Testing for Victory: Nuclear weapons tests, brinkmanship, and nuclear crisis outcomes

Bryce Beschorner, Ph.D. candidate

University of Kentucky Department of Political Science

Abstract: Nuclear weapons tests are an overlooked and understudied phenomenon within international relations. I argue that nuclear-armed states use nuclear weapons tests as part of a brinkmanship strategy. More specifically, nuclear weapons tests generate costs and signal a high degree of resolve. As a result, nuclear tests can encourage less resolved opponents to back down and provide testing states with a greater likelihood of crisis victory. I test this argument on nuclear crisis outcomes from 1945 to 2016 using the International Crisis Behavior data set. The results confirm that states that test more nuclear weapons than their crisis adversaries are more likely to win that crisis. Additional analysis reveals that nuclear weapons-possessing states that do not test as much as their crisis adversary do not have a greater probability of crisis victory when compared to states that do not possess nuclear weapons. The analysis provides an important next step in evaluating brinkmanship strategies and the findings have important implications for the study of nuclear weapons and policymaking.

On 30 October 1961, the Soviet Union detonated the most powerful weapon in the history of mankind. The nuclear weapon later nicknamed the "Tsar Bomba," had an explosive yield of fifty megatons[1] (equivalent to fifty million tons of TNT) and immediately generated intense global backlash against the Soviet Union and the testing of nuclear weapons (Wellerstein 2021). While the explosive yield of the Tsar Bomba makes it unique among nuclear weapons tests, the international community's response to the test and the Soviet rationale for conducting the test was not unique. An obvious reason why the Soviets decided to test a nuclear weapon so powerful, or any nuclear weapons for that matter, is to gather data to maintain and improve existing weapon systems or develop new ones. However, historical evidence indicates that the global number of nuclear weapons tests exceeds the number of weapons tests states would have conducted if they only cared about using these tests to gather data. This raises the following question: why have states conducted so many nuclear weapons tests?

I argue that nuclear weapons tests serve a dual-purpose for the state carrying out the tests. Nuclear weapons tests allow states to learn about and develop new capabilities but can also be a part of a brinkmanship strategy for crisis participants. In brinkmanship, states exert pressure on their adversaries by taking steps and carrying out provocative actions that raise the risk that events will accidentally spiral out of control and lead to an exchange of nuclear weapons (Schelling 1966). According to brinkmanship then, the state most willing to take provocative actions that "leave something to chance" and run the risk of nuclear war is most likely to emerge victorious from a nuclear crisis, with the level of risk-acceptance of each crisis participant directly determined by their level of resolve (Powell 2003; Schelling 1966). I argue that during a

[1] The Tsar Bomba had a potential yield of 100 megatons but was only tested at 50 megatons because of concerns about the number of radioactive particles released in the atmosphere that would result from a 100-megaton detonation (Wellerstein 2021).

nuclear crisis, states use nuclear weapons tests to reveal the balance of resolve between themselves and their adversary, as well as manipulate the shared risk of war by pushing crisis opponents closer to war, with the winner of the crisis being determined by which state conducts more nuclear weapons tests. The Tsar Bomba test, for example, was one of several tests that the Soviet Union conducted during the Berlin Wall Crisis, a crisis in which they conducted more nuclear weapons tests than the United States and won.

While the brinkmanship model has existed for almost six decades and received intense theoretical examination through formal models and qualitative studies (e.g., Jervis 1989; Powell 1988; Powell 1990; Powell 2003; Schelling 1966), the empirical literature concerning brinkmanship is lacking for such an important model.[2] This project expands the existing empirical literature of brinkmanship, presents a novel theoretical and empirical analysis of nuclear weapons tests, while also offering a more direct empirical test of the brinkmanship model than past studies. Additionally, this project adds to the study of crisis outcomes by examining how a previously overlooked factor, nuclear weapons tests, influences crisis outcomes.

The paper proceeds as follows: I begin with a general discussion of nuclear weapons tests before introducing a theory that fits nuclear weapons tests within the brinkmanship model. I then test my theory on all crisis dyads between 1945 and 2016 where both states in the dyad possess nuclear weapons. I find that states that conduct more nuclear weapons tests than their opponent during a crisis are more likely to win the crisis, while also showing that the possession of nuclear weapons alone is not enough to achieve victory in a crisis. I conclude with a discussion of the important implications these results have for policy-making and future academic research.

[2] For a notable exception see Kroenig (2013).

The basics of nuclear weapons tests

Since the United States tested the world's first nuclear weapon in the "Trinity" test, a total of 2081 nuclear weapon tests have been conducted between 1945 and 2017 (Cha 2013; Johnston 2009; Nagdy and Roser 2013; Nitkin 2021; U.S. Department of Energy 2015). This includes 2052 tests conducted by single states, as well as 28 joint United States and United Kingdom tests and 1 unconfirmed test that is thought to be a joint South African-Israeli test.[3] According to the 1974 Threshold Test Ban Treaty (TTBT), a nuclear weapons test is defined as "either a single underground nuclear explosion conducted at a test site, or two or more underground nuclear explosions conducted at a test site within an area delineated by a circle having a diameter of two kilometers and conducted within a total period of time of 0.1 second" (Threshold Test Ban Treaty 1974: section 1.2).

