Nuclear Superiority Is What States Make of It

The Importance of Methods, Information, and Beliefs*

David C. Logan
Princeton University†

June 25, 2020

Abstract

Does nuclear superiority offer political or military benefits? And, importantly, do these benefits accrue past the acquisition of a secure second-strike capability? International relations theory has long held that nuclear superiority does not confer significant advantages, a conclusion supported by much of the qualitative literature on bargaining and crisis interactions between nuclear-armed states. More recently, scholars have turned to Large-N quantitative methods to answer this question. I argue that this quantitative work, despite some considerable strengths, has suffered from at least two significant and unrealistic assumptions: warhead prominence and complete information. I demonstrate that states 1) use multiple quantitative and qualitative characteristics to evaluate the nuclear balance; and 2) often have inaccurate or incomplete information about the size, composition, and configuration of the nuclear forces of other states. In a replication of previous work, I show that relaxing these assumptions produces systematically different results for the significance of nuclear superiority in interstate crises. Instead of using warhead counts or measures of nuclear forces, I propose an alternative way of measuring state perceptions of the nuclear balance by examining state beliefs about the outcomes of a nuclear exchange. Drawing on archival and interview data from the United States and the Soviet Union during the Cold War, I find little evidence that states possessed consistent, symmetric, and meaningful understandings of the nuclear balance.

*For helpful comments on earlier drafts, I thank Mark Bell, Leyatt Betre, Thomas J. Christensen, Chris Chyba, Francis J. Gavin, Mindy Haas, Rovert Jervis, Pavel Podvig, Travis Sharp, Eric Snyder, Sam Winter-Levy, and members of the Princeton University Security Studies Graduate Research Seminar. Research for this article was supported by the Center for International Security Studies at Princeton University.

†Ph.D. Candidate in Security Studies, dlogan@princeton.edu, https://scholar.princeton.edu/dlogan

In the second winter of his presidency, Donald Trump again traded nuclear-tinged warnings with North Korean leader Kim Jong-un.[1] After Kim reminded the U.S. president of North Korea's burgeoning nuclear capabilities, Trump attempted to return the sentiment, tweeting "North Korean Leader Kim Jong Un just stated that the 'Nuclear Button is on his desk at all times.' Will someone from his depleted and food starved regime please inform him that I too have a Nuclear Button, but it is a much bigger & more powerful one than his, and my Button works!" With this afternoon missive, Donald Trump staked out a position in a debate that has raged between academics and policy-makers since the advent of the nuclear age.

Does nuclear superiority offer political or military benefits? And, importantly, do these benefits accrue past the acquisition of a secure second-strike capability? These have been central questions to the study of interstate conflict. The issue of nuclear superiority has also had important policy implications as U.S. policymakers have debated the political and military utility of the bomb. For decades, much of the theoretical and qualitative scholarship rejected the idea that nuclear superiority mattered, so long as each state enjoyed a secure second-strike capability, though some analysts did find evidence that superiority might have mattered in certain historical cases.

Questions of nuclear superiority have recently reemerged within policy debates across the globe. China's ongoing nuclear modernization project has spurred concerns that Beijing might attempt a sprint to parity which could undermine the U.S.'s nuclear superiority.[2] Nuclear superiority has featured in discussions about how the U.S. should respond to North Korea's growing nuclear capabilities.[3] Some observers, seeing elements of U.S.

1. Peter Baker and Michael Tackett, "Trump Says His 'Nuclear Button' Is 'Much Bigger' Than North Korea's," *New York Times*, January 2, 2018,

2. See, for example, discussion in David C Logan, "Hard Constraints on a Chinese Nuclear Breakout," *The Nonproliferation Review* 24, nos. 1-2 (2017): 13–30.

3. Matthew Kroenig, *The Logic of American Nuclear Strategy: Why Strategic Superiority Matters* (Oxford University Press, 2018), chapter 2.

nuclear modernization as a push for nuclear primacy, have alternatively lauded the efforts as bolstering U.S. security or criticized them as foolish and dangerous.

Scholars have also begun to return to the question of whether nuclear superiority matters by applying statistical methods to data on nuclear crises and compellent threats. This new body of literature is small and its findings mixed, though it has made important contributions to the body of literature by, for instance, theorizing mechanisms through which nuclear superiority can confer bargaining or other political advantages and by remedying some of the methodological weaknesses of their predecessors by, for instance, better matching cases with theory.[4] However, despite these strengths, the empirical components of this literature have generally suffered from at least two crucial weaknesses.

First, the work has relied on overly simplistic measures of the nuclear balance between states, measures which sometimes fail to capture either the actual nuclear balance (assuming such an objective balance even exists and is measurable) or, more importantly, decisionmakers' perceptions of the balance. Analysts of nuclear weapons have long sought methods for assessing, measuring, and comparing the nuclear forces of states. Throughout the Cold War, the U.S. government and independent analysts dedicated significant resources to measuring the strategic nuclear balance between the United States and the Soviet Union. They developed complex measures of nuclear forces: static quantitative measures, such as counts of warheads and delivery vehicles; composite measures, such as counter military potential (CMP) and hard target kill capability (HTK); qualitative measures assessing the accuracy, throwweight, and survivability of delivery systems; and dynamic nuclear exchange models. However, the recent quantitative scholarship investigat-

4. Matthew Kroenig, "Nuclear Superiority and the Balance of Resolve: Explaining Nuclear Crisis Outcomes," *International Organization* 67, no. 1 (2013): 141–171; Kroenig, *The Logic of American Nuclear Strategy: Why Strategic Superiority Matters*; Todd S Sechser and Matthew Fuhrmann, "Crisis Bargaining and Nuclear Blackmail," *International organization* 67, no. 1 (2013): 173–195; Todd S Sechser and Matthew Fuhrmann, *Nuclear weapons and Coercive Diplomacy* (Cambridge University Press, 2017); Matthew Kroenig, Miriam Krieger, and Hans Noel, "Dare to Fail: Nuclear Superiority, Threat Initiation, and Compellent Success," *Typescript, Georgetown University*, 2015,

ing whether nuclear superiority matters has relied on simple counts of the total warheads controlled by a state in a given year. In doing so, they have both elided the myriad other factors which effect the capabilities of a state's nuclear arsenal and, most significantly, they have ignored the information, perceptions, and beliefs of state leaders. Researchers have long noted the important psychological dimensions of nuclear deterrence.[5] But what should and do states consider when assessing nuclear superiority?

Second, the quantitative literature has largely assumed that states share the same (accurate) information about the relative nuclear balance between them. The statistical work has used measures of nuclear capabilities using data published by independent organizations years after the state interactions which form their observations. In doing so, they discount the importance of state perceptions of the nuclear balance at the time of the interaction. Further, they have analyzed this data using symmetric state-year dyads, implicitly assuming that states share the same perceptions of the nuclear balance. However, while states may possess accurate information about their own nuclear arsenals–though for some recent nuclear weapon states, even this may not be true–they may often have imperfect information about the nuclear arsenals of others, whether due to normal intelligence failures or counterintelligence efforts by other states. How accurate are states' intelligence about the nuclear capabilities of other states?

Here, I argue that these two assumptions of warhead prominence and complete information in the quantitative literature often do not hold. First, policymakers across different states and different time periods have focused on different metrics in assessing the relative nuclear balance, metrics which are not captured by recent scholarship. Using an original dataset on state nuclear capabilities covering nearly 60 years, I demonstrate that the results of empirical tests of nuclear superiority depend strongly on how one operationalizes the

5. Philip E Tetlock, Charles B McGuire, and Gregory Mitchell, "Psychological perspectives on nuclear deterrence," *Annual Review of Psychology* 42, no. 1 (1991): 239–276; Robert Jervis, "Cooperation under the Security Dilemma," *World politics* 30, no. 2 (1978): 167–214.

concept. Second, by returning to the historic record and using evidence of actual state information and perceptions, I demonstrate that the assumption of complete information is untenable. Comparing data from declassified U.S. intelligence estimates of Soviet nuclear deployments to the actual number of those deployments, I show that U.S. perceptions of the nuclear balance were often substantially incorrect. In short, the measures used by scholars to assess the nuclear balance have not been adequately coupled to either theories of nuclear superiority or to states' actual experience in assessing the relative nuclear balance.

Having demonstrated the weaknesses of existing measures of the relative nuclear balance, I investigate a measure of nuclear superiority which is more tightly coupled to its empirical referent: states' actual estimates of the relative damage they would suffer in a nuclear exchange. Building on newly published scholarly work, my own archival work, and recently declassified interviews with top Soviet leaders, I show that not only were official relative damage estimates, at best, weakly correlated with their dyadic crisis outcomes, I also demonstrate that state damage assessments varied considerably depending on state assumptions and had a tenuous connection to the technical military features of the nuclear balance. In the end, the findings reinforce a key theoretical tenet of the nuclear revolution: that once states have achieved a secure second-strike capability, nuclear superiority does not confer additional political benefits.

More significantly, however, I draw on the evidence presented here to highlight a key but under-appreciated feature of the relative nuclear balance: that it depends not only on technical military factors, but also on state perceptions and beliefs. These perceptions and beliefs shape how states see and experience the nuclear balance and imbue that balance with meaning. In a nuclear exchange, things like megatonnage and throwweight and the destruction they wreak are real and unchallengeable. But in domains short of nuclear exchanges, such as the brinksmanship contests which have come to characterize the nuclear

age, state beliefs are paramount. In short, here, nuclear superiority is what states make of it.

The article begins with an overview of the major theoretical and empirical works on nuclear superiority. Next, it proceeds to examine the most recent bout of quantitative work examining whether or not nuclear superiority conveys advantages. In doing so, the article highlights two key assumptions in the literature: that warhead counts are an appropriate measure of the strategic nuclear balance, and that states enjoy complete information about the relative nuclear balance. I demonstrate that these assumptions often fail to hold and, in replicating past work on nuclear superiority, I show that results are sensitive to how one operationalizes measures of the nuclear balance. Finally, I propose states' perceptions of anticipated damage in a future nuclear exchange as an improved alternative measure of nuclear superiority. Drawing on archival and testimonial evidence, I conclude that within the U.S.-Soviet Cold War dyad, there is little evidence that superiority mattered.

Three Strands of Nuclear Superiority Research

Work examining the question of whether nuclear superiority matters may be divided into three categories. Early theoretical work on the impact of nuclear weapons on interstate relations largely argued that nuclear superiority was irrelevant so long as the states in question all enjoyed a secure second strike. As the Cold War drew on, the interactions between a growing number of nuclear-armed states provided case studies against which to test these theories. Much of this work appeared to confirm the theoretical arguments that nuclear superiority was largely irrelevant, though there were some findings which suggested the relative nuclear balance could sometimes matter in nuanced ways. Finally, and most recently, large-N statistical work has attempted to determine whether there is a

systematic relationship between nuclear superiority and the outcomes of interstate relations. This work, investigating the outcomes of interstate crises, the initiation of interstate disputes, and the success of coercive diplomacy, has produced somewhat mixed results about the value of nuclear superiority.

The dominant theories of nuclear strategy have largely rejected the concept of nuclear superiority. Writing just after the U.S. use of atomic bombs against Japan, early nuclear strategist Bernard Brodie dismissed the concept, writing: "superiority in numbers of bombs is not in itself a guarantee of strategic superiority in atomic bomb warfare."[6] For Brodie, it was the acquisition and maintenance of a secure-second strike which mattered, not nuclear superiority.[7] Schelling, noting the degree to which nuclear weapons and strategic delivery systems had compressed the speed with which violence could be inflicted, argued that interstate conflict in the modern era would be characterized not by competitions of military strength, but rather by "competitions in risk taking."[8] Because of the speed and assurance with which they could deliver destruction, "Nuclear weapons make it possible to do monstrous violence to the enemy without first achieving victory."[9] Within this world, the relative nuclear balance was insignificant–all states were vulnerable. Jervis, refining the theory, argued that superiority might only matter in a war of attrition, which seemed very unlikely in the nuclear era.[10] Waltz, continuing in this strain, similarly rebuffed the notion that nuclear superiority could convey advantages, arguing that "So long as two or more countries have second-strike forces, to compare them is point-

6. Bernard Brodie, *The Absolute Weapon: Atomic Power and World Order* (New York, Harcourt, 1946), 46.

7. Bernard Brodie, "The Development of Nuclear Strategy," *International Security*, 1978, 65-66; Elsewhere, Brodie illustrates the concept of deterrence by proposing that a hypothetical state which, though conventionally inferior to the Soviet Union, could nonetheless threaten it with "a single thermonuclear bomb" could still enjoy significant deterrent power. See Bernard Brodie, *Strategy in the Missile Age* (Princeton University Press, 1959), 275.

8. Thomas C Schelling, *Arms and influence* (New Haven, CT: Yale University Press, 1966), 92-125.

9. Ibid., 22.

10. Robert Jervis, "Why Nuclear Superiority Doesn't Matter," *Political Science Quarterly* 94, no. 4 (1979): 617-633; Robert Jervis, *The Illogic of American Nuclear Strategy* (Cornell University Press, 1985).

less."[11] He acknowledged that some U.S. policymakers did not share this view but argued this was a failure to appreciate the changes wrought by the nuclear revolution.[12]

Proponents of the benefits of nuclear superiority have argued that states with nuclear superiority will enjoy greater success in coercive bargaining with nuclear inferior states because they expect to suffer less damage in a prospective nuclear exchange.[13] All things being equal, the state which expects to suffer less damage is more likely to press its claims, while the state which expects to suffer more damage is more likely to capitulate.[14] Nuclear superiority, it is argued, might decrease expected damage in one of two ways. First, it might permit the nuclear superior state to launch a successful counterforce strike and eliminate much or all of the adversary's nuclear forces, thereby insulating itself from a nuclear retaliatory strike. Second, in a purely countervalue exchange, the nuclear superior state would be able to inflict more damage than its adversary, again providing it with greater resolve.

