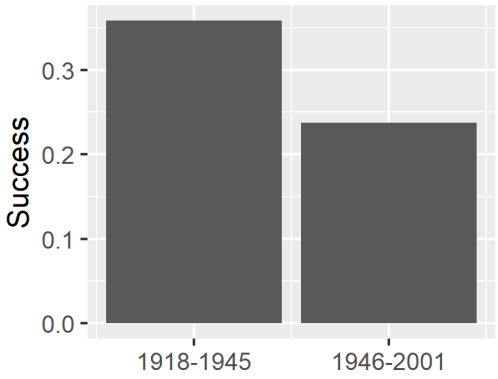
In the full specifications used by Kroenig, the superiority independent variable achieves an in-sample p-value of .003; the nuclear ratio variable achieves .001. Recoding the Soviet invasion changes these to .405 and .057, neither of which achives significance at conventional levels. Unsurprisingly, a model with superiority included has an in-sample AUC of .8235 while a model excluding this covariate has a nearly identical outcome of .8170. For the nuclear ratio covariate, the results are .835 versus .8170. Again, we are faced with a fragile model where the inclusion of nuclear covariates adds very little to overall performance.

Sechser and Fuhrman (2013, 2017)

Sechser and Fuhrmann, in their 2013 article and later in their 2017 manuscript, use the bespoke Militarized Compellent Threat dataset. Their main hypothesis is that "*Compellent threats from nuclear states are no more effective, on average, than threats from nonnuclear states*". This dataset contains 242 compellent dyads between 1918 and 2001 stemming from 210 distinct threats. Of the 242 dyads, 118 of them occurred after 1945 (and the advent of nuclear weapons) and 124 of them were in 1945 or earlier.

Obviously, this runs afoul of the problem identified when comparing different regimes, but even larger issues exist with the sample. Using Sechser and Furhmann's codings, threats were significantly more successful before 1946 than after, and the threats made in the 1946-2001 time period are also substantively different in nature. For example, the

average distance between the states involved in a threat is about 800 miles in the 1918-1945 period, and 1350 in the 1946-2001 period.



If we limit the sample to the post-1945 period where one can make a reasonable case that observations are IID, the list of crises in Sechser and Fuhrmann's data is quite odd. For states that are involved in more than two crises, the distribution in the MCT data are:

Country	Threats
United States	17
United Kingdom	10
Soviet Union/Russia	9
South Africa	7
Iraq	6
India	6
Libya	5
Turkey	5
Israel	5
France	4
Iran	4
China	4

It is worth noting here that 10 of the United States's 17 threats are made after 1989, and the United States only threatens the Soviet Union twice during the Cold War in the MCT data. Even more odd, the Soviet Union *never* threatens the United States. And Israel, in addition to only making five threats, is only twice the target of a threat. We must admit that we are flummoxed at this representation of history after the Second World War.

Turning to Sechser and Fuhrmann's main regression models, we note that in some sense they are on firmer ground than other research in this area: they find null effects for state possession of nuclear weapons. Like the models above, rerunning their regression using L1 regularization reveals that *none* of the covariates are selected. It is, however, difficult to infer overmuch from this finding. Null results can occur for a variety of reasons, many of which we have identified here. And, relying on the MCT sample, given the additional problems it creates, is even more problematic – one might simply arrive a null finding due to lack of signal in the sample (e.g., the Cold War does not exist in MCT).

Summary

The problems we have identified here are common throughout the larger quantiative interantional relations literature. In part, we conducted reliability tests on these examples because they are influential and use the main datasets relied on by interantional relations scholars today. We also chose these models because there were carefully done according to the standards of the field at the time they were published.

Limitations in the quality of the data and the hard fact that nuclear arsenals are both relatively recent innovations and held by a tiny number of states makes inference exceedingly hard in this area. Statistical problems involving non-IID interstate crises, the rare events nature of the main dependent variables, and selection into both treatment and outcome further complicates matters.