Two parts of this definition need clarification. First, this definition only mentions underground nuclear weapons tests because atmospheric testing was banned with the signing of the Partial Test Ban Treaty (PTBT) in 1963. Not all nuclear powers immediately signed onto the PTBT, and some, like France and China, continued to test nuclear weapons in the atmosphere until 1974 and 1980, respectively (Johnston 2009). The count of total nuclear weapons tests presented here includes any and all nuclear weapons tests, including tests in the atmosphere, underground, and underwater. The second part of the nuclear weapons test defintion that needs clarification is the second clause describing "two or more underground nuclear explosions" only counting as one test. This was adopted because by 1974, the United States and Soviet Union had increasingly begun to use salvo shot testing. This is a form of nuclear weapons test in which the testing state detonates multiple nuclear weapons devices simultaneously at the same testing site,

[3] This is the infamous *Vela Incident*, which is thought to have most likely been a joint South African-Israeli nuclear weapons test in the south Atlantic Ocean in 1979 (Krzyzaniak 2019).

but in separate testing holes. Even though multiple nuclear devices were being detonated, both states decided that salvo shots would only count as a single test. This part of the definition is reflected in the total count of nuclear weapons tests, with salvo shot tests being counted as a single test. The following states have conducted nuclear weapons tests, with the year of the first test in parentheses: United States (1945), Soviet Union/Russia (1949), United Kingdom (1952), France (1960), China (1964), India (1974), Israel[4] (1979), South Africa[5] (1979), Pakistan (1998), and North Korea (2006). All of these states still possess nuclear weapons, with the exception of South Africa, who gave their weapons up in 1990 (de Villiers, Jardine, and Reiss 1993). Table 1 (below) gives the total number of tests for each testing state.

The dual-purpose of nuclear weapons tests

The primary reason states test nuclear weapons is to gather data in order to maintain and improve weapons systems or develop new capabilities. However, as previously stated, states have conducted more nuclear weapons tests than would be necessary if they were solely interested in using the tests to gather data. For example, before the Tsar Bomba test, the lead Soviet scientist on the project, Andrei Sakharaov, stated that he believed that if the test was successful then the Soviet Union would not need to conduct any more tests in the atmoshpere (Wellerstein 2021). He reiterated this belief after the test was successfully conducted, but the Soviet Union went on to conduct over 100 more atmospheric nuclear weapons tests in the next few months before the PTBT entered into force. These additional tests, as well as other historical evidence (Carr 2020; Wellerstein 2021), demonstrates that there is a secondary reason why states test nuclear weapons: to signal to their adversaries as part of a brinkmanship strategy during a crisis. These two reasons for testing nuclear weapons are not in competition with each other, as a

[4] This would be the *Vela Incident*.
[5] This would be the *Vela Incident*.

single test (like the Tsar Bomba) can serve both purposes. Additionally, since other states cannot easily discern the purpose of a single nuclear weapons test, we must treat each test as being theoretically and empirically the same.

Table 1: **Global nuclear weapons tests, 1945-2017**

State	Number of nuclear weapons tests	Number of tests during all crises
United States	1026	216 (21.1%)
Soviet Union/Russia	715	186 (26.0%)
France	210	29 (13.8%)
China	45	3 (6.7%)
United Kingdom	45	5 (11.1%)
North Korea	6	3 (50%)
India	3	1 (33.3%)
Pakistan	2	2 (100%)
Joint U.S.-UK	(28)	--
Joint South Africa-Israel	(1)	--
Total	2052 (2081)	445 (21.7%)

Sources: Cha (2013); Johnston (2009); Nagdy and Roser (2013); Nitkin (2021); U.S. Department of Energy (2015).

Table 1[6] lists the total number of nuclear weapons tests conducted by each testing state between 1945 and 2017 and also how many of those tests took place within international crises[7]

[6] The United States has conducted half of all single-state nuclear weapons tests, but the United States does not drive the results presented in the main analysis. See Supplementary table 4 in the supplementary file.
[7] Data on international crises comes from the International Crisis Behavior (ICB) project (Hewitt 2003).

in which the testing state was a participant. 21.7%[8] of all nuclear weapons tests[9], excluding joint tests, took place during crises. This ranges from 6.7% of all Chinese nuclear weapons tests to 100% of Pakistan's tests. Since over a fifth of all nulear tests taking place during crises, this suggests that testing may be important for crisis participants and their goals. As I will explain in later sections of the paper, nuclear weapons tests can be an effective part of a state's brinkmanship strategy, with states that conduct more tests than their adversary having a much greater chance of victory. I now briefly discuss brinkmanship.