The theory of the nuclear revolution and the relative unimportance of the nuclear balance appear to have been largely supported by qualitative case studies. Trachtenberg, in analyzing the Cuban Missile Crisis, concludes that American nuclear superiority did not influence U.S. decisions and that fears of escalation and general nuclear war predominated in American discussions.[15] Betts, in reviewing nearly a dozen examples of nuclear coercion concluded that nuclear superiority was at times an ancillary factor influencing state be-

11. Kenneth N Waltz, "Nuclear Myths and Political Realities," *American Political Science Review* 84, no. 3 (1990): 738.

12. Ibid., 741.

13. Kroenig, "Nuclear Superiority and the Balance of Resolve: Explaining Nuclear Crisis Outcomes," 146-152.

14. Of course, things are never equal.

15. Marc Trachtenberg, "The Influence of Nuclear Weapons in the Cuban Missile Crisis," *International Security* 10, no. 1 (1985): 147-156. Trachtenberg also argues (156-161) that there is some limited evidence to support the idea of asymmetric effects of nuclear superiority. That is, though U.S. officials were not emboldened, Soviet officials may have been cowed by concerns of their relative nuclear inferiority. However, his argument is largely speculation about the Soviet motivations for not alerting their nuclear forces and does not, likely due to lack of sources, point to any evidence that the nuclear balance definitively influenced Soviet thinking.

havior but that, "In crisis decisions the United States exhibited a proclivity toward nuclear coercion that was not strongly governed by the nuclear balance of power."[16] In a review of threatened and actual uses of force since World War II, Blechman and Kaplan concluded that the relative nuclear balance between the United States and the Soviet Union had no impact.[17] Statements from senior U.S. officials also appear to cast doubt on arguments about the importance of the relative nuclear balance. Writing 20 years after the Cuban Missile Crisis, several former Kennedy administration officials rejected the notion that American nuclear superiority–then at its all-time greatest, as measured in warheads–had played any role in the resolution of the crisis.[18] Former Secretary of Defense Robert McNamara has written of a "recognition by U.S. civilian and military officials that NATO's vastly superior nuclear capabilities, measured in terms of numbers of weapons, did not translate into usable military power."[19] Former National Security Adviser McGeorge Bundy, in reviewing the U.S. record of "atomic diplomacy" concluded that nuclear superiority was not a significant factor.[20]

Other scholars have pointed to some episodes as potentially demonstrating the importance of nuclear superiority. For example, Snyder argued that on the question of whether nuclear superiority matters, "Pure logic gives a clear negative to this question" but that "real world experience does not quite follow this logic."[21] However, Diesing, in a footnote at the end of the discussion on nuclear superiority, rebuts Snyder's analysis, arguing that

16. Richard K Betts, *Nuclear Blackmail and Nuclear Balance* (Brookings Institution Press, 1987), 213.

17. Barry M Blechman, Stephen S Kaplan, and David K Hall, *Force without War: US Armed Forces as a Political Instrument* (Brookings Institution Press, 1978), 132.

18. Dean Rusk et al., "The Lessons of the Cuban Missile Crisis," *Time* 27 (1982): 89–92.

19. Robert S McNamara, "The Military Role of Nuclear Weapons: Perceptions and Misperceptions," *Foreign Affairs* 62, no. 1 (1983): 64.

20. McGeorge Bundy, "The Unimpressive Record of Atomic Diplomacy," *The Choice: Nuclear Weapons Versus Security*, 1984, 44-47.

21. Glenn Herald Snyder and Paul Diesing, *Conflict among Nations: Bargaining, Decision Making, and System Structure in International Crises* (Princeton University Press, 1977), 459. The section analyzing nuclear superiority (459-462) rightly notes that what truly matters in questions of the nuclear balance are the beliefs of decisionmakers.

"the empirical evidence on whether nuclear superiority confers any bargaining advantage is so weak that no conclusions can be reached."[22] As noted, both Trachtenberg and Betts have granted the possibility that nuclear superiority (or inferiority, in the case of Trachtenberg) may matter, though they circumscribe when and how it may matter.

Most recently, a handful of studies have attempted to investigate whether nuclear superiority matters by subjecting large-N datasets to statistical regressions. Kroenig analyzes a new set of interstate crises between nuclear-armed states and finds that states which possess a greater proportion of the nuclear weapons in the dyad are more likely to achieve their goals.[23] Sechser and Fuhrmann investigate whether nuclear superiority influences the success of coercive threats issued by states.[24] Using a dataset on militarized compellent threats, they find no evidence to support this. They argue nuclear weapons are uniquely inappropriate for issuing compellent threats because they cannot be used to seize the disputed item and because they entail such high costs for the user. Finally, Kroenig, Krieger, and Noel, in a replication and rebuttal of the Sechser and Fuhrmann work, argue that nuclear superior states are more likely to succeed with compellent threats in part because they are more willing to issue them.[25]

The recent statistical work has made notable contributions. Kroenig formalizes the logic connecting nuclear superiority to relative damage assessments to interstate crisis outcomes. Sechser and Fuhrmann identify and resolve two systematic weaknesses of earlier efforts to assess the effects of nuclear superiority: indeterminate research designs (largely through selection on the dependent variable) and inappropriate quantitative data (includ-

22. Snyder and Diesing, *Conflict among Nations: Bargaining, Decision Making, and System Structure in International Crises*, 462.

23. Kroenig, "Nuclear Superiority and the Balance of Resolve: Explaining Nuclear Crisis Outcomes." For a fuller development of the argument, including qualitative case studies and policy implications, see Kroenig, *The Logic of American Nuclear Strategy: Why Strategic Superiority Matters*.

24. Sechser and Fuhrmann, "Crisis Bargaining and Nuclear Blackmail."

25. Kroenig, Krieger, and Noel, "Dare to Fail: Nuclear Superiority, Threat Initiation, and Compellent Success."

ing observations which do not qualify as coercive threats). This has added to our understanding of how to conceive of, measure, and test, nuclear superiority.

Two Central Assumptions: Warhead Prominence and Complete Information

However, despite the analytical strengths discussed above, the recent quantitative work on nuclear superiority has generally suffered from two related weaknesses.[26] First, it assumes that total warhead counts are the most important metric of nuclear superiority. All three statistical analyses discussed above use a simple total of each state's nuclear warheads. However, this is not the only potential metric for nuclear superiority. Kroenig uses a rationalist model in which nuclear superiority allows a state both to limit damage to itself and increase the damage to the adversary. But the ability to limit damage in a nuclear exchange does not accrue simply in proportion to the aggregate number of warheads possessed by a state. Those warheads must be deployed and made operationally available in ways that permit damage limitation strikes. Even a modest damage limitation capability requires that a state's nuclear arsenal meet a slew of demanding technical and operational criteria, from miniaturized warheads with sufficiently high yield-to-weight ratios to strategic launchers with highly accurate guidance systems. Simple warhead counts do not take into account variation in whether and how those warheads are deployed or even in the characteristics of the warheads themselves, to say nothing of the variation in the ISR and targeting capabilities required for damage limitation strikes.[27]

26. In analyzing the recent quantitative literature on nuclear superiority, I typically focus my discussion to three representative works: Kroenig, "Nuclear Superiority and the Balance of Resolve: Explaining Nuclear Crisis Outcomes"; Kroenig, *The Logic of American Nuclear Strategy: Why Strategic Superiority Matters*; and Sechser and Fuhrmann, "Crisis Bargaining and Nuclear Blackmail."

27. I thank Eric Snyder for this point.

Second, the work has assumed that states and their leaders possess perfect information about the nuclear arsenals of other states. However, states may often have imperfect information about the size and configuration of the nuclear forces of other states. In addition to the general risk of intelligence failures, states often have incentives to obscure or misrepresent their military capabilities, especially in the nuclear realm. What matters in determining state crisis behavior is not the size and capability of the adversary's nuclear forces but, rather, the state perceptions of those forces. Without demonstrating that states possess reasonably accurate information about the nuclear arsenals of the adversary, the quantitative literature on nuclear superiority falls short.

Next, I interrogate these two assumptions, assess their validity, and discuss the implications for the recent statistical work on nuclear superiority.

Assumption of Warhead Prominence

All recent works use the number of nuclear warheads possessed by a state to measure the nuclear balance. In measuring the impact of nuclear superiority on crisis outcomes between nuclear-armed states, Kroenig employs two measures of nuclear superiority: one a dichotomous variable measuring whether the state possessed more nuclear warheads than its counterpart and a second variable measuring the proportion of the nuclear warheads controlled by the state in the dyad. In assessing the impact of nuclear superiority on the success of compellent threats issued by nuclear-armed states, Sechser and Fuhrmann employ three variables measuring nuclear superiority: a dichotomous variable measuring whether the state possessed more nuclear warheads than its target; a "nuclear ratio" variable measuring, like Kroenig, the proportion of nuclear warheads controlled by the state in the dyad; and a variable measuring the difference in the number of warheads possessed by

each state.[28] Kroenig, in, assessing whether or not nuclear superiority matters in the success of compellent threats, similarly utilizes measures of total warhead stockpiles. What all these measures have in common is they measure the total nuclear warhead stockpile of the country. While this measure may present advantages in terms of measurement and conceptual simplicity, it is an inappropriate operationalization of nuclear superiority.[29]

However, as I demonstrate below, a raw counting of a state's nuclear warhead stockpile, however, fails to capture these intuitions for several reasons. In assessing other potential metrics of nuclear superiority, I focus on the nuclear arsenals of the United States and the Soviet Union. This focus presents theoretical, evidentiary, and analytical advantages. First, the Cold War superpower dyad represents a hard case of my claim that nuclear superiority is not mechanically reflected by the technical features of arsenals, but is instead mediated by state beliefs and information. These states invested considerable resources in both knowing the characteristics of the adversary's nuclear forces and in developing concepts for responding to those forces. If any states could claim to know, conceptualize, and respond to the nuclear balance, it is the superpowers. Second, in evidentiary terms, the declassification of U.S. government documents from the Cold War provides a rich source of data on the nuclear forces of the two states. Third, these two states represent a large proportion of the observations in the statistical studies of nuclear superiority. In Kroenig's analysis of nuclear crisis outcomes, either the Soviet Union or the United States participated in 16 of the 20 nuclear crises in the dataset. Both countries appeared in 13 of the crises. In the Sechser and Fuhrmann analysis, either country appears in 24 of 48 of the cases in which a nuclear-armed state issued a compellent threat.

Here, focusing on the U.S. and Soviet Union, I demonstrate why stockpile-based mea-

28. Sechser and Fuhrmann, "Crisis Bargaining and Nuclear Blackmail," 185-186.

29. Though not discussed in detail here, Erik Gartzke, J Kaplow, and R Mehta, "Nuclear Deterrence and the Structure of Nuclear Forces," *Unpublished manuscript, Univ. Calif. San Diego*, 2015, use the same binary and ratio measures of total stockpiles as Kroenig, "Nuclear Superiority and the Balance of Resolve: Explaining Nuclear Crisis Outcomes" to operationalize nuclear superiority.

sures of nuclear superiority (those which merely count the number of nuclear warheads possessed by a state) are potentially inappropriate for four reasons.

First, counting total stockpiles generates measures which will include tactical warheads with little if any relevance to the overall strategic balance. Non-strategic (tactical) nuclear warheads are generally lower-yield and mated to shorter-range delivery systems. They would almost never be used in a direct attack on the adversary's home territory and could not be used in a counterforce strike. Stockpile-based measures of nuclear superiority will systematically misrepresent the nuclear arsenals of states which deploy significant numbers of tactical nuclear weapons, such as the United States and the Soviet Union. For example, using total stockpile size, the U.S. enjoyed greatest relative superiority in 1964, when it possessed 30,751 warheads. However, nearly 23,000 of those warheads were non-strategic warheads, meaning they would likely have had little if any bearing on the strategic nuclear balance. Measures of stockpile size not only misrepresent the size and configuration of an individual state's nuclear arsenal; they can also distort relative measures of nuclear superiority within a state dyad. For example, the total size of the Soviet Union's nuclear warhead stockpile surpassed that of the United States in 1978, when the U.S. ratio of nuclear weapons in the U.S.-Soviet dyad dropped to 48 percent. However, measuring in terms of only strategic warheads, the U.S. still enjoyed significant nuclear superiority over the Soviet Union, possessing 72 percent of strategic bombs within the dyad.

Second, stockpile-based measures of nuclear superiority include both deployed and non-deployed warheads, though the latter should have little if any impact on most interstate negotiations. Non-deployed weapons could, in theory, be relevant over the long term. However, they are unlikely to influence a state's behavior in a short-term crisis. To influence the strategic balance between states, non-deployed weapons would have to be deployed by being mated with delivery vehicles and assigned to military units. Simple stockpile-based measures of nuclear superiority will distort the nuclear arsenals of states

which possess significant numbers of non-deployed weapons. The differences between total warhead stockpile and deployed (or loaded) weapons can again be seen in the U.S. example. In 1966, for instance, the United States had an estimated 11,232 strategic warheads. However, barely half of those were associated with delivery vehicles.[30] In this year, counting all warheads results in a doubling of the relevant U.S. nuclear arsenal.

Third, and related to the inclusion of non-strategic warheads discussed above, stockpile-based measures do not account for weapon yield. If, as suggested by theories of nuclear superiority and demonstrated in Kroenig's formal model, nuclear-superior states are more successful in interstate negotiations because they anticipate suffering less damage than their adversary, then states should care greatly about the destructive power of the adversary's nuclear arsenal. An arsenal composed of low-yield fission weapons would theoretically confer weaker bargaining advantages than one composed of multi-megaton thermonuclear bombs. Stockpile-based measures elide this potentially significant difference by treating all warheads the same. For example, in 1969, the U.S. possessed 16,000 more nuclear warheads than the Soviet Union; but the total megatonnage of the Soviet arsenal had already surpassed that of the U.S. force.

Fourth, stockpile-based measures do not account for the delivery vehicles necessary for executing a nuclear attack. If states are concerned about their relative ability to limit the damage they will suffer in a prospective nuclear exchange, then they should care more about the relative balance of delivery vehicles, not nuclear warheads. For a given number of nuclear warheads, the greater the number of associated delivery systems, the more survivable the nuclear force. This intuition is illustrated by two expressions of U.S. concern about relative inferiority in the nuclear realm compared to the Soviet Union: the so-called

30. Robert Stan Norris and Thomas B Cochran, *US-USSR/Russian Strategic Offensive Nuclear Forces, 1945-1996* (Natural Resources Defense Council, 1997), 12. The 1997 Databook and the Archive of Nuclear Data resources largely agree though where there are differences, I use the data in the Archive of Nuclear data, which was compiled and published more recently.

bomber gap and the later missile gap (these episodes are discussed in more detail below). Again, the distortive effects of stockpile-based measures in ignoring delivery vehicles can be seen in the Cold War arsenals of the United States and the Soviet Union. In 1972, the U.S. controlled nearly two-thirds of all the warheads possessed between it and the Soviet Union. But that same year was also the first in which the Soviet Union deployed more strategic delivery vehicles than the United States.

Some of the statistical work appears to recognize that nuclear capabilities may be measured in other ways but ultimately dismisses these measures as either unimportant or redundant. For example, Kroenig observes that "Nuclear analysts often consider additional factors when calculating the nuclear balance between states including: total megatonnage, numbers and accuracy of delivery vehicles, and the ability of command-and-control systems to execute war plans in a crisis."[31] However, he argues that using a stockpile-based measure is nonetheless appropriate both because "detailed information on these variables is not available for every nuclear weapon state in every year" and because "there is good reason to believe that simple warhead counts and more complicated assessments of nuclear capabilities are highly correlated."[32] Kroenig's assumption of correlation between different measures of nuclear capabilities likely holds true across many nuclear-armed state dyads in many years. For example, there is no year in which, say, China could be said to have enjoyed technical nuclear superiority over the United States or the Soviet Union.