Brinkmanship and Crisis Outcomes

The dawn of the nuclear age in the aftermath of World War II dramatically changed the nature of interstate disputes, especially for the few states who possessed nuclear weapons. Scholars have argued that due to the increased costs of war brought by nuclear weapons, states are deterred from engaging in direct military combat with one another, leading to a change in the form of interstate disputes among nuclear powers from interstate wars to international crises (Hoffman 1965). Additionally, as the form of interstate disputes among nuclear powers changed from war to crises, the nature of crises have shifted along with it, moving from competitions in military capabilities to "competitions in risk-taking" (Schelling 1966). This means that the outcome of interstate disputes, when both states possess nuclear weapons, is not going to be determined by which state has greater military capabilities but instead, by which state is the most resolved and risk-accepting when it comes to the likelihood of the dispute escalating to nuclear war. States reveal their levels of resolve and take part in this "competition in risk-taking" by using a strategy of brinkmanship.

[8] This decreases to 9.9% of all tests when focusing on crises in which both participants are nuclear powers.
[9] I only count tests that take place between the day after the crisis officially starts and the day before the crisis officially ends. This ensures that each test that is counted takes place clearly within the crisis window.

Brinkmanship is a strategy that attempts to gain an advantageous outcome for a state by exerting coercive pressure on its' adversary through taking steps and carrying out provocative actions that raise the risk that events will accidentally spiral out of control and lead to a nuclear exchange (Powell 2003; Schelling 1966). More simply put, brinkmanship is about manipulating the shared risk of nuclear war to figure out which of the crisis participants is more resolved (Schelling 1966: 99). According to the brinkmanship model then, there are 2 ways to win a crisis: first, states can take actions that reveal their relative levels of resolve to their adversary and remove the uncertainty over resolve that is underlying the crisis; or, states can take actions that bring both states closer to war, with the more risk-accepting (and resolved) state having the advantage as they are more likely to let events spiral out of control and get closest to nuclear war. As I will discuss in the following sections, some actions, like the testing of nuclear weapons, accomplish both of these tasks and can be part of an effective brinkmanship strategy.

Brinkmanship evidence

Numerous scholars have examined how different actions or nuclear capabilities can influence the balance of resolve and fit into a brinkmanship model.[10] Powell (2003) uses formal theory to examine how a national missile defense system would influence a state's level of resolve and their likelihood of victory in a nuclear crisis. Powell argues that an effective national missile defense system would give the United States more freedom of action to raise the level of risk in order to get an adversary to back down in a crisis because an effective national missile defense system will prevent some or all of the adversary's missiles from striking the United States, reducing the expected costs the United States would pay if the crisis escalates to nuclear

[10] Not all scholars accept the brinkmanship model and some even question the relevance of nuclear weapons in international politics. See Tannenwald's (1999) discussion of the nuclear taboo for an example. For evidence against the existence of the nuclear taboo see Sagan and Valentino (2017) or Horschig (2022).

war. Other scholars have used systematic qualitative analyses to evaluate brinkmaship. In a series of case studies, Betts (1987) examines explicit nuclear threats made by the United States and Soviet Union during the Cold War and how those threats influenced policymakers' decisionmaking and crisis outcomes. The results are inconclusive, as Betts concludes that neither the balance of resolve nor the nuclear balance between crisis participants provides a satisfactory explanation for crisis outcomes.

Kroenig (2013, 2018[11]) offers one of the first systematic empircial examinations of brinkmanship.[12] Kroenig argues and finds support for the notion that nuclear superior[13] states have a greater probability of victory in crises because of the effect that the balance of nuclear warheads has on the balance of resolve. States with more warheads than their adversary are thought to have a greater level of resolve because they will have lower expected costs if nuclear war does occur. Nuclear superior states have lower expected costs of war than their opponents because they are better able to engage in a counterforce targeting strategy that limits their opponents' ability to inflict damage on them. These lowered expected costs of nuclear war increase the nuclear superior state's level of resolve in a crisis and allows them to get closer to the "brink" in a nuclear crisis, giving them greater bargaining leverage over their crisis opponent.This gives the state with a larger nuclear arsenal the ability to be more risk-accepting than their adversary and gives them a greater probability of crisis victory.

By focusing on nuclear weapons tests states during a crisis, this study probes the brinkmanship argument further. A focus on nuclear weapons tests is an important next step in

[11] In Kroenig (2018) the author expands on the argument from Kroenig (2013) and renames it the superiority-brinkmanship synthesis theory.
[12] For a systematic empirical analysis that questions the validity of brinkmanship see Sechser and Fuhrmann (2013, 2017).
[13] Kroenig (2018) defines nuclear superiority as a military nuclear advantage over an opponent. It is operationalized according to a state's expected cost of nuclear war and considers the balance of nuclear warheads between adversaries.

analyzing the efficacy of brinkmanship because nuclear weapons tests capture dynamic provocative actions that take place during a crisis and are at the heart of brinkmanship strategies, as opposed to more static features of the crisis actors that do not provide new information to their opponents. To motivate the analysis, in the next section I decribe how nuclear weapons tests can form an effective part of states' brinkmanship strategy and help them prevail in crises.

Nuclear weapons tests as brinkmanship

States have used nuclear weapons tests in order to coerce and intimidate other states into backing down during a crisis. As stated earlier, the Soviet Union's Tsar Bomba test had a dual-purpose; not only did it serve as a test for high-yield nuclear weapons, but it was also part of the Soviet Union's strategy for prevailing in the Berlin Wall Crisis of 1961. Officials who were around Soviet Premier Nikita Khruschev during the discussions surrounding the decision to test the Tsar Bomba noted that some of the motivation for the timing of the test was the Berlin Wall Crisis, which was reaching its' zenith when Tsar Bomba was detonated, and Khruschev's wish to look more resolved to the United States (Wellerstein 2021). American officials reacted harshly to the test, calling it a "diplomatic bomb" whose purpose was "intimdiation, fear and blackmail" (United Press International 1961). Clearly, officials in both the Soviet Union and the United States viewed the Tsar Bomba test as having a purpose beyond just gathering data, with both sides seeing it as part of the Soviet strategy to get the U.S. and its' allies to back down in Berlin. Additionally, the International Crisis Behavior (ICB) Project considers the Berlin Wall Crisis to be a victory for the Soviet Union (Brecher et al. 2021).

How do nuclear weapons test fit into our understanding of brinkmanship? I argue that nuclear weapons tests can form part of a state's brinkmanship strategy during a crisis in two complementary ways, each of which I will discuss in the following paragraphs. First, nuclear weapons tests reveal the balance of resolve between crisis adversaries. Second, nuclear weapons tests during a crisis between two nuclear-armed states bring both sides closer to war.

Nuclear weapons tests and the balance of resolve

I argue that through nuclear weapons tests states update each other about their respective levels of resolve, which helps remove uncertainty over the overall balance of resolve. Crises only occur when there is uncertainty about the balance of resolve between two adversaries (Powell 2003). Uncertainty over the balance of resolve causes crises because if states knew where the balance of resolve stood before the crisis begins, then the less resolved state could back down and avoid the crisis entirely. Nuclear weapons tests during a crisis are informative to the testing state's opponents, because each time a state tests a nuclear weapon during a crisis, they raise the level of risk they and their opponent must pay and push the states closer to nuclear war. Additionally, because of the costs associated with nuclear weapons testing, nuclear weapons tests during a crisis serve as a form of sunk cost signalling that reveals the testing state's level of resolve. Costs associated with nuclear weapons tests include economic costs[14], politcal and reputational costs[15], health costs[16], and environmental costs.[17] The costs of nuclear weapons testing are paid ex ante and are informative in that it is assumed that states invest less in signalling when their resolve is lower because they do not value whatever the "good" or policy in dispute is enough to invest enough resources to acquire it (Fearon 1997; Gartzke et al. 2017).

[14] The United States spent about $2.3 billion dollars (in 2020 dollars) during the Crossroads test series in 1946 (Schwartz 1998).
[15] After India and Pakistan exchanged nuclear weapons tests in 1998, the United States, Japan, and other states enacted sanctions on both India and Pakistan and cut off other forms of cooperation and aid (CNN 2001).
[16] A 1991 study by the International Physicians for the Prevention of Nuclear War predicted that eventually 2.4 million people could die from cancer because of atmospheric nuclear weapons testing (IPPNW 1991).
[17] Former inhabitants of Bikini Atoll, a U.S. testing site in the 1950s, and their descendants have been unable to move back to the islands because the groundwater is contaminated due to nuclear weapons tests (Scott 2017).

During a crisis, after State A conducts a nuclear weapons test, State B needs to decide whether to raise the level of risk through their own test or back down and admit defeat. Since State B now knows that State A is somewhat resolved in the crisis because of the nuclear weapons test, some of the uncertainty around the balance of resolve between the two states is reduced. If State B decides to conduct its' own nuclear weapons test after State A's test, then both sides know that the other side has some level of resolve and they need to continue to test their nuclear weapons to remove more of the uncertainty surrounding the balance of resolve. If State B backs down after State A's test, then State A is clearly more resolved and wins the crisis. Since states incur costs for each subsuquent nuclear weapons test, during the duration of a crisis each additional nuclear weapons test reveals more information about the testing state's resolve to their adversary. The more a state tests during a crisis, the more resolved they appear to be to their opponent because of the greater costs they must pay. This is useful because the difference in the number of nuclear weapons tests conducted by crisis opponents reveals which side is more resolved, allowing the less resolved state to back down short of nuclear war. Eventually, a state will bid up the amount of risk and incur enough costs through nuclear weapons testing to reveal that they are more resolved than their opponent, thus removing the underlying uncertainty over the balance of resolve that caused the crisis to develop in the first place. Since a state's resolve, along with uncertainty about their adversary's level of resolve, directly determines a state's level of risk-acceptance; and, since the balance of resolve is revealed through nuclear weapons testing during a crisis, the state that conducts the most nuclear weapons tests during a crisis is most likely to win.