However, the assumption of correlation fails to hold in important instances, especially for the U.S.-Soviet dyads. To illustrate this, I construct a new database of nuclear capabilities for the two superpowers across five measures. Figure 1 shows the nuclear balance between the United States and the Soviet Union from 1945 to 2002 across five different measures: nuclear warheads stockpile (stockpile), strategic warheads stockpile (strategic

31. Kroenig, "Nuclear Superiority and the Balance of Resolve: Explaining Nuclear Crisis Outcomes," 155.
32. Ibid.

stock), strategic delivery vehicles (launchers), warheads loaded on strategic delivery vehicles (loaded), and total megatonnage (megatonnage). The figure plots the number of U.S. capabilities for a given measure minus the number of Soviet capabilities. In other words, the higher the line above zero, the greater the U.S. superiority in that measure; the lower the line is below zero, the greater the Soviet superiority in that measure. The solid black line represents the relative nuclear balance as measured by the overall warhead stockpiles of the United States and the Soviet Union. This is the measure that has been employed in statistical analyses of nuclear superiority. As shown in the figure, there appears to be some co-movement in the relative balances of stockpile size and strategic delivery vehicles, though the latter remains at a much lower level. However, the relationship between overall stockpile and other measures is more tenuous. The balance in overall megatonnage reversed a full decade earlier than the reversal in overall stockpile size. Conversely, the United States maintained superiority in both the size of its strategic warhead stockpile and the number of warheads loaded onto strategic delivery vehicles. Throughout this period, the outcomes of crises and threats between the United States and the Soviet Union would have been overdetermined by U.S. nuclear superiority in strategic warheads and loaded strategic warheads.

To better illustrate the degree to which stockpile-based measures can distort assessments of the nuclear balance, I also construct a measure of the relative nuclear balance across each of these five measures. The measure represents the proportion of capabilities possessed by the United States in a given year. For example, a value of 1 indicates that the United States controlled all of the relevant nuclear capabilities in a given year (or, said differently, the Soviet Union deployed none of those capabilities). A value of 0.5 indicates that both the United States and the Soviet Union possessed the exact same level of capabilities. Values closer to 1 indicate higher degrees of U.S. nuclear superiority while values closer to 0 indicate higher degrees of Soviet nuclear superiority. In Figure 2 I plot

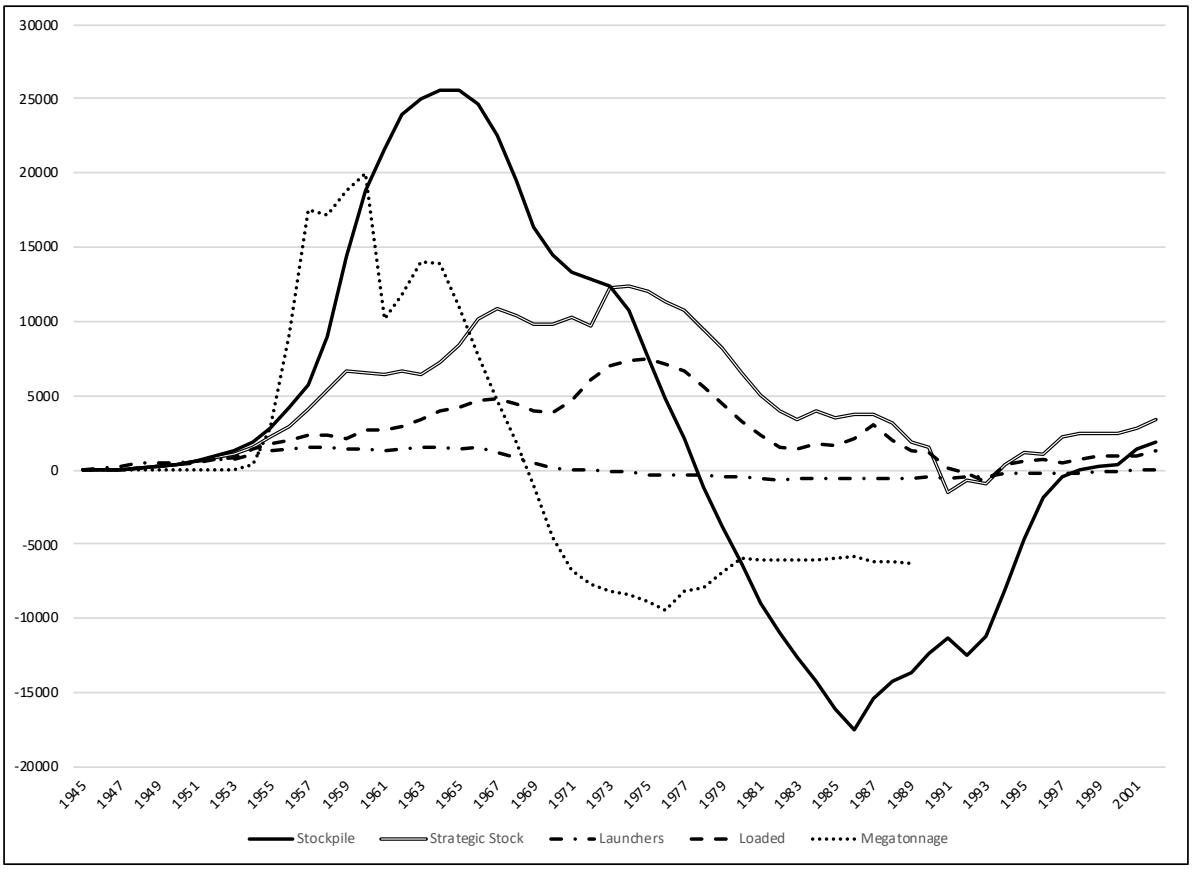

Figure 1: U.S.-Soviet Union absolute nuclear balance, 1945-2002

the relative nuclear balance along each of the five measures of nuclear capabilities.

The figure shows that depictions of the U.S.-Soviet nuclear balance often depended significantly on the measure that is used. Though the trend of each measure is generally toward greater parity over the course of the Cold War, they do not always move in the same direction. Most significantly, even when the measures do move together, they move at dramatically different rates. By the mid- to late-1960s, assessments of the nuclear balance would have depended heavily on what measure one used. This disparity only increased with time. In 1976, a stockpile-based measure reported a value of 0.55, indicating almost perfect parity. However, in the same year, the United States had more than three-quarters of all strategic warheads and strategic warheads loaded onto delivery vehicles, while the Soviet Union had a majority of strategic launchers and nearly three-quarters of total mega-

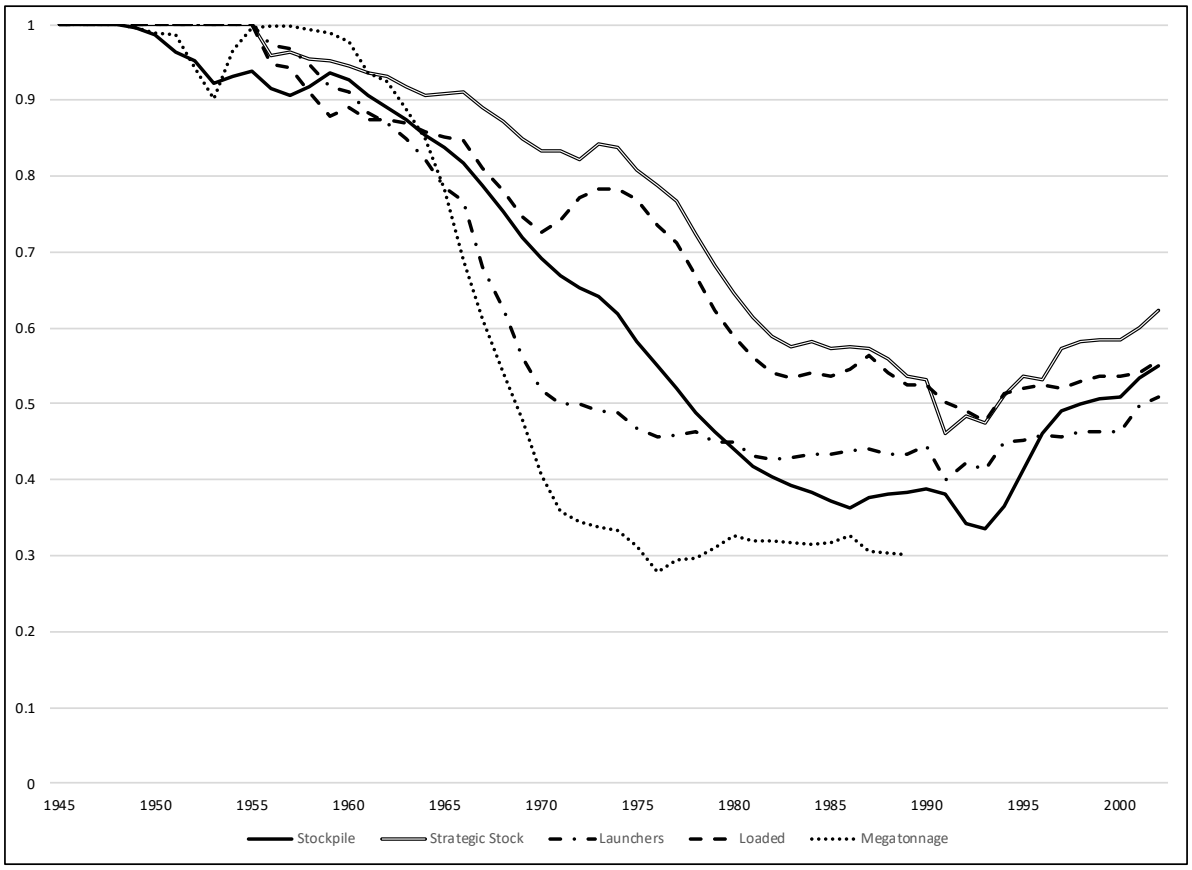

Figure 2: U.S.-Soviet Union relative nuclear balance, 1945-2002

tonnage. It is true that until the early 1960s, the U.S. enjoyed significant superiority across all five measures. However, observations of interstate relations under conditions of such extreme superiority raise questions about whether it is nuclear superiority that matters or whether the inferior state believes that it no longer possesses a secure second-strike capability.[33]

The historical record demonstrates that U.S. analysts and policymakers were more attuned to more comprehensive measures of nuclear capability than simple aggregate war-

33. By "nuclear superiority" here, I mean "numerical superiority in warheads," as it has been operationalized by some of the statistical work. For other works finding that survivability (and diversification) is more important than superiority, see Gartzke, Kaplow, and Mehta, "Nuclear Deterrence and the Structure of Nuclear Forces"; and Dana Higgins, Connor Huff, and Anton Strezhnev, "Survivability not Superiority: A Critique of Kroenig (2013)," 2013,

head counts. The use of more nuanced measures of nuclear capabilities can be seen in at least two features of the Cold War. First, the early warnings in the United States of an impending bomber gap and a later missile gap demonstrate that U.S. analysts and policymakers focused on delivery systems, not on warheads. From the late-1940s to the mid-1950s, U.S. officials became (mistakenly) concerned that the Soviet Union had gained a significant strategic advantage over the United States by deploying large numbers of jet-powered long-range nuclear-capable bombers.[34] In the late 1950s, analogous fears emerged of a "missile gap," following the Soviet launches of the SS-6 ICBM and the Sputnik satellite, which were seen as signaling Soviet technological superiority.[35] Those fears were further heightened by the publication by the President's Science Advisory Committee of the so-called Gaither Report, which warned that "By 1959, the USSR may be able to launch an attack with ICBMs carrying megaton warheads, against which [Strategic Air Command] will be almost completely vulnerable under present programs."[36] Throughout both episodes, U.S. officials were much more concerned with a relative inferiority in delivery vehicles (bombers and, later, missiles) than with nuclear warheads. The 1948 Finletter Report warning of an impending bomber gap, for instance, focused on the development and production of nuclear-capable aircraft, with little reference to the bombs they would deliver.[37] Similarly, the measures recommended by the Gaither Report to ensure the survivability of the U.S. deterrent focused on delivery systems, including guidance to increase bomber alert status, harden and disperse Strategic Air Command facilities, and ramp up production of ICBMs.[38] Concerns in the United States in the 1970s of an alleged "window of vul-

34. Luke Benjamin Wells, "The 'Bomber Gap': British Intelligence and an American Delusion," *Journal of Strategic Studies* 40, no. 7 (2017): 963–989.

35. Greg Thielmann, "The Missile Gap Myth And Its Progeny," *Arms Control Today* 41, no. 4 (2011): 44.

36. Office of Defense Mobilization, "Deterrence and Survival in the Nuclear Age," *Washington, DC: Government Printing Office*, 1957, 14.

37. Thomas Knight Finletter, "Survival in the Air Age: A Report by the President's Air Policy Commission," *Washington, DC: Government Printing Office*, 1948,

38. Defense Mobilization, "Deterrence and Survival in the Nuclear Age," 6-7.

nerability" centered around estimates of the throwweight and accuracy of Soviet strategic delivery vehicles.[39]

Second, national intelligence estimates on Soviet nuclear capabilities produced during the Cold War focused almost exclusively on delivery systems. At the start of the Cold War, U.S. policymakers recognized the need for better strategic intelligence and tasked the burgeoning intelligence community with the production of intelligence assessments, including several series of National Intelligence Estimates.[40] The U.S. intelligence community consistently produced two series of NIEs addressing the nuclear capabilities of the Soviet Union. One, the 11-2 series, reported on developments in Soviet atomic energy, including progress on nuclear reactors, nuclear materials production, nuclear weapons development, and atomic aid provided to other states. Estimates of the Soviet nuclear arsenal did not feature prominently in these analyses. For example, the 1959 NIE on the Soviet atomic energy program noted that "We lack sufficient evidence to support a firm estimate of the Soviet weapons stockpile by number, by type, by mission, or otherwise."[41] Similarly, the 1963 NIE estimated cumulative Soviet U-235 production at 130,000 kilograms but that, considering margins of error, the true number could have fallen anywhere between

39. Pavel Podvig, "The Window of Vulnerability That Wasn't: Soviet Military Buildup in the 1970s?A Research Note," *International Security* 33, no. 1 (2008): 118-138.

40. For a brief history of the origins of the NIEs, see Donald Paul Steury, *Intentions and Capabilities: Estimates on Soviet Strategic Forces, 1950-1983* (History Staff, Center for the Study of Intelligence, Central Intelligence Agency, 1996), xi-xxii. For a summary of the process by which contemporary National Intelligence Estimates are written and their role in the policy process, see Greg Bruno and Sharon Otterman, "National Intelligence Estimates," *Council on Foreign Relations* 14 (2008).

41. National Intelligence Estimate 11-2-59 1959, 59. See Document A1 in the Appendix for active citation. For more on active citation and increasing transparency in qualitative international relations and security studies scholarship, see Andrew Moravcsik, "Active Citation: A Precondition for Replicable Qualitative Research," *PS: Political Science & Politics* 43, no. 1 (2010): 29–35, Andrew Moravcsik, "Active Citation and Qualitative Political Science," *Qualitative and Multi-Mmethod Research* 10, no. 1 (2012): 33–37, Andrew Moravcsik, "Trust, but Verify: The Transparency Revolution and Qualitative International Relations," *Security Studies* 23, no. 4 (2014): 663–688; Diana Kapiszewski and Dessislava Kirilova, "Transparency in Qualitative Security Studies Research: Standards, Benefits, and Challenges," *Security Studies* 23, no. 4 (2014): 699–707; and Elizabeth N Saunders, "Transparency without Tears: A Pragmatic Approach to Transparent Security Studies Research," *Security Studies* 23, no. 4 (2014): 689–698.

80,000 and 180,000 kilograms.[42] Though the 1959 report presents "illustrative allocations" of Soviet nuclear forces, it cautions that "In the light of the range and complexity of the factors discussed above, we do not believe it is possible or desirable to arrive at any single 'most probable' estimate of the Soviet weapons stockpile at selected periods."[43] These NIEs largely focused on qualitative changes in the Soviet nuclear weapons program and do not appear to provide any usable estimate of the size of the Soviet stockpile until 1964 and even these are presented with significant caveats.[44]

A second set of annual NIEs, the 11-8 series, focused on Soviet nuclear forces which could be used to attack the continental United States. The substantive focus of these NIEs is demonstrated by the sections into which they are divided, with each estimate reporting developments on, in order: Soviet ICBMs, Soviet space-based capabilities, Soviet submarine-launched ballistic missiles, and Soviet long-range nuclear-capable bombers. In fact, by late 1955, DCI Allen Dulles declared that collecting reliable intelligence on Soviet ICBMs was "of the highest priority, probably of even greater ultimate importance to our national security than atomic energy intelligence."[45] In short, U.S. intelligence estimates on Soviet nuclear capabilities generally discounted warhead counts both because of the difficulty of developing such estimates and because of their relative unimportance compared to estimates of delivery systems.

Even the five measures discussed here fail to encompass the universe of measurements of nuclear forces. As early as the 1960s, observers were attempting to measure and compare the capabilities of strategic nuclear forces to predict how the strategic balance might

42. National Intelligence Estimate 11-2A-63 1963, 37. See Document A2 in the Appendix for active citation.
43. National Intelligence Estimate 11-2-59 1959, 62. See Document A1 in the Appendix for active citation.
44. References to the estimates are made at National Intelligence Estimate 11-2-64 1964, 29, though the estimates themselves appear to have been excised from the declassified document. It is possible that similar estimates were made in earlier NIEs and had merely been excised from the declassified versions, though there is no indication of this. Some earlier NIEs included "illustrative allocations," though the authors were careful to emphasize that these were notional and not meant to provide any meaningful estimate of Soviet nuclear forces. See Document A3 in the Appendix for active citation.
45. As quoted at Steury, *Intentions and Capabilities: Estimates on Soviet Strategic Forces, 1950-1983*, 55

change under any future arms control agreements.[46] Over the years, analysts have concocted numerous complex metrics by which to measure and compare nuclear capabilities, most of which enjoy their own acronym, including: equivalent megatons (EMT), lethal area potential, counter military potential (CMP), hard target kill capability (HTK), equivalent weapons (EW), joint throwweight and warhead measures, standard weapon stations (SWS), and distinct blasts index (DBI).[47] The five measures of nuclear forces discussed above (total warheads, strategic warheads delivery vehicles, loaded warheads, and total megatonnage) are, in fact, quite crude by comparison. Applying these complex measures to the same nuclear forces often produces very different assessments of the nuclear balance and some observers may cherry pick their measure to lobby for a preferred policy.[48]

It is difficult to say, ex ante, which measure of nuclear superiority is most appropriate. The metrics that matter may vary depending on what kind of nuclear exchange is anticipated. In a counterforce nuclear exchange, perhaps the number of strategic delivery vehicles matters most, since they better represent the survivability of the force.[49] However, in a countervalue exchange, loaded warheads or total megatonnage may matter more since they better measure the destructive capability of the state's force. The importance of dif-

46. See, for example, Glenn A Kent, *On the Interaction of Opposing Forces under Possible Arms Agreements*, 5 (Center for International Affairs, Harvard University, 1963); Thomas A Brown, "US and Soviet Strategic Force Levels: Problems of Assessment and Measurement," *The Annals of the American Academy of Political and Social Science* 457, no. 1 (1981): 18–27; Jeffrey T Richelson, "Evaluating the Strategic Balance," *American Journal of Political Science*, 1980, 779–803; Jeffrey T Richelson, "Static Indicators and the Ranking of Strategic Forces," *Journal of Conflict Resolution* 26, no. 2 (1982): 265–282; and Fred A Payne, "The Strategic Nuclear Balance: A New Measure," *Survival* 19, no. 3 (1977): 107–110.

47. For a review of most of these measures, see Sydell P Gold, "Report of the Chairman: What Role Can Analysis Play in Illuminating Arms Control Objectives and Options?," in *Modelling and Analysis in Arms Control* (Springer, 1986), 132-135. The Soviet Union developed its own measures by which to evaluate the effectiveness of nuclear forces. See Claire M Levy, *Soviet Strategic Nuclear Measures of Effectiveness*, technical report (RAND Corp. Santa Monica, CA, 1992)

48. Richelson, "Static Indicators and the Ranking of Strategic Forces."

49. Assumptions of other variables, such as the performance characteristics of the delivery vehicles and the nature of the nuclear strike, would also significantly change assessments of the strategic nuclear balance. For a demonstration of how predicted outcomes of a nuclear exchange change along some of these variables, see Steinbruner and John D Steinbruner and Thomas M Garwin, "Strategic Vulnerability: The Balance between Prudence and Paranoia," *International Security*, 1976, 138–181.

ferent metrics may also vary with changes in technology. For instance, in the early years of the Cold War, when delivery vehicles could only carry a single warhead, the number of strategic launchers may have been most significant. But following the development of multiple independently targetable reentry vehicles, warhead loadings may have gained prominence. The salience of different metrics might also shift with changes in political or military leadership. For instance, for political leaders with relatively "unsophisticated" understandings of nuclear weapons, "pre-attack" indicators such as the number of deployed warheads may be most salient, while military professionals may be most focused on more complex exchange models.[50] However, without first interrogating the historical record, it is difficult to say which measures will be most salient for decisionmakers and, therefore, difficult to construct a plausible test.

Evidence from official U.S. documents suggests that analysts have, at times, been interested in all these and other measures of nuclear capabilities. Government documents suggest substantial nuance in considering Soviet nuclear forces, incorporating measures of number of delivery vehicles, on-line throw weight, missile accuracy, and total megatonnage. At times, these assessments seemed to implicitly reject the bean-counting approach when assessing the relative nuclear balance and its strategic implications. For example, 1969 was the first year in which Soviet strategic delivery vehicles were estimated to outnumber those of the United States.[51] However, the NIE distributed that year was careful to acknowledge the diversity within this force and, significantly, its inability to execute a counterforce strike against the United States, noting that "The SS-9 is the only [Soviet] ICBM with the combination of payload and accuracy to attack hard targets effectively, but in its present numbers with single warheads it could attack no more than a small percent

50. Richelson, "Evaluating the Strategic Balance," 781, 801; Snyder and Diesing, *Conflict among Nations: Bargaining, Decision Making, and System Structure in International Crises*, 459.

51. I include ICBMs, SLBMs, and long-range bombers.

of the US ICBM force."[52] This kind of nuanced analysis as to the actual operational capabilities of a state's nuclear arsenals is lost in the recent statistical work purporting to examine whether or not nuclear superiority matters. The difficulties presented in conceptualizing and measuring nuclear superiority should itself suggest that leaders may not be attuned to it in the midst of crises. Even the measures which presumably best reflect the strategic nuclear balance–strategic stockpile and warheads loaded on strategic launchers–suggest that nuclear superiority played little if any role between the United States and Soviet Union as the United States maintained superiority across both measures for the entirety of the Cold War.

As proponents of large-N analyses have rightly argued, questioning the coding of individual cases in large-N datasets does not necessarily invalidate analyses which rely on those coding decisions.[53] However, if the operationalizations of key concepts are systematically biased, then analyses which use them will produce unreliable results. To demonstrate how different operationalizations can produce substantially different results, I conduct a sensitivity analysis of some key claims about the political impact of nuclear superiority. First, I reconstruct the outcomes of crises involving the Soviet Union and the United States in the Kroenig 2013 dataset. The outcomes involving these dyads are significant both because they represent much of the world's experience with nuclear crises and because they therefore constitute such a large portion of the relevant datasets. For instance, Kroenig identifies 13 crises involving the U.S.-Soviet dyad, out of a total of 20 in the entire dataset.[54] The U.S.-Soviet pairings account for 26 of the 52 total directed dyads in the dataset making them extremely significant for the paper's overall conclusions. Below, I

52. National Intelligence Estimate 11-8-69 1969, 3. See Document A4 in the Appendix for active citation.

53. See, for example, Matthew Fuhrmann, Matthew Kroenig, and Todd S Sechser, "The Case for Using Statistics to Study Nuclear Security," in *H-Diplo/International Security Studies Forum*, vol. 2 (2014), 37–54. Though, in some instances, the recoding of individual cases, can change results. See Mark S Bell and Nicholas L Miller, "Questioning the Effect of Nuclear Weapons on Conflict," *Journal of Conflict Resolution* 59, no. 1 (2015): 74–92.

54. In re-examining the data, I use the original coding decisions of the International Crisis Behavior Project.

	U.S. Success	Soviet Success	Total Warheads	Strategic Warheads	Strategic Delivery Vehicles	Loaded Strategic Warheads	Total Megatonnage
Korean War	*N*	*Y*	*US*	*US*	*US*	*US*	*US*
Suez Crisis	*Y*	*Y*	*US*	*US*	*US*	*US*	*US*
Berlin Deadline	*N*	*N*	*US*	*US*	*US*	*US*	*US*
Berlin Wall	*N*	*Y*	*US*	*US*	*US*	*US*	*US*
Cuban Missile Crisis	*Y*	*N*	*US*	*US*	*US*	*US*	*US*
Congo War	*Y*	*N*	*US*	*US*	*US*	*US*	*US*
Six Day War	*Y*	*N*	*US*	*US*	*US*	*US*	*SOV*
Cienfuegos	*Y*	*N*	*US*	*US*	*US*	*US*	*SOV*
Yom Kippur	*Y*	*N*	*US*	*US*	*SOV*	*US*	*SOV*
Angola	*N*	*Y*	*US*	*US*	*SOV*	*US*	*SOV*
Afghanistan	*N*	*Y*	*SOV*	*US*	*SOV*	*US*	*SOV*
Able Archer	*N*	*N*	*SOV*	*US*	*SOV*	*US*	*SOV*
Nicaragua	*N*	*N*	*SOV*	*US*	*SOV*	*US*	*SOV*

	Total Warheads	Strategic Warheads	Strategic Delivery Vehicles	Loaded Strategic Warheads	Total Megatonnage
Superior Victory Proportion	0.53	0.46	0.53	0.46	0.38
Inferior Victory Proportion	0.31	0.38	0.31	0.38	0.53

Table 1: Sensitivity test on conceptualizations of strategic nuclear superiority

list each crisis and note whether the U.S. or Soviet Union achieved its objectives (Y) or did not (N).[55] I also indicate which state possessed superiority along one of the five measures of nuclear capabilities discussed above. Finally, I record the overall "win" rates of the superior and inferior states according to each those measures of nuclear capabilities.

Table 1 (above) indicates that, in the U.S.-Soviet context, determinations about which state enjoyed nuclear superiority depend on how one operationalizes the concept. As noted above, when superiority is measured according to number of strategic warheads or number of strategic warheads loaded onto delivery vehicles, the U.S. enjoyed superiority in every crisis with the Soviet Union, meaning there would be no change in the independent variable. Assessing the causal impact of one variable on another requires that the dependent and independent variables co-vary. Without variation along the independent

55. According to the original coding rules, it is possible for both or neither state to achieve its objectives and "win" the crisis.

variable (nuclear superiority as measured by strategic warheads or loaded strategic warheads), no conclusions can be drawn about its relationship to the dependent variable. By contrast, if superiority is measured according to total megatonnage, the Soviet Union enjoyed superiority in a majority of crises with the United States. In addition, conclusions about the impact of superiority vary with the operationalization. The results also indicate that broader conclusions about the importance of superiority depend on how that superiority is defined. If, as in the original dataset, superiority is measured by aggregate warheads, then the nuclear superior state "wins" a majority of its crises, whereas the inferior state "wins" only a third of the time. By comparison, measuring by total megatonnage, the nuclear inferior state wins a majority of its crises and is actually more likely to "win" a nuclear crisis than the nuclear superior one.

To more fully investigate the extent to which recent quantitative findings vary with how the nuclear balance is measured, I replicate the results of Kroenig's work examining the impact of nuclear superiority on crisis bargaining outcomes. I begin with Kroenig's dataset and make two alterations. First, following Sechser and Fuhrmann, I recode the identifier used for clustering standard errors by crisis dyad.[56] Second, I construct an original dataset consisting of five separate measures of the nuclear balance: strategic stockpile, strategic delivery systems, loaded strategic warheads, and total megatonnage.

Introducing these alternative measures of the nuclear balance clearly demonstrates the significance of how nuclear superiority is measured. When measuring the nuclear balance with total stockpile, there are 52 relevant directed dyads. However, when using one of the alternative measures, the number of relevant directed dyads drops. For example, there are only 32 directed dyads in which both states possessed at least one strategic nuclear delivery vehicle. This is because while researchers have typically coded a country as a

56. Todd S Sechser and Matthew Fuhrmann, "Debating the Benefits of Nuclear Superiority, Part III," *The Duck of Minerva (March 28). Available at http://www. whiteoliphaunt. com/duckofminerva/2013/03/debating-the-benefits-of-nuclear-superiority-part-iii. html*, 2013,

nuclear weapon state in the first year in which it explodes a nuclear device, it may take much longer for the state to develop other components of its nuclear arsenal.[57] It's not clear how these observations should be treated. Should they be included in models with the relevant nuclear capabilities zeroed out? Or should they be dropped entirely from the models? According to the most commonly used definition in the literature, these are nuclear-armed states, and yet they may not be relevant to the analysis. Here, for the purposes of robustness, I estimate models both including and censoring these observations, though for brevity I only report results from the more inclusive models.[58]

Using this new data and retaining all 52 original dyads, I re-estimate the full models reported in the original Kroenig 2013 analysis, including all control variables. In total, I run 10 models, two for each of the five measures of nuclear balance. Following the original analysis, for each measure, I run two separate models: one measuring superiority as a binary variable, and another measuring superiority as the ratio of nuclear capabilities possessed by one state divided by the total nuclear capabilities possessed by both states in the dyad. The results of the models for the variable of interest (nuclear superiority) are reported in Table 2.

The results motivate two conclusions. First, the results of the analysis depend significantly on how nuclear superiority is measured. In the two models in which the nuclear balance is determined by total nuclear stockpiles, the results are statistically significant at the .05-level. In the model employing the ratio of loaded strategic warheads, the results are significant at the .10-level. In the remaining seven models, however, the results are not sta-

57. Jacques EC Hymans, "When Does a State Become a "Nuclear Weapon State"? An Exercise in Measurement Validation," *Nonproliferation Review* 17, no. 1 (2010): 161–180; Mark S Bell, "Beyond Emboldenment: How Acquiring Nuclear Weapons Can Change Foreign Policy," *International Security* 40, no. 1 (2015): 87–119; Mark S Bell, "Nuclear Opportunism: A Theory of How States Use Nuclear Weapons in International Politics," *Journal of Strategic Studies* 42, no. 1 (2019): 3–28.

58. In the "inclusive" model, states with zero values on a given metric are included. If both states possess zeroes, neither is coded as possessing binary superiority and the ratio of capabilities is set as equal to 0.5. See the appendix for these results.