Nuclear weapons tests and manipulating the risk of war

Nuclear weapons tests fulfill one of the key elements of brinkmanship in that they reveal the balance of resolve between crisis adversaries. The second way that nuclear weapons tests conducted during a crisis fit into the brinkmanship model is that they manipulate the risk of war, moving opponents closer to nuclear war. The steps-to-war framework is useful for explaining how testing nuclear weapons during a crisis moves crisis participants closer to war. The steps-to-war framework argues that war is "chosen" step-by-step, with each step increasing the probability that the next step will be taken (Sample 2016). The core of the steps-to-war framework is the notion that territorial disputes are more likely to escalate to war when compared to other types of disputes (Senese and Vasquez 2005; Vasquez 1993). This is especially the case when states use realpolitik measures, such as forming alliances and arms racing, to attempt to win the territorial dispute by demonstrating their capability and will. The reason why these realpolitik measures make war more likely is that each subsequent step produces a security dilemma that leads each state to feel more threatened and more hostile towards the other state and causes each state to take additional steps aimed at their opponents in order to increase their own security, creating a hostile spiral (Senese and Vasquez 2005). Additionally, this vicious cycle of hostile steps solidifies the position of hardliners in domestic politics and tends to increase the number of hardliners and their influence, pushing leaders to take more realpolitik measures to deal with the threat from the opposing state (Senese and Vasquez 2005).

How does the steps-to-war approach explain why nuclear weapons tests during a crisis move states closer to nuclear war? First, in both territorial disputes and nuclear crises, neither side are completely in control of events, and there is a chance of hostilities spiraling out of control (Schelling 1966; Vasquez 1993). Not only are control of events out of the hands of participants in both territorial disputes and nuclear crises, but both situations are characterized by an overt tension and uneasiness that increase the chances of an accident or misinterpretation that leads to war.

Second, and most importantly, nuclear weapons tests can be considered a realpolitik measure. Nuclear weapons tests serve a dual-purpose that includes helping states develop new types of nuclear capabilities. Thus, it is easy to see nuclear weapons tests as a realpolitik measure that either increase the testing state's nuclear capability, like how forming an alliance or investing in an arms race increases a state's traditional military capabilities or demonstrates a previously unrevealed nuclear capability. Like other realpolitik measures then, nuclear weapons tests create a security dilemma where the crisis opponent of the testing state feels more threatened after a nuclear weapons test because of the testing state's perceived increase in nuclear capabilities, and so, undertakes their own nuclear test to improve their nuclear capabilities and overall security. This creates a hostile spiral that can escalate to war with each subsequent nuclear weapons test then as another "step" closer to war. A state that tests can push the crisis disputants closer to nuclear war which can encourage the less resolved participant to back down short of war, leading to crisis victory for the testing superior and more resolved state.

Detecting nuclear weapons tests

Finally, nuclear weapons tests would not be an effective part of a brinkmanship strategy if crisis adversaries could not detect and confirm each other's tests. The ability to detect other states' nuclear weapons tests are necessary because leaders have an incentive to misrepresent their resolve in a crisis in order to try and get the other state to back down. This means leaders may purposely lie about conducting nuclear weapons tests during a crisis in order to appear more resolved. These detection methods allow other states to confirm that their crisis opponents have actually conducted nuclear weapons tests.

There is a fine line with how much information states want to reveal with their nuclear weapons tests. On the one hand, the tests will not be an effective brinkmanship signal if other states cannot detect them. On the other hand, states are wary about revealing classified information about their nuclear programs, such as specific weapon design, that may give their crisis adversaries an advantage. States sometimes make it easy for their adversaries by announcing tests ahead of time, like with the Tsar Bomba test (Wellerstein 2021), or immediately confirm their nuclear weapons tests after they have been conducted, like with the series of Indian nuclear weapons tests in 1998. If the testing state's leaders do not publicly confirm that they have conducted a nuclear weapons tests, states have numerous methods of detection, including espionage, "sniffer" planes[18], national technical means[19], and the Comprehensive Test Ban Treaty Organization's (CTBTO) international monitoring system.[20]

Overall, nuclear weapons tests can be an effective part of a state's brinkmanship strategy during a crisis. Nuclear weapons tests fulfill the two main processes associated with brinkmanship: revealing the balance of resolve and removing the uncertainty surrounding it and moving crisis opponents closer to war. This leads to the following hypothesis:

Nuclear Testing Superiority hypothesis: During a crisis, states that conduct more nuclear weapons tests than their adversaries will be more likely to win that crisis.

[18] See Rhodes (1995: 368-372) for a description.
[19] National technical means include spy satellites and spy planes.
[20] The CTBTO's monitoring system includes seismic, hydroacoustic, infrasound, and radionuclide monitors all around the world and can detect tests that take place underground, underwater, in outer space, or in the atmosphere (CTBTO Preparatory Commission 2012).

Research Design

To test my theory of nuclear crisis outcomes, I use the International Crisis Behavior Project's (ICB) dyadic-level data set (Hewitt 2003). In combination with the ICB's system-level and actor-level data sets, the ICB dyadic-level data set provides plenty of information on crisis outcomes, the trigger and termination dates of crises, and state-level characteristics related to crisis outcomes. The ICB defines a crisis as an interstate dispute that threatens at least one state's values, has a heightened probability of military escalation, and has a finite time frame for resolution (Brecher and Wilkenfield 2000). Additionally, according to Schelling, "it is the essence of a crisis that the participants are not fully in control of events" (Schelling 1966). A nuclear crisis can occur whether or not nuclear weapons are used, are explicitly threatened, or are the disputed issue in the crisis (Kroenig 2013).

The temporal domain for this analysis is from 1945 to 2016. The unit of analysis is the dyad crisis, as the main independent variable, nuclear testing superiority, can only be measured at the dyadic-level. I employ directed-dyads because the dependent variable, crisis outcomes, varies for each state in the dyad. Another difference between my analysis and the existing literature, is that I only include two-sided crises in my sample, while other authors (Kroenig 2013) include both one-sided and two-sided crises.[21] This leaves me with a sample of 62 crisis-dyads across 24 different crises[22] in which both states possess nuclear weapons[23] at the time of the crisis.

---

[21] This is necessary because the ICB actor-level data set, from which I derive my dependent variable and many of the controls, only includes information on states who are considered crisis actors. If I included one-sided crises in my sample, I would be left with missing data for at least one state in each of one-sided crises.
[22] For a list of the nuclear crises analyzed and the victors of each crisis see supplementary table 1 in the supplementary files.
[23] To determine when each state first possesses nuclear weapons, I use ICB's variable *Nuclear*. I code states as possessing nuclear weapons if they score a '3' or above, except for Israel who I recode as possessing nuclear weapons from 1970 on (Nuclear Threat Initiative 2014).

Dependent Variable: Crisis Outcomes

The dependent variable in this analysis is a binary variable called *Victory. Victory* measures whether a country achieves victory in a crisis and is drawn from the variable *OUTCOM* from the ICB's actor-level data set. The original ICB variable is an ordinal indicator that codes whether a state achieves a victory, a compromise, a stalemate, or is defeated in a crisis. Following past research (Beardsley and Asal 2009, Kroenig 2013), I dichotomize this variable to code whether a state achieves a victory in a crisis or not. *Victory* equals 1 when a state wins a crisis and is coded as a 0 for all other types of outcomes. The ICB defines a victory as a crisis in which an actor achieves its "basic goals" (Brecher et al. 2021). There can be multiple victors in a crisis, as well as cases where no clear victor is produced. Of the 62 crisis-dyads in the sample, a clear victory is achieved in 17 of them. Since the dependent variable is binary, I use probit models throughout. I also employ robust standard errors that are adjusted for clustering by crisis dyad to account for the interdependence of observations.

Main Independent Variables: Nuclear testing superiority and nuclear testing ratio

The two main independent variables are *Test Superiority* and *Test Ratio.* Both *Test Superiority* and *Test Ratio* are used to evaluate the *Nuclear Testing Superiority hypothesis*. *Test Superiority* is a binary variable that measures whether State A in each crisis-dyad conducted more nuclear weapons tests during the duration of the crisis than the other state in the dyad. To begin to construct this variable, I first counted the number of nuclear weapons tests that each state in the dyad conducted, beginning with the day after the crisis begins and ending with the day before the crisis ends. This ensures that each nuclear weapons test that is counted clearly occurs after the crisis begins and before it is resolved. I do not include nuclear weapons tests in which the sources disagree about whether it took place or not[24] (denoted by an "X" in Johnston's Archive). I only count tests that were conducted by a single state, since joint tests make it difficult to attribute the test, and the subsequent signals it sends to a single state. As previously discussed, I also count salvo shot tests as a single test, even though multiple nuclear devices are detonated during these tests. There are 14 instances (out of 62) where State A conducts more nuclear weapons tests than State B during a crisis. *Test ratio* is used to examine whether states that enjoy greater levels of nuclear weapons testing superiority are more likely to win crises. *Test ratio* is calculated as the number of nuclear weapons tests conducted by State A during the crisis divided by the total number of tests conducted by both State A and State B during the crisis. This variable can theoretically range from 0 to 1.