	Total Stockpile	Strategic Stockpile	Strategic Delivery Vehicles	Strategic Loadings	Total Mega-Tonnage
Nuclear Superiority (binary)	2.005** (0.953)	0.763 (0.739)	1.124 (0.715)	0.763 (0.739)	0.458 (0.634)
Nuclear Ratio	4.252** (2.160)	2.268 (1.526)	1.376 (1.441)	2.672* (1.551)	1.315 (1.293)

Standard errors in parentheses. Control variables not shown. * p <0.1, ** p <0.05, *** p <0.01

Table 2: Logit analyses of impact of various forms of nuclear superiority on crisis success

tistically significant at any conventional threshold. Without a well formulated argument about which indicator is most relevant to states' actual understandings of the nuclear balance, it's not clear how to interpret these results. Second, and relatedly, the overall results suggest that nuclear superiority was not a significant factor in crisis outcomes. In the majority of models, the measure of nuclear superiority is statistically insignificant, including most of those which, a priori, would seem the most relevant for state perceptions, such as strategic delivery vehicles and strategic warhead loadings.

Assumption of Complete Information

The second major weakness in the statistical nuclear superiority literature is that it has given insufficient attention to the role of perceptions. All the quantitative work discussed here has used data on nuclear warheads which has been published by non-governmental sources years or decades after the fact.[59] In doing so, the work makes two crucial implicit assumptions. First, it assumes that states possessed accurate information about the relative nuclear balance. Second, it assumes that this data accurately represents the nuclear arsenals of the time. Here, I review each assumption and demonstrate why they may not

59. All three works rely on data published by either the Natural Resources Defense Council or the Bulletin of the Atomic Scientists. In both instances, as is discussed below, these independent analyses draw on unclassified and declassified U.S. government sources.

hold.

First, the work assumes that states have accurate information and, in doing so, assumes that what matters is the actual nuclear balance and not what states subjectively perceive. By using undirected dyads, the statistical work has implied that states possess the same information about the nuclear balance. In reality, though, each state possesses, at best, perfect information about only its own arsenal and holds imperfect information about the arsenal of the adversary (though some states, especially ones with relatively recent nuclear forces, may not even have complete information about their own capabilities). In this way, both states in a dyad might simultaneously believe they are in a state of nuclear superiority or inferiority, depending on both their intelligence about the nuclear forces of the adversary and their beliefs about what determines the nuclear balance. For some more balanced dyads which appear frequently in the datasets–such as the U.S.-Soviet Union and India-Pakistan dyads–small errors in intelligence assessments could produce importance differences in perceptions of the nuclear balance. Information on nuclear arsenals is often kept secret and guarded by elaborate counterintelligence operations.[60]

Second, the quantitative work uses data which may not accurately represent states' nuclear arsenals. All the work uses data on warhead counts published by non-governmental sources. However, much of this data is, at base, compiled from U.S. government estimates. This means that while data on the U.S. nuclear arsenal is likely accurate, data on the nuclear arsenals of other states is, at best, a composite estimate.

I now assess the accuracy of these assumptions. First, I examine the extent to which states' real-time perceptions of the nuclear balance reflected reality. To determine the accuracy of those perceptions requires information both on what states believed about the nuclear arsenals of their adversaries and a "true" baseline against which to compare those

60. Oleg A Bukharin, "The Cold War Atomic Intelligence Game, 1945-70: From the Russian Perspective," *Studies in Intelligence* 48, no. 2 (2004): 1–11.

beliefs. Here, I examine the real-time assessments of the United States. I construct a new dataset containing estimates by the U.S. intelligence community of the size and composition of Soviet strategic nuclear forces from 1960 to 1977.[61] I draw from the 11-8 series of the National Intelligence Estimates which, in each year, provided estimates of the number of operational Soviet ICBMs, SLBMs, and long-range nuclear-capable bombers. This series of NIEs may be taken as the authoritative view of the entire U.S. intelligence community on Soviet nuclear capabilities and would have been the government source most likely to inform and influence policymakers. According to a review of declassified Cold War NIEs, "the appellation 'national' indicated that the intelligence analysis in question was produced with the concurrence–or at least the informed dissent–of the government organizations that made up the US intelligence community."[62]

Next, I construct a "true" baseline of data on Soviet nuclear forces from the same time period. Past efforts at assessing the accuracy of U.S. intelligence estimates of Soviet nuclear capabilities have often suffered from a lack of authoritative data on the Soviet nuclear arsenal and been forced to rely on "independent" assessments, such as those produced by the Natural Resources Defense Council (NRDC).[63] The NIE and NRDC both produced estimates of the total number of Soviet strategic delivery vehicles. Here, I construct a two measures of U.S. intelligence errors using the NRDC data as a baseline, one reporting the error as the percent of the actual number of delivery vehicles reported by the NRDC and the second reporting the error as the difference in the number of delivery vehicles between the NRDC and NIE estimates. The results are illustrated in Figures 3 and 4.

Each figure plots both the annual estimation and the average estimation error across

61. Unfortunately, it does not appear possible to construct similar datasets for other measures of nuclear balance, such as warhead stockpile, strategic warhead stockpile, or total megatonnage. The NIEs produced during this period either did not provide estimates of these characteristics or, if the estimates were provided, they have been excised or redacted in the declassified versions.

62. Steury, *Intentions and Capabilities: Estimates on Soviet Strategic Forces, 1950-1983*, xi.

63. This is the same data I analyze in the earlier section examining the different ways of conceptualizing and measuring the strategic nuclear balance. Norris and Cochran 1997.

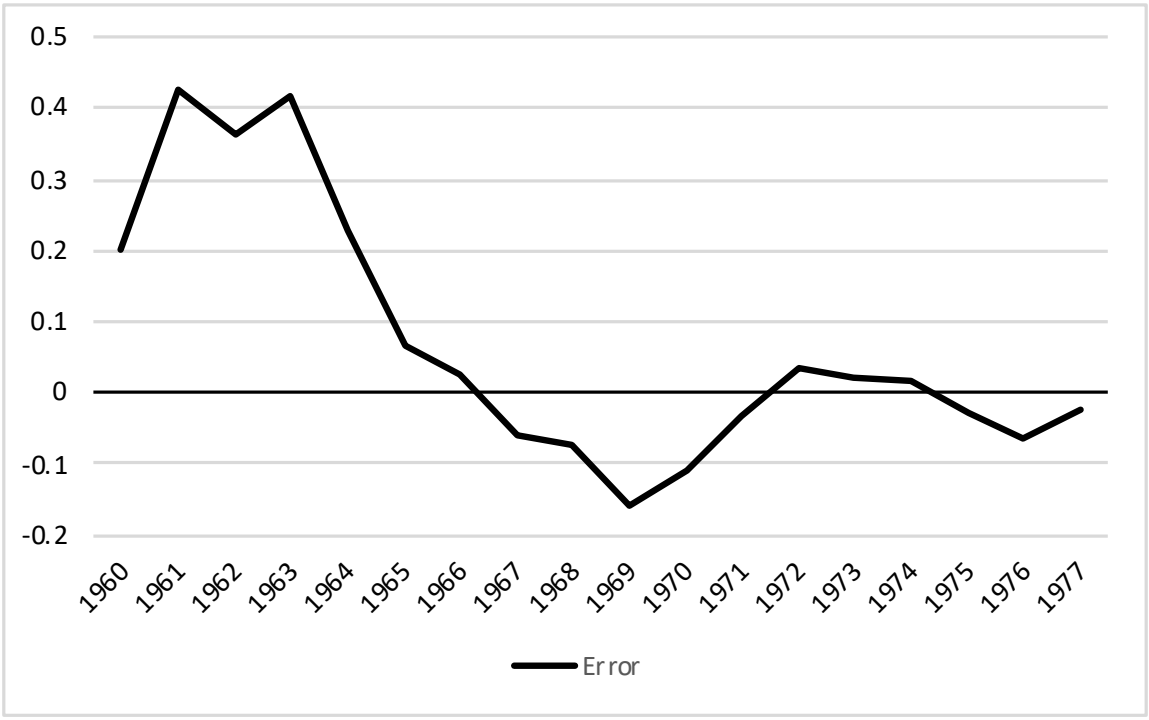

Figure 3: Delivery vehicles estimate error as percent of actual delivery vehicles, 1960-1977

all years in the dataset. We see that U.S. intelligence estimates of Soviet strategic delivery vehicles exhibited non-trivial errors throughout the 1960s and 1970s, averaging an error of roughly 15 percent (without regard to sign) and 100 bombs (without regard to direction). Though these errors were significantly reduced in the early 1970s, they at times reached above 40 percent and 250 delivery vehicles.

Unfortunately, however, the NRDC data makes for an imperfect baseline for two reasons. First, little if any of the data comes directly from Soviet sources; rather, much of it is built from U.S. government and non-governmental sources. Thus, the numbers represent merely a best estimate rather than an actual baseline. Second, and most significant, both the baseline and the estimates rely on some of the same sources.[64] For instance, the baseline relies, in part, on the declassified National Intelligence Estimates.[65] Due to this

64. Norris and Cochran, *US-USSR/Russian Strategic Offensive Nuclear Forces, 1945-1996*, 10-11.

65. See, for example, Daryl G Press, *Calculating Credibility: How Leaders Assess Military Threats* (Cornell University Press, 2005), 91-94. To construct his "true baseline" of Soviet nuclear forces, Press relies on data

Figure 4: Delivery vehicles estimate error as difference in number of estimated and actual delivery vehicles, 1960-1977

duplication of sources, the "estimates" and the "baseline" will naturally exhibit a higher than "natural" degree of correlation.[66] To avoid these problems, I rely on data published by Podvig which draws from the personal papers of a senior Soviet official.[67] These papers include data on the number of Soviet ICBMs deployed in each year from 1970 to 1990. I then construct two measurements of the estimation error of the intelligence assessments: an absolute error, measured by the difference in the number of warheads; and a relative error, measured in the percent by which the intelligence assessment over- or underestimated the size of the actual ICBM force.

published in Podvig 2001. The relevant data in Podvig comes from Norris and Cochran 1997, which relies, in part, on U.S. government estimates.

66. I thank J. Robert Logan, V for emphasizing this point. Some analysts apparently fail to realize the extent to which their own "true baselines" are built from the same data sources as U.S. intelligence estimates, though they do sometimes supplement their analyses with additional sources.

67. Podvig's data comes from the personal papers of Vitalii Leonidovich Kataev, who served as senior adviser to the secretary for the Defense Industry of the Central Committee of the Communist Party in the Soviet Union from 1974 to 1990.

To address the second weakness discussed above, I also attempt to determine the degree to which data from independent estimates, such as those compiled by the NRDC, reflect the actual size and composition of nuclear arsenals. I do this by comparing NRDC estimates for Soviet ICBMs to the data published in Podvig. As above, I again calculate absolute and relative errors of these estimates. I plot the absolute (Figure 5) and relative (Figure 6) estimation errors of both the NIE and NRDC estimates.

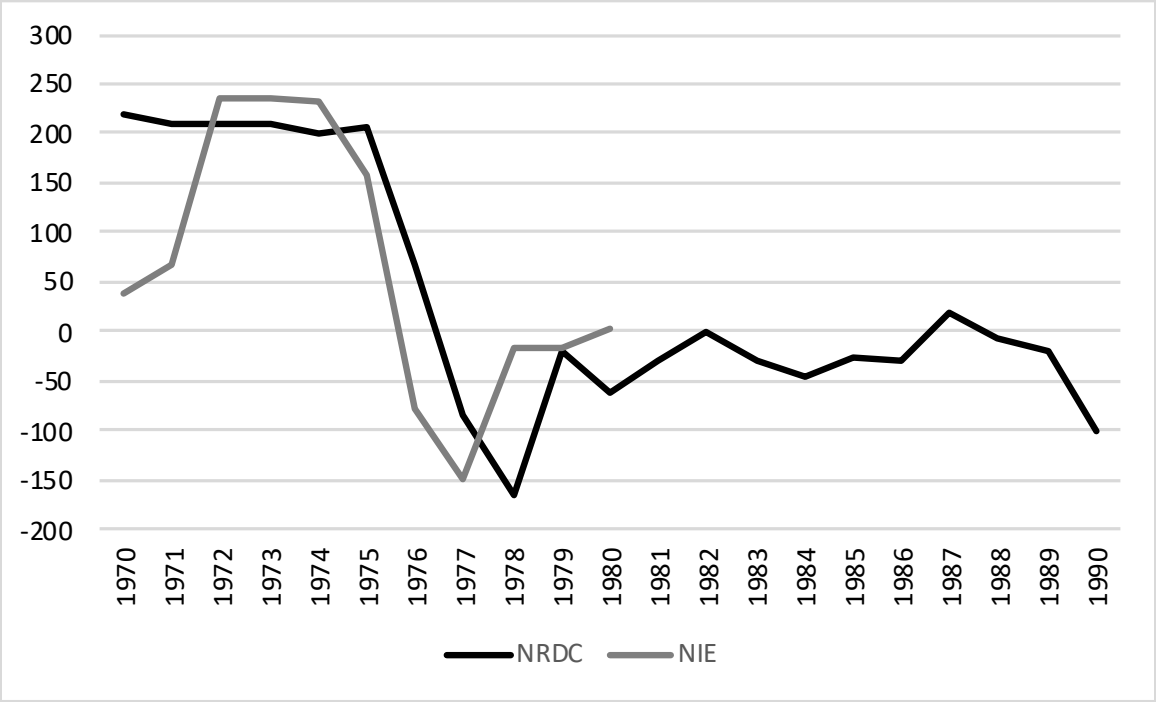

Figure 5: ICBM estimate error as difference in number of estimated and actual ICBMs, 1970-1990

The figures motivate several conclusions. First, over the period for which comparisons are possible, U.S. intelligence estimates of the number of Soviet ICBMs exhibited significant errors, sometimes overestimating the force by as much as 20 percent and underestimating it by as much as 10 percent. Without regard to sign, the average U.S. estimate of Soviet ICBMs was off by more than 10 percent. Though these errors appear smaller than the differences between the NIE and NRDC estimates, they are still significant for

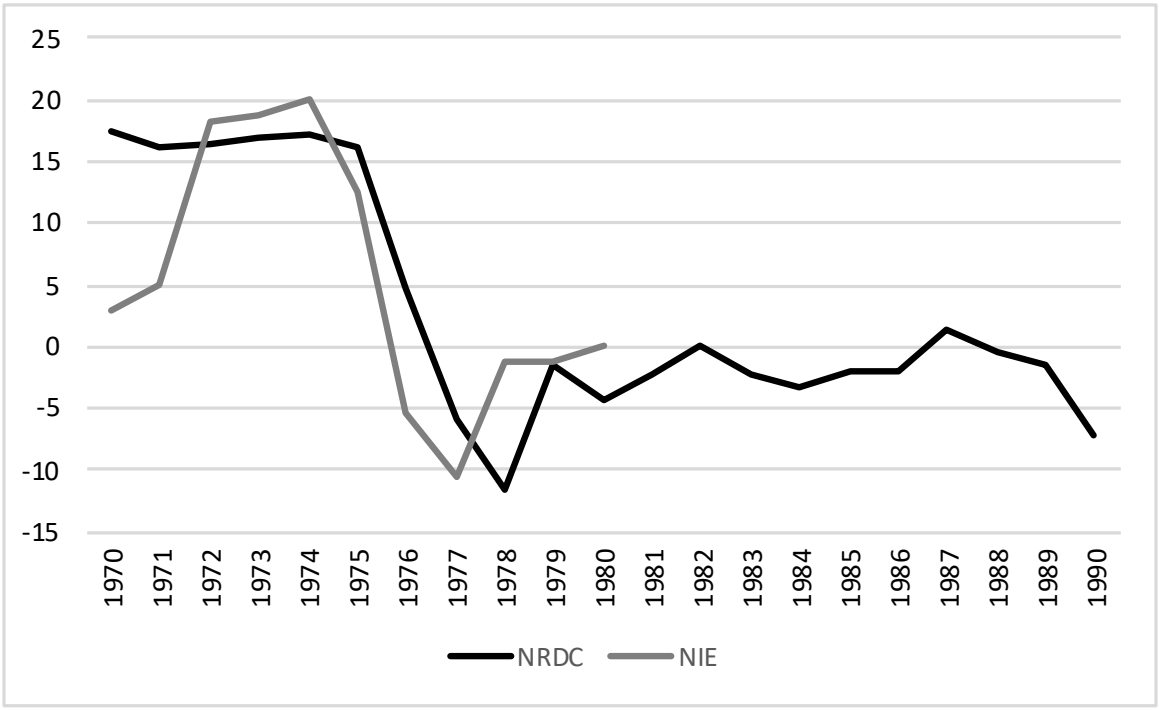

Figure 6: ICBM estimate error as percent of actual ICBMs, 1970-1990

the "easy" case of ICBM estimates. Second, data from independent sources largely track the data in the National Intelligence Estimates and exhibit similar errors in estimation. Finally, by the 1980s, estimation errors had been significantly reduced to under 50 ICBMs, or about 5 percent of the actual size of the Soviet arsenal.