Data on nuclear weapons tests comes from multiple sources. For the United States, I rely on a list of tests provided by the U.S. Department of Energy (DOE) (2015). Unfortunately, similar government reports are not readily available for the other testing states. For the other testing states, I mostly rely on Johnston's Archive (Johnston 2009). To verify the accuracy of Johnston's Archive, I compare the list of United States' tests provided by the U.S. DOE and Johnston's Archive. The only differences between the two sources are that Johnston's Archive counts salvo shot tests as multiple tests instead of a single test and the DOE's list includes the 28 joint U.S.-UK tests. Once these differences are accounted for, Johnston's Archive lists the exact same number of tests, under the same codenames, as the DOE's list for the United States. This gives me confidence that Johnston's Archive is an accurate source for all nuclear weapons tests. Additionally, once salvo shot tests are accounted for, Johnston's Archive lists the exact same

[24] The exception to this rule is France. Johnston's Archive questions the July 21st, 1966 French test codenamed "Ganymede", but other sources, including the U.S. Department of Energy (2015) and Kimball (2020) provide evidence that France conducted 210 nuclear weapons tests and so I count this test as being part of the French testing regime.

number of total tests for every testing state (through 2009) as other sources like the U.S. DOE (2015) report and Kimball (2020). Johnston's Archive uses other sources to compile their nuclear weapons testing data.[25] I also used information from the *Center for Strategic and International Studies* (Cha 2013) and *Congressional Research Service* (Nitkin 2021) to fill in missing data related to North Korea's post-2009 tests since Johnston's Archive ends in 2009.

Control Variables

I control for several additional factors that may influence crisis outcomes, following previous studies (Beardsley and Asal 2009; Kroenig 2013). I control for whether a state enjoys nuclear warhead superiority over the other state in the crisis-dyad, as Kroenig (2013) finds that states that possess more nuclear warheads than their adversary are more likely to prevail in crises. I first gathered data on each states' total number of nuclear warheads each year from Our World in Data (Roser and Nagdy 2013). I then created a binary variable, *Warhead Superiority*[26], which equals 1 when State A possesses more nuclear warheads than State B during a crisis.

Next, I control for whether violence was used by either side during a crisis, as high levels of violence during a crisis may be more likely to produce a clear winner. I use the *Violence* (called *VIOL* by the ICB) variable from the ICB system-level data set. *Violence* is a four-point ordinal variable that ranges from 1 (No violence) to 4 (Full-scale war).

I also include two variables that control for the stakes of the crisis. The first, *Gravity*, measures the relative gravity of the crisis for the actors involved. The graver the threat a crisis is to an actor, the more likely they are to push harder for victory in the crisis when compared to other actors who may not be facing as grave a threat. *Gravity* uses data from the ICB actor-level

---

[25] A list of sources used by Johnston's Archive can be found by following this link: https://www.johnstonsarchive.net/nuclear/tests/ and navigating to each state's individual web page.
[26] The correlation between *Nuclear Superiority* and *Test Superiority* is 0.0946 in the main sample.

data set that codes the severity of the crisis for each actor involved, from 0 (Economic threat) to 6 (Threat to existence).[27] *Gravity* equals 1 when State A is facing a graver threat than State B. The second control variable I include to control for the stakes of the crisis is the binary variable *Proximity. Proximity* measures whether the crisis took place closer to State A or State B. The closer to home a crisis is for an actor, the greater the stake the actor has in the crisis and the outcome, and the harder they are likely to push in the crisis. *Proximity* uses data from the ICB actor-level data set that codes the location of a crisis relative to each actor and ranges from 1 (crisis took place in the actor's home territory) to 4 (elsewhere).[28] *Proximity* equals 1 when the crisis takes place closer to State A than State B.

Additionally, I control for the traditional military capabilities in the crisis dyad, using the Correlates of War's (COW) National Material Capabilities data set (Singer et al. 1972). It is possible that states with greater traditional military capabilities are more likely to win crises than their weaker foes. I create a variable *Capabilities* that equals the capabilities of State A divided by the total combined capabilities of both State A and B. Each state's military capabilities are assessed using CINC scores.

I also include the variable, *Population*, that measures State A's population in a given year. States with larger populations may be better able to absorb a nuclear attack and remain intact than states with smaller populations, and thus, may be more likely to push harder in a crisis. I take the natural log of the population data from COW's National Material Capabilities data set to create *Population* (Singer et al. 1972).

---

[27] Called *GRAVTY* in the ICB actor-level data set. I recode all cases where the ICB data indicates the gravest threat to be "other" (GRAVTY=7) to equal -1, with the assumption being that if there is not a single, clear threat to a crisis actor, then the crisis is probably less grave for them than other crisis actors facing threats coded from 0-6.

[28] Called *CRACTLOC* in the ICB actor-level data set.

The domestic regime type of each actor in a crisis may play a role in influencing crisis outcomes, as it has been argued that democracies are more likely to win crises because they select into crises that they are more likely to win (Gelpi and Griesdorf 2001). I create the variable *Democracy,* that measures whether State A is a democracy. Data on domestic regime types comes from ICB's actor-level data set, which contains an ordinal variable[29] that categorizes each crisis actor's domestic regime during the crisis. *Democracy* equals 1 when State A is a democracy during a crisis.