Though existing data only permits a direct assessment of U.S. intelligence on Soviet ICBMs from 1970 to 1977, there is reason to believe that U.S. assessments about other aspects of the Soviet nuclear arsenal, including warhead counts, were subject to even much greater inaccuracies. First, and most importantly, the error estimates reported above may be considered minimum expected error rates compared to errors in other types of estimates. Within the domain of assessing an adversary's nuclear forces, counting silo-based delivery vehicles constitutes one of the easiest tasks. It is much easier to count delivery vehicles than warheads and easier to evaluate quantitative than qualitative aspects of nu-

clear arsenals.[68] Warheads are smaller and more easily concealed and transported than their associated launchers. Indeed, the relative ease of counting launchers, in comparison to warheads, is one of the primary reasons arms control agreements restricted the former. Second, as discussed earlier, U.S. intelligence assessments themselves directly acknowledged the difficulty in counting warheads and at times either declined to provide such an estimate or provided very wide ranges.[69] Analysts seemed reluctant to provide an estimate of the stockpile size given the inherent difficulties.

Finally, evidence suggests Western estimates of the Soviet warhead stockpile have exhibited more significant inaccuracies. For example, in 1993, then-head of the Russian Ministry of Atomic Energy stated that the Soviet Union's nuclear warhead stockpile had peaked in 1986 at about 45,000 warheads, as much as 17,000 more bombs or more than 60 percent higher than had been estimated at the time.[70] U.S. estimates of the technical features of Soviet nuclear forces were similarly flawed. U.S. estimates from the mid- and late-Cold War incorrectly placed the accuracy of key Soviet ICBMs at almost half (250 meters) of their true accuracy (400 meters or more).[71] Other reports overestimated the hardness of Soviet ICBM silos by as much as a factor of ten (U.S. estimated hardness ratings of 15,000 to 25,000 psi compared to an actual maximum Soviet hardness rating of 1500 psi).[72] These same challenges will confront independent analyses of nuclear arsenals, especially to the extent that they rely on U.S. government documents. These estimation errors suggest that U.S. decisionmakers frequently did not have accurate information about the Soviet nuclear

68. Steve Fetter, "A Comprehensive Transparency Regime for Warheads and Fissile Materials," 29, no. 1 (1999): 3–7.

69. See, for example, National Intelligence Estimate 11-2-59 1959, 59; and National Intelligence Estimate 11-2A-63 1963, 37. See Document A1 and Document A2 in the Appendix for active citations.

70. William J Broad, "Russian Says Soviet Atom Arsenal Was Larger Than West Estimated," *The New York Times*, 1993, 1; Houston T Hawkins, *Rethinking the Unthinkable*, technical report (Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2014), 10.

71. Podvig, "The Window of Vulnerability That Wasn't: Soviet Military Buildup in the 1970s?A Research Note," 125-129.

72. Ibid., 129-132.

arsenal. As detailed information on the size and configuration of Soviet nuclear forces have yet to be released, it is impossible to conduct a similar analysis of Soviet perceptions. However, there is little reason to believe that Soviet intelligence products would have been significantly more accurate than those of the United States.

It's not clear how to incorporate potential intelligence errors into a replication of the statistical analysis. Just examining the U.S. data, error rates differed by year, the type of system being assessed (warheads, ICBM, SLBM, heavy bomber), and the characteristics being estimated (number of systems, performance characteristics). Intelligence error rates likely include both a systematic component (related to political influence, bureaucratic pathologies, standard operating procedures, and adversary countermeasures) and a natural random element. However, the evidence presented here indicates that intelligence estimates frequently varied from both the actual capabilities of the adversary and from the independent estimates used by scholars.

A failure to account for incomplete information may systematically bias analyses of nuclear superiority. When the differences between nuclear arsenals are clearly large, assumptions of complete information may not bias the analysis. These systematic biases will be most significant when the nuclear capabilities of the two states are more similar and when the states possess relatively smaller arsenals. At least two sets of cases from the large-N studies stand out as potentially vulnerable to bias from assuming complete information. First are the series of interactions between the United States and the Soviet Union from the late 1960s to the late 1970s when the nuclear arsenals of those two states approached parity and the Soviet Union began to obtain superiority across some nuclear measures.[73] The errors during this period may be large enough to misrepresent

73. Kroenig's dataset includes crises over the Cienfuegos submarine base (1970), the Yom Kippur War (1973), the war in Angola (1975), and the invasion of Afghanistan (1979), while the Sechser and Fuhrmann database include the Cienfuegos base incident as an instance of coercion by the U.S. directed against the Soviet Union.

even which side possessed superiority. For example, in 1972, just before the crisis over the Yom Kippur War, the U.S. National Intelligence Estimate assessed that the Soviet Union possessed 2,238 strategic nuclear delivery vehicles, in comparison to the U.S.'s 2,167. It is still unclear exactly how many strategic delivery vehicles the Soviets deployed. However, retrospective estimates by the NRDC estimate only 2,164 delivery vehicles, less than the U.S. deployed. Assuming that estimates of total strategic delivery vehicles suffered from (merely) the same error rates as estimates of ICBMs, the Soviets may have only deployed 1,895 delivery vehicles, nearly 300 fewer than the U.S.

Second are the several crises involving India and Pakistan from 1990 to 2001.[74] In each of these observations, estimates of the size of the two nuclear arsenals are almost identical, differing by, at most, only two warheads in some years.[75] It is unlikely that decisionmakers at the time were aware of, much less sensitive to, differences between the size of the two arsenals. As one scholar has rightly observed, "Indeed, it is not even clear that Indian and Pakistani civilian leaders knew the size of their own nuclear arsenals during this period."[76]

Towards Superior Nuclear Superiority Research

Here I propose two means by which to improve the quantitative literature on nuclear superiority and offer some provisional evidence that, at least for the U.S. and the Soviet Union, nuclear superiority did not matter. First, to incorporate state perceptions, research on nuclear superiority should use data from real-time intelligence assessments and em-

74. The Kroenig dataset includes four crises involving India and Pakistan (1990, 1998, 1999, and 2001), while the Sechser and Fuhrmann date include the 2001 attempt by India to coerce Pakistan into controlling terrorist organizations.

75. Data taken from Hans M Kristensen and Robert S Norris, "Nuclear Notebook: Nuclear Arsenals of the World," *Bulletin of the Atomic Scientists*, 2018, The Bulletin data only identifies India and Pakistan as initially possessing nuclear arsenals in 1998.

76. Francis J Gavin, "What We Talk About When We Talk About Nuclear Weapons: A Review Essay," in *H–Diplo/International Security Studies Forum*, vol. 2 (2014), 11–36.

ploy them within directed dyads. Second, scholars must move beyond simple warhead counts and develop measures of nuclear capabilities which better reflect both the actual strategic balance and, most importantly, the beliefs of state leaders.

First, to address the issue of perceptions, the literature should employ data reflecting state perceptions of the nuclear balance within directed dyads. As discussed, up until now the quantitative work has relied on third party estimates and assumed that each state in a dyad shares the same (accurate) information about the relative nuclear balance. Employing directed dyads would also have the added benefits of both increasing the sample size and better modeling real-world dynamics. Specifically, perceptions of the nuclear balance may generate asymmetric effects, depending on what side of the balance a state believes it is. That is, a state's response to a perceived condition of nuclear superiority may not simply be the mirror of its response to a perceived condition of nuclear inferiority.[77] Employing directed dyads helps to capture these different effects.

When employing directed dyads, research should not assume symmetric information but, rather, attempt to capture state perceptions at the time. As discussed, research could reasonably assume that states possess perfect information about their own arsenals and imperfect information about the arsenals of other states. When reliable information is not available about the nuclear arsenals of a state, scholars should employ sensitivity tests to ensure that their findings are robust to misperceptions generated by intelligence errors.

Second, if the literature will continue to apply quantitative methods to questions of nuclear superiority, in order to address the problems associated with using total warhead counts as a proxy for nuclear capabilities, the literature must develop more robust measures of nuclear capabilities. Significantly, this cannot be done based on what researchers believe is the most important measure of a state's nuclear capabilities or what kinds ca-

77. For some evidence of this, see Trachtenberg, "The Influence of Nuclear Weapons in the Cuban Missile Crisis," 147-161

pabilities are easiest to measure but, rather, must reflect the beliefs and perceptions of decisionmakers at the time. In short, nuclear superiority will generate political effects only when and how leaders believe it does.

To illustrate these points, I conduct an analysis of perceptions of the nuclear balance in the U.S.-Soviet dyad. In order to investigate the potentially asymmetric effects of nuclear superiority, I employ a directed-dyad approach, examining first the perceptions of U.S. decisionmakers and then of their Soviet counterparts. In order to best capture state perceptions of nuclear superiority, I eschew static measures of nuclear arsenals and instead attempt to capture decisionmakers' actual perceptions of the nuclear balance and their expectations of the results of a nuclear exchange.

Instead of aggregate stockpiles or other static technical indices, I propose states' expected relative fatality counts as a better real-time measure of assessments of the nuclear balance. Arguments about the significance of nuclear superiority turn on state assumptions about future nuclear exchanges and the relative damage that each side would receive. These damage assessments can then work to either bolster or undermine state resolve which, in turn, impacts interstate signaling, bargaining, and crisis outcomes. In this way, the best measure of states' perceptions of the nuclear balance is not raw counts of the weapons themselves but, rather, the anticipated effects of using them.[78] Historically, expected damage from suffering a nuclear strike was measured in fatalities and damage to industrial production. However, these kinds of fatality and damage estimates are not easily constructed solely from warhead counts. Fatality estimates can vary widely depending on assumptions about, inter alia, force postures, exchange details, weather patterns, pop-

78. Even using raw fatality estimates may oversimplify and distort state perceptions. For instance, would states care more about the number of fatalities inflicted or, instead, what proportion of the state's population those fatalities represent? During the Cold War, U.S. officials often defined "unacceptable damage" to the Soviet Union in terms of percentage of population and economic activity lost. U.S. nuclear exchange models often highlighted the "recovery time" of a society as a key variable impacting the relative strategic balance.

ulation density, missile accuracy, and effectiveness of civil defense measures.[79] In fact, one analyst argued that, even at roughly symmetrical strategic nuclear forces, these climatic, socioeconomic, and policy factors provided the Soviet Union an enduring advantage in its strategic competition with the United States.[80] These kinds of data may not be readily available for all nuclear-armed states. Many states may not have even produced these kinds of damage assessments or, if they did, the assessments may not be declassified or otherwise available to researchers. Indeed, in the directed-dyad analysis below, systematic fatality assessments are only available from the U.S. perspective. Soviet perspectives of the nuclear balance must be deduced through more qualitative work on leaders' perceptions. However, the challenges of data collection are not sufficient reason for resorting to poor conceptual measures. This is especially true when such data is readily available to investigate some of the most substantively important cases; in this case, the U.S.-Soviet Cold War dyad.

Forecasting Armageddon: U.S. Nuclear Exchange Fatality Expectations

Throughout the Cold War, U.S. government agencies produced dozens of assessments estimating the damage which would be suffered by the United States and the Soviet Union in the event of a nuclear exchange between the two. Many of these estimates were prepared by the Net Evaluation Subcommittee of the National Security Council (NESC), which was tasked with estimating the damage which would occur under various hypothetical nuclear attacks.[81] Similar estimates were produced by other government organizations, such as the Defense Department's Weapons Systems Evaluation Group (WSEG), the Arms Con-

79. Desmond Ball, *Can Nuclear War Be Controlled?* (International Institute for Strategic Studies London, 1981), 27.

80. Francis P Hoeber, "How Little is Enough?," *International Security* 3, no. 3 (1978): 56.

81. National Security Council Directive 5511. 1955. Net Evaluation Subcommittee. Available at https://history.state.gov/historicaldocuments/frus1950-55Intel/d207.

trol and Disarmament Agency (ACDA), the CIA's Office of Research and Estimates (ORE), the Air Force, and the Joint Chiefs of Staff.[82] These fatality estimates naturally incorporate many of the other technological, strategic, meteorological, and demographic variables which would otherwise be difficult to incorporate. More significantly, they provide a clear measure of the variable which should matter in theories of nuclear superiority: expected damage. These assessments were more than just bureaucratic exercises. They reached and influenced the highest levels of the U.S. government. The results were regularly briefed to the president and senior officials including the Secretary of Defense and National Security Advisor. Personal memos, meeting minutes, and private diaries are peppered with references to these reports' gruesome conclusions.[83]

Based on an analysis of declassified U.S. documents and adding to recently published by work by Milne, I construct a dataset of U.S. expected fatalities.[84] Table 3 shows the number of fatalities in a large-scale nuclear exchange between the United States and the Soviet Union as estimated by U.S. government assessments.[85] For each assessment, I record the year in which it was published and the estimated range of U.S. and Soviet fatalities. For assessments which include estimates of both U.S. and Soviet fatalities, I then record at the low, mean, and high estimates, which of the two states was predicted to "win" a large-scale nuclear exchange by suffering the fewest casualties.[86]

82. For an excellent review of the NESC and other government reports estimating the damage of prospective nuclear exchanges, including the impact of such reports on the thinking and decisions of policymakers, see Caroline Reilly Milne, "Hope Springs Eternal: Perceptions of Mutual Vulnerability between Nuclear Rivals," 2017, For a discussion of the NESC and its reports, see William Burr, "Studies by Once Top Secret Government Entity Portrayed Terrible Costs of Nuclear War," *The National Security Archive Electronic Briefing Book No. 480*, 2014,

83. Milne, "Hope Springs Eternal: Perceptions of Mutual Vulnerability between Nuclear Rivals," chapters 2-5.

84. For illustrative excerpts from some of these reports, especially those not captured in the Milne work, see Document B1 to Document B5 in the Appendix.

85. Data taken from Milne, "Hope Springs Eternal: Perceptions of Mutual Vulnerability between Nuclear Rivals," with some supplementation from primary sources.

86. Reporting simply the ranges of the fatality estimates necessarily simplifies the predictions of fatalities. When estimates provide a range of possible fatalities, I produce a "mean" estimate by simply taking the mean of the highest and lowest estimates. For estimates which only produce one estimate of fatalities (for

Assessment	Year	US fatalities	Soviet fatalities	Low Estimates	Mean Estimates	High Estimates
Harmon Report	*1949*	–	*2.7-6.7*	–	–	–
Air Force estimate	*1950*	*3-6*	–	–	–	–
CIA (ORE 91-49)	*1950*	*10*	–	–	–	–
NSC-114/2	*1951*	*10*	–	–	–	–
RAND study	*1952*	*22-35*	*22-25*	*EQUAL*	*USSR*	*USSR*
NESC	*1953*	*4.5-6.3*	–	–	–	–
NESC	*1954*	*3.1-9.6*	–	–	–	–
WSEG R-12	*1955*	–	*60*	–	–	–
NESC	*1955*	*114*	–	–	–	–
NESC	*1956*	*71*	–	–	–	–
NESC	*1957*	*46-95*	*81*	*USA*	*USA*	*USSR*
NESC	*1958*	*50*	*114*	–	*USA*	–
NESC	*1959*	*62-82*	–	–	–	–
NESC	*1960*	*61*	*99*	–	*USA*	–
WSEG 50	*1961*	*60-90*	–	–	–	–
NESC	*1961*	*68-83*	*67*	*USSR*	*USSR*	*USSR*
JCS	*1961*	*68-83*	*108*	*USA*	*USA*	*USA*
ISA-NSC	*1961*	*0-140*	*0.5-1*	*USA*	*USSR*	*USSR*
DOD (ISA)	*1961*	*12-177*	*3-68*	*USSR*	*USSR*	*USSR*
NESC	*1962*	*47-93*	*69-93*	*USA*	*USA*	*EQUAL*
DPM	*1962*	*30-95*	*17-83*	*USSR*	*USSR*	*USSR*
DOD Report	*1962*	*4-85*	*10-95*	*USA*	*USA*	*USA*
NESC	*1963*	*63-134*	*136-143*	*USA*	*USA*	*USA*
DOD Report	*1976*	*5-125*	*40-75*	*USA*	*USSR*	*USSR*
ACDA	*1978*	*69-131*	*23-94*	*USSR*	*USSR*	*USSR*
DOD Report	*1981*	*20-165*	*23-100*	*USA*	*USSR*	*USSR*

Table 3: U.S. government estimates of U.S. and Soviet fatalities (millions) in a nuclear exchange

Next, I assess whether the outcomes of nuclear crises between the United States and the Soviet Union were impacted by the nuclear balance or estimates of relative fatalities. For each assessment with estimates of both U.S. and Soviet fatalities, I calculated the value of the expected number of U.S. fatalities minus the expected number of Soviet fatalities.[87] I then plot each estimate on Figure 7. For each assessment, I also indicate whether a crisis with the Soviet Union occurred in the same year of or the year following the assessment. For assessments that were proximate to nuclear crises, I then mark whether the U.S. succeeded (solid diamonds) or failed (hollow diamonds) to achieve its goals in the crisis. On the same figure, I also plot the difference between the number of warheads in the U.S. arsenal and the Soviet arsenal.

example, Soviet fatalities in the 1957 NESC assessment), I use the same number for low, mean, and high estimates.

87. Here I use the mean values provided by the estimates.

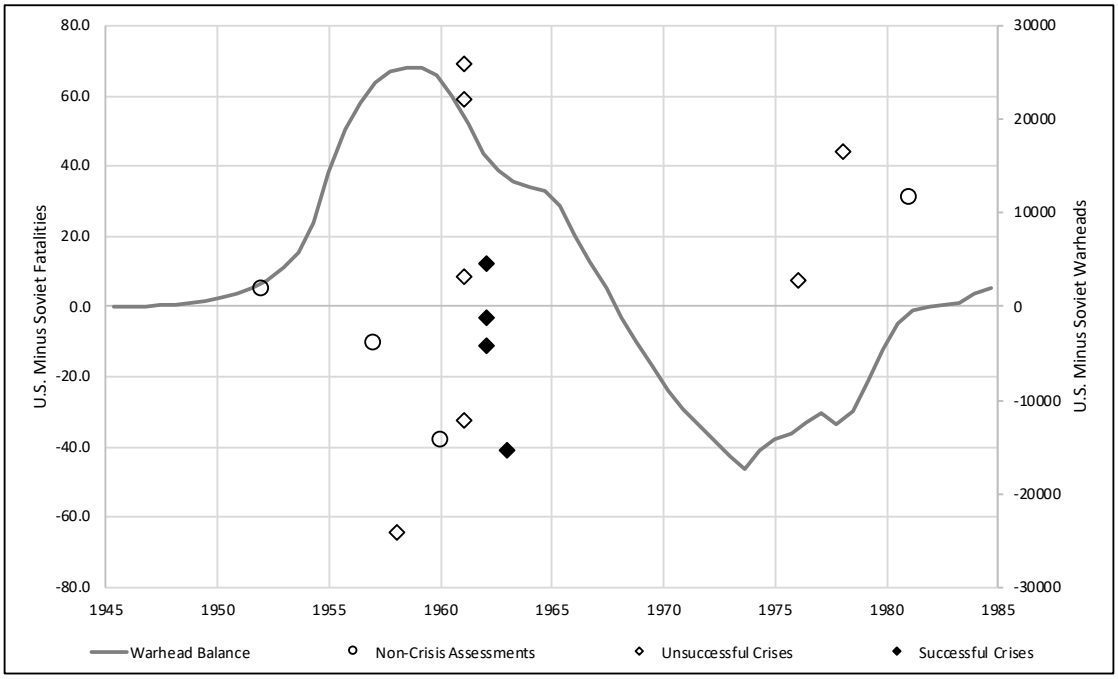

Figure 7: U.S.-Soviet crisis outcomes and nuclear exchange fatality estimates

Three findings emerge from the data. First, there appears to be a little relationship between the relative size of U.S. and Soviet nuclear stockpiles and the relative balance in expected fatalities in each state following a large-scale nuclear exchange. The late-1950s to mid-1960s was a period of considerable stockpile superiority for the United States, during which it possessed at least 20,000 more nuclear weapons than the Soviet Union. However, the estimates produced during this period show little agreement on which side would suffer more fatalities, as several estimates even predicted that, despite its significant stockpile superiority, the United States would suffer more than the Soviet Union.

Second, there appears to be a weak relationship between the expected outcome of a nuclear exchange and the outcomes of interstate bargaining. The U.S. succeeded in two of the three crises in which assessments generally suggested it would suffer fewer fatalities than the Soviet Union. By comparison, it succeeded in none of the three crises in which proximate assessments predicted it would suffer more fatalities than the Soviet Union.

This data might be interpreted as weak evidence in favor of the underlying assumption of nuclear superiority: that states which expect to suffer less damage will be more likely to prevail in interstate bargaining.

Third, and most significantly however, there appears to be little if any relationship between, on the one hand, the material nuclear balance between the Soviet Union and the United States and, on the other hand, expectations about the relative damage each state would suffer in a nuclear exchange. Here, it may be instructive to focus on the damage assessments conducted just prior to the Cuban Missile Crisis. In the two years prior to the Crisis, offices of the U.S. government created at least six separate assessments of the outcome of a nuclear exchange between the United States and the Soviet Union. In 1960, when the United States possessed roughly 25,000 more warheads than the Soviet Union, the Net Evaluation estimated that, in such an exchange, the United States would suffer 61 million fatalities compared to 99 million Soviet Union fatalities. The next year, five assessments were produced in the U.S. government, one by the NESC, two by the Joint Chiefs of Staff, and two by the office of the Secretary of Defense for International Security Affairs (ISA). The relative damage estimates produced by these assessments varied widely. The most optimistic of these assessments, produced by the JCS, estimated that the U.S. would suffer 32 million fewer fatalities than the Soviet Union. By contrast, the two ISA assessments produced that year estimated the United States would suffer 60 to 70 million more fatalities than the Soviet Union. Estimates varied significantly even within individual assessments, with one ISA assessment predicting that U.S. fatalities could range anywhere from "negligible" to 140 million.[88] These estimates, nearly all issued within a year of one another, all analyzed an essentially unchanging material balance. And, yet, they predicted dramatically different outcomes.

88. Here, I assume that "negligible" means close to or at zero. Other assessments offered similarly low predictions. A separate DOD (ISA) assessment conducted in 1961 estimated U.S. fatalities might be as "low" as 12 million, while a DOD report conducted the following year estimated there might be "only" 4 million.

Another way of appreciating the variability and uncertainty in U.S. expectations about fatalities in a nuclear exchange with the Soviet Union is to directly consider the variation in official U.S. estimates. Figure 8 plots, for each of the 15 nuclear exchange reports that produced estimates of both U.S. and Soviet fatalities, the median expected number of U.S. fatalities minus the median number of expected Soviet fatalities.[89] Error bars extend to the highest expected differences favorable to the United States (lower on the figure) and the Soviet Union (higher on the figure). As the figure illustrates, there was tremendous variation in the expected relative fatalities at both the level of individual assessments and at the level of the year in which assessments were produced. In almost every case, it was not even clear which side would suffer more fatalities.

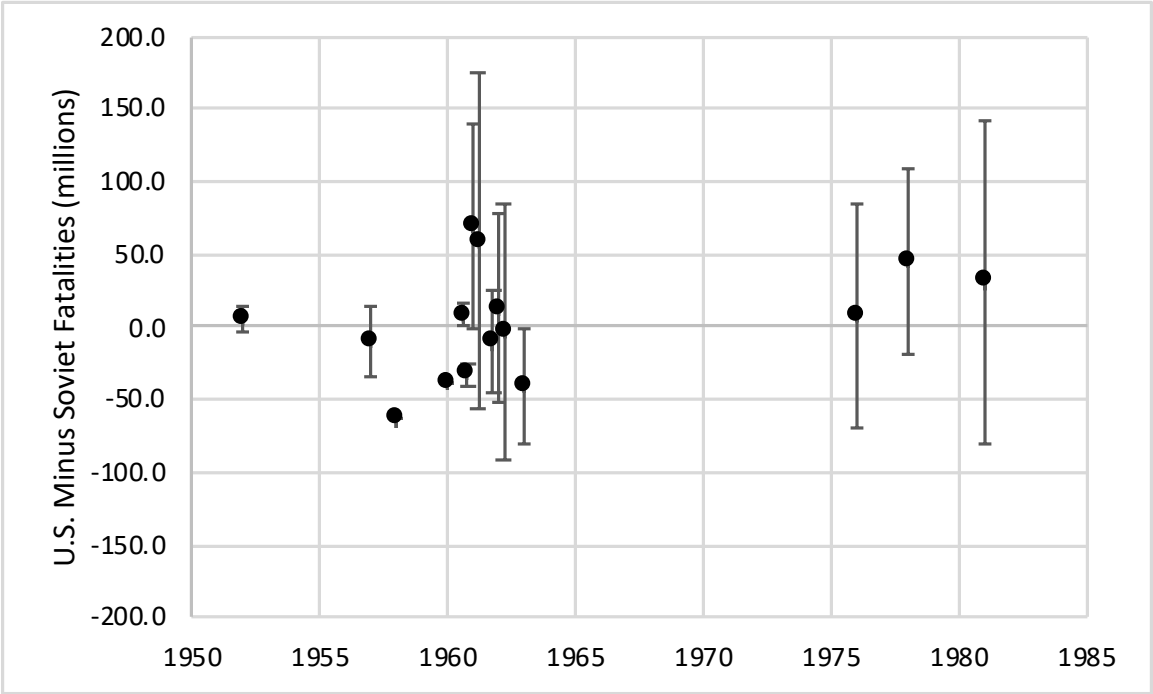

Figure 8: U.S. minus Soviet nuclear exchange fatality estimates

This variation is driven by a range of assumptions about how such a nuclear exchange would occur: which side would strike first, the size and configuration of each side's nu-

89. To facilitate discrimination, data points for reports produced in the same year are slightly offset horizontally.

clear forces, targeting approaches, the degree to which nuclear forces are hardened, the effectiveness of civil defense procedures, the accuracy of early warning capabilities, proximity of population centers, and (literally) which direction the winds will blow. All of these are subject to uncertainty and most are hidden by the adversary. These gaps are filled by the assumptions, perceptions, and beliefs of states. The entire process is then compounded by parallel processes taking place on the other side.

Ideological Blinders: The Political Obstacles to Soviet Nuclear Assessments

Analysis within the Soviet Union of the relative nuclear balance was similarly shaped by the perceptions and beliefs of leaders, first as political ideology suppressed efforts at conducting "objective" studies of the results of nuclear war and, later as state beliefs about the fragility of command and control structures prevented the Soviets from enjoying the fruits of their quantitative superiority. As shown earlier, the U.S. intelligence and defense communities had developed various (and at times sophisticated) models for measuring and comparing nuclear capabilities. By comparison, independent assessments of the strategic nuclear balance did not impact Soviet decisionmaking until decades into the Cold War. Until the late 1960s and early 1970s, the Soviet Union possessed practically no means of systematically measuring the relative nuclear balance.[90] Throughout this time period, nuclear exchange models were often discounted because to accept their conclusions of a Soviet defeat would have been to undermine the prevailing ideology at the time.[91] According to Marshal Sergei F. Akhromeev, who rose through the General Staff

90. Levy, *Soviet Strategic Nuclear Measures of Effectiveness*, 1-3; Milne, "Hope Springs Eternal: Perceptions of Mutual Vulnerability between Nuclear Rivals."

91. Milne, "Hope Springs Eternal: Perceptions of Mutual Vulnerability between Nuclear Rivals," 122-161; John G Hines, M Mishulovich Ellis, and F Shull John, *Soviet Intentions 1965–1985, Volume I: An Analytical*

and eventually served as Gorbachev's national security advisor, "modeling and analysis did contribute" to operational and strategic planning but that this "was more true in the mid-1970s and later [and that m]any other factors, however, went into such decisions."[92] And, according to Milne, whose recent work examined Soviet perceptions of the strategic balance, "Credible calculations of the effects associated with large-scale nuclear exchanges did not make an impression on the decision-making process until the early 1980s; prior to this point models of nuclear war were either nonexistent or misrepresented to make the costs seem more palatable."[93]

When the Soviets did develop assessments of the strategic nuclear balance, their assessments incorporated more than just warhead counts. According to one Senior Soviet official intimately involved with nuclear weapons issues, "Nuclear power [iadernaia moshch'] in our assessments, is a function of yield, nuclear weapons, and accuracy."[94] A RAND report produced at the end of the Cold War identified four distinct measures of effectiveness (MOEs) for assessing and comparing nuclear capabilities.[95] The first, Quantitative Correlation of Nuclear Forces, measured the relative numbers of similar types of weapons in states' arsenals. The second, Equivalent TNT Correlation of Nuclear Forces adjusted arsenal size to account for the yield of the weapons. A third, the Anureyev Correlation of Nuclear Forces, incorporated technical features of the arsenals as well as assumptions about how they would be used. Finally, the Destruction Potential Correlation of Nuclear Forces which sought to explicitly model the effects of a nuclear exchange with the United States. It was this final method which analysts believed to have been employed by So-

Comparison of US-Soviet Assessments During the Cold War, 1995, 25-27.

92. Interview with Sergei F. Akhromeev in John G Hines, M Mishulovich Ellis, and F Shull John, *Soviet Intentions 1965–1985, Volume II: Soviet Post-Cold War Testimonial Evidence*, 1995, 4. See Document C1 in the Appendix for active citation.

93. Milne, "Hope Springs Eternal: Perceptions of Mutual Vulnerability between Nuclear Rivals," 122.

94. Interview with Vitalii Leonidovich Kataev in Hines, Ellis, and John, *Soviet Intentions 1965–1985, Volume II: Soviet Post-Cold War Testimonial Evidence*, 100. See Document C2 in the Appendix for active citation.

95. Levy, *Soviet Strategic Nuclear Measures of Effectiveness*, v-vii. See Document C3 in the Appendix for active citation.

viet assessments. Colonel-General Andrian A. Danilevich, who led the Soviet group in charge of strategic and operational planning from 1977 to 1986, recognized the poverty of bean-counting arsenals, saying: "[A]nalysis of quantity alone provides only half of the analytical picture. Because of qualitative deficiencies, one side could have a tenfold quantitative advantage and still be behind."[96]

Although there is evidence that Soviet officials eventually conducted assessments of the results of a nuclear exchange with the United States, there are few details on the findings of those assessments. At least two strands of evidence provide some insight into Soviet assessments of the strategic nuclear balance and how those assessments impacted decisionmaking.

First, even when Soviet assessments of the relative nuclear balance concluded that the Soviet Union enjoyed numerical superiority across various aspects of the arsenals, they still did not believe they enjoyed overall superiority because of the weakness of their command and control arrangements. The Soviets actively sought strategic superiority throughout much of the Cold War.[97] The Soviet Union eventually did perceive superiority in several areas, including number of launchers, silo protection, yield of warheads, and range and power of missiles.[98] However, this numerical superiority did not translate into any meaningful strategic superiority from the Soviet perspective. A report from the Soviet General Staff concluded that command and control infrastructure was so fragile that "after sustaining an all-out nuclear strike the Soviets would be able to launch only 2% of their missiles."[99] As late as 1982, Soviet officials were still bemoaning their inability to validate the survivability of either nuclear forces or their related command and control

96. Interview with Andrian A. Danilevich in Hines, Ellis, and John, *Soviet Intentions 1965–1985, Volume II: Soviet Post-Cold War Testimonial Evidence*, 22. See Document C4 in the Appendix for active citation.

97. Hines, Ellis, and John, *Soviet Intentions 1965–1985, Volume I: An Analytical Comparison of US-Soviet Assessments During the Cold War*, 2-13.

98. Interview with A. S. Kalashnikov in Hines, Ellis, and John, *Soviet Intentions 1965–1985, Volume II: Soviet Post-Cold War Testimonial Evidence*, 89-90. See Document C5 in the Appendix for active citation.

99. Interview with A. S. Kalashnikov in ibid., 90.

systems. A March 1982 report from the head of the Central Committee's Defense Industry Department to Chief of the General Staff and other senior defense officials argued that, despite the importance of verifying the survivability of the nuclear forces, a combination of bureaucratic hassles, financing shortfalls, and technical challenges meant that, "Notwithstanding the importance of carrying out the verification of the stability of various forms of military equipment, until now the general five-year [verification] plan has not worked out."[100] A separate report high-level report issued in February that year by the Central Committee's Defense Industry Department observed the failure of the Soviets to evaluate and ensure their nuclear command and control systems, noting that "a comprehensive comparative assessment of the necessity and sufficiency of the work ongoing in the country with the goal of ensuring control of nuclear forces in war (especially in conditions of nuclear strike) has not yet been conducted. This does not make it possible to identify weak links in the systems being created or to eliminate shortcomings."[101] Without the ability to actually launch the weapons, their greater numbers were useless. According to one former senior official, "Soviet superiority in the number of launchers did not give them any real advantage. This numerical superiority reflected a mechanistic, wasteful approach to force building."[102]

Second, even when predicting that the United States would suffer more in a nuclear exchange than the Soviet Union, Soviet officials did not embrace the supposed benefits of nuclear superiority. The Soviet Union conducted three exercises in the early 1970s aimed at predicting the consequences of a large-scale nuclear exchange with the United States.[103] At the time, the Soviet Union had obtained a moderate lead in overall megatonnage, had

100. "Vitalii Leonidovich Kataev Papers," box 5, folder 8, document 1. See Document C6 in the Appendix for active citation.
101. "Vitalii Leonidovich Kataev Papers," box 5, folder 8, document 8. See Document C7 in the Appendix for active citation.
102. Interview with A. S. Kalashnikov in Hines, Ellis, and John, *Soviet Intentions 1965–1985, Volume II: Soviet Post-Cold War Testimonial Evidence*, 90.
103. Interview with Andrian A. Danilevich in ibid., 27.

reached complete parity in number of strategic launchers, and faced moderate inferiority in total number of nuclear weapons. Despite the somewhat mixed strategic balance, the models depicted devastating consequences for the Soviet Union. Danilevich, who helped oversee strategic nuclear planning in the Soviet Union at that time, described the results of the final exercise: "We explained our conclusions that after the strike the Armed Forces would be reduced to 1/1,000 of their previous strength; 80 million citizens would be dead; 85% of the industrial capability of the Soviet Union would be destroyed; the European part of the USSR would be contaminated by radiation at extremely lethal levels of 3,000 roentgens. Given all of this, the consequences of a retaliatory strike against the U.S. would be even more lethal to that country."[104]

Despite the apparent assumption that the Soviets enjoyed relative "superiority" in the expected damage from a large-scale exchange with the U.S., the assessment terrified senior officials. According to Danilevich, "Brezhnev and Kosygin were visibly terrified by what they heardâĂę During the exercise three launches of ICBMs with dummy warheads were scheduled. Brezhnev was actually provided with a button in the exercise and was to 'push the button' at the appropriate timeâĂę When the time came to push the button, Brezhnev was visibly shaken and pale and his hand trembled and he asked Grechko several times for assurances that the action would not have any real-world consequences."[105] Testimonial evidence from a range of Soviet officials confirms that from at least the late 1960s onward, the Soviet Union concluded it could not meaningfully "win" a nuclear war.[106] Despite believing that the Soviet Union would emerge from a large-scale nuclear exchange relatively better off than the United States, officials were still loathe to invite the catastrophic damage of a nuclear war.

104. Interview with Andrian A. Danilevich in Hines, Ellis, and John, *Soviet Intentions 1965–1985, Volume II: Soviet Post-Cold War Testimonial Evidence*, 27.
105. Interview with Andrian A. Danilevich in ibid.
106. Hines, Ellis, and John, *Soviet Intentions 1965–1985, Volume I: An Analytical Comparison of US-Soviet Assessments During the Cold War*, 26.

Conclusion

To claim that nuclear superiority matters is to claim that states have a meaningful, coherent, and stable conception of what nuclear superiority is, how it can be measured, and why it matters. States must have a clear understanding of the component features of nuclear superiority and of which components of the nuclear balance matter and which don't. They must have accurate information about the distribution of those components between themselves and the adversary. And they must have a coherent theory of how that distribution generates political effects. But if states conceive of, perceive, measure, and respond to the relative nuclear balance in ways which are not fully captured by the "objective" technical military features, analyses which rely on those features will necessarily fall short.

The argument presented here, that there is a crucial subjective component to how states perceive and respond to the relative nuclear balance, is not entirely new. Though he may not have employed the language of constructivism, none other than Jervis recognized the ways in which the meaning of the nuclear revolution could be an intersubjective creation of nuclear-armed states by observing that propositions about the utility of nuclear superiority "rest on decision makers' beliefs–beliefs, furthermore that can be strongly influenced by American policy and American statements. âĂę Although the Russians stress war-fighting ability, they have not contended that marginal increases in strategic forces bring political gains; any attempt to do so could be rendered less effective by an American assertion that this is nonsense."[107] The material balance matters, but it is not everything. Through their statements, perceptions, and beliefs, states create and shape the meaning of the nuclear balance.

This article illuminated two key weaknesses in some of the recent quantitative litera-

107. Jervis, "Cooperation under the Security Dilemma," 209-210.

ture on nuclear superiority: assumptions of warhead prominence in assessments of the relative nuclear balance and assumptions of symmetric and complete information about states' nuclear arsenals. Using examples from the U.S.-Soviet relationship during the Cold War, the article demonstrated why these assumptions do not hold. Warhead counts, though easily available, are neither accurate measures of the relative strategic balance, nor necessarily correlated with other measures of a state's nuclear capabilities. Using more accurate measures of state perceptions of the strategic nuclear balance, the article demonstrates that there is little evidence that nuclear superiority matters. Finally, and most significantly, it highlights the role of non-material factors in determining how states conceive of, measure, and, ultimately, respond to the relative nuclear balance.

These findings have important theoretical and methodological implications. Theoretically, these findings support one of the central tenets of the "nuclear revolution," that nuclear superiority does not convey political benefits. This theoretical contribution is significant given much recent work challenging some of the principles of the nuclear revolution.[108] It may also be the case that, as Bell and Macdonald argue, the inconclusive empirical findings about the political benefits of nuclear superiority emerge from a failure to appreciate the variation in nuclear crises.[109] This may be particularly true if states are more likely to select into certain nuclear crises based on the risks they present.

Methodologically, it suggests avenues for improving research on the implications of nuclear superiority. Scholars have hotly debated the appropriateness of applying different methodologies to questions of nuclear weapons.[110] The analysis presented here suggests

108. See, for example, Austin Long and Brendan Rittenhouse Green, "Stalking the Secure Second Strike: Intelligence, Counterforce, and Nuclear Strategy," *Journal of Strategic Studies* 38, nos. 1-2 (2015): 38–73; Brendan R Green and Austin Long, "The MAD Who Wasn't There: Soviet Reactions to the Late Cold War Nuclear Balance," *Security Studies* 26, no. 4 (2017): 606–641; and Bell, "Nuclear Opportunism: A Theory of How States Use Nuclear Weapons in International Politics"

109. Mark S Bell and Julia Macdonald, "How to Think About Nuclear Crises (February 2019)," *Texas National Security Review*, 2019,

110. See, especially, the debates in Scott D Sagan, "Two Renaissances in Nuclear Security Studies," in *H-Diplo/ISSF Forum*, vol. 2 (2014); Gavin, "What We Talk About When We Talk About Nuclear Weapons: A

skepticism about the ability of some large-N methods to satisfactorily investigate some of the research questions surrounding nuclear weapons. Indeed, much past research applying large-N methods to questions of nuclear weapons has been shown to be either quite fragile to coding decisions or to shed little light on real-world dynamics.[111] Some concepts, such as perceptions of nuclear superiority, may not be amenable to quantification, especially given the importance of states' information and beliefs, which may vary significantly over time and space. The use of large-N regression techniques, such as those scrutinized here, flattens this data, assuming that each observation holds equal weight and that the impact of nuclear superiority is similar across all observations. However, the evidence does not suggest that nuclear superiority exerts an average probabilistic effect across all observations. Instead, it is likely to be highly contingent on the information available to states in times of crisis and their beliefs about what that information means for their odds of success.[112] If state beliefs about the importance (or unimportance) of nuclear superiority can be learned and un-learned, this may not be captured by large-N methods.[113] If state responses to nuclear superiority vary with the state's political leadership, this may not be captured.[114] Nuclear superiority and its effects are not objective and mechanistic. Rather, nuclear superiority may be, to some extent, what states make of it.[115] Assessing the beliefs and information of states, and how states perceive or respond to nuclear superiority is perhaps best done through careful attention to the particular circumstances of the state, a

Review Essay"; Fuhrmann, Kroenig, and Sechser, "The Case for Using Statistics to Study Nuclear Security"; and Francis J Gavin, *Nuclear Weapons and American Grand Strategy* (Brookings Institution Press, 2020).

111. See, for example, Alexander H Montgomery and Scott D Sagan, "The Perils of Predicting Proliferation," *Journal of Conflict Resolution* 53, no. 2 (2009): 302–328; Bell and Miller, "Questioning the Effect of Nuclear Weapons on Conflict"; and Mark S Bell, "Examining Explanations for Nuclear Proliferation," *International Studies Quarterly* 60, no. 3 (2016): 520–529.

112. Gavin, "What We Talk About When We Talk About Nuclear Weapons: A Review Essay," chapter 3.

113. Joseph S Nye, "Nuclear Learning and US–Soviet Security Regimes," *International Organization* 41, no. 3 (1987): 371–402.

114. Campbell Craig and Sergey Radchenko, "MAD, not Marx: Khrushchev and the Nuclear Revolution," *Journal of Strategic Studies* 41, nos. 1-2 (2018): 208–233.

115. Jervis, "Cooperation under the Security Dilemma," 208-210.

task to which focused qualitative work may be better suited.

If, however, the field persists in applying quantitative methodologies to questions about the political effects of nuclear weapons, more work needs to be done to ensure that those methods best measure the phenomena to which they are applied. So far, scholars may have used the relative size of warhead stockpiles to measure the relative nuclear balance between states because such data is readily available. However, such data may not adequately represent the dynamics scholars are interested in modeling. Scholars ought to do the hard work to ensure that their data and methods best resemble the true mechanisms they are investigating. This means first interrogating the historical record to determine state beliefs and perceptions about how to evaluate nuclear capabilities and whether those perceptions of nuclear superiority matter. Then, based on an understanding of those perceptions, scholars should construct datasets which better reflect the variables that matter. In short, theories based on the historical record should guide the compilation of data, not the other way around.

Research on whether or not nuclear superiority matters also has important policy implications. As North Korea has strengthened its nuclear capabilities, observers have debated whether U.S. nuclear superiority will matter vis-a-vis Pyongyang, while U.S. President Donald Trump has boasted of having a "bigger [nuclear] button" than his North Korean counterpart, suggesting he believes in the advantages of nuclear superiority.[116] Some scholars have seized on evidence that nuclear superiority matters to lobby for a further expansion and modernization of the U.S. nuclear arsenal.[117] The recent body of research purporting to find the benefits of nuclear superiority is too small, too indeterminate, and too flawed. We will need far more numerous and far more compelling works demonstrating the benefits of nuclear superiority before casting aside longstanding tenets of interna-

116. Baker and Tackett, "Trump Says His 'Nuclear Button' Is 'Much Bigger' Than North Korea's."
117. Kroenig, *The Logic of American Nuclear Strategy: Why Strategic Superiority Matters*.

tional relations theory and potentially putting U.S. policy on the path to arms racing once again.