I also control for whether crisis participants possess second-strike nuclear capabilities. A state that possesses second-strike capabilities may be able to better resist nuclear coercion, as they could theoretically absorb a first strike and still have enough surviving nuclear forces for a retaliatory strike. Drawing once again on the ICB actor-level data set, which includes a variable measuring crisis actors' nuclear capability that ranges from 1 (no foreseeable nuclear capability) to 4 (developed nuclear capability with second strike capability)[30], I create *Second Strike*, which equals 1 when State A in the dyad possesses second-strike capabilities during a crisis.

Finally, I control for a state's security environment. A competitive security environment may influence both a state's likelihood to develop and possess nuclear weapons (and thus conduct nuclear weapons tests) and crisis outcomes. I create *Security*, which measures the average number of crises a state experiences each year. *Security* is the total number of crises a state was involved in during the years preceding the current crisis, divided by the number of years it has been since 1918 (the first year in the ICB data) or since a state has entered the international system, if later than 1918.[31]

---

[29] See *REGIME* in the ICB actor-level data set.

[30] See *NUCLEAR* in the ICB actor-level data set.

[31] See Beardsley and Asal (2009) for another explanation of how to construct this variable.

Results

To begin evaluating my hypothesis I use a series of probit models with robust standard errors that are clustered by crisis dyad. I predicted that states that conduct more nuclear weapons tests than their adversary during a crisis are more likely to win that crisis. Table 1 presents the results from the main analysis, with model 1 featuring *Test Superiority* and model 2 testing *Test Ratio*. The results across both models 1 and 2 lend support for the *Nuclear Testing Superiority hypothesis*. In model 1, *Test Superiority* has a positive coefficient and is statistically significant.[32] In model 2, *Test Ratio* has a positive coefficient and is statistically significant, as well.[33] Taken together, these results indicate that states that conduct more nuclear weapons tests that their opponents during a crisis are more likely to win that crisis. These results are supported by robustness checks in the supplementary file that controls for whether State A in the dyad initiates the crisis and removes the United States from the sample. I will discuss the substantive effects of *Test Superiority* and *Test Ratio* in models 1 and 2 below, but first, will briefly discuss the results for some of the control variables.

Interestingly, the variable *Gravity* has a negative coefficient and is statistically significant across both models 1 and 2. This is surprising because it indicates that states that are facing a graver threat than their adversaries, because of the crisis, are less likely to win the crisis. These results make more sense when the results from *Proximity* are also considered. The results from *Proximity* in both models 1 and 2 indicate that states that are geographically closer to the crisis location than their crisis adversary are more likely to win the crisis, confirming prior expectations. Altogether, the results from *Gravity* and *Proximity* indicate that states facing more

[32] The p-value for *Test Superiority* in model 1 equals 0.010.
[33] The p-value for *Test Ratio* in model 2 equals 0.052.

severe threats than their crisis adversary are less likely to win that crisis, but that this could change depending on the location of the crisis.

Table 2: **Probit models with only nuclear dyads**

	Model 1	Model 2
Test Superiority	1.93**	
	(0.75)	
Test Ratio		1.53*
		(0.79)
Warhead Superior	0.13	0.01
	(0.52)	(0.63)
Violence	0.68**	0.61**
	(0.21)	(0.17)
Gravity	-1.47**	-1.38**
	(0.73)	(0.67)
Proximity	2.16**	2.00**
	(0.62)	(0.59)
Capabilities	2.45	2.37
	(1.67)	(1.54)
Population	-0.13	-0.16
	(0.22)	(0.21)
Democracy	1.56**	1.25**
	(0.60)	(0.61)
Security	2.87**	2.42**
	(1.04)	(0.87)
Second Strike	-0.94*	-0.63
	(0.48)	(0.42)
Constant	-4.95*	-3.97
	(2.63)	(2.65)
Observations	62	62

Standard errors in parentheses
* p<.10, ** p<.05

The other interesting results are from *Democracy* and *Security.* The coefficient for *Democracy,* across both models, is positive and statistically significant. The results indicate that democracies are more likely to win crises than non-democracies. This lends additional support to the notion that democracies are more likely to be victorious in crises because they select into crises that they are more likely to win (Gelpi and Griesdorf 2001). Finally, *Security* is positive and statistically significant in both models 1 and 2. *Security* measures the average number of crises that states participated in, in the years preceding whatever the current crisis is. The results for *Security* indicate that as the average number of crises a state is involved preceding the current crisis increases, the likelihood of victory in the current crisis also increases. This likely reflects some underlying and unmeasured factor, such as great power status, that causes states to be participants in more crises but also more likely to prevail in those crises. I now move towards interpreting the substantive effects from models 1 and 2.

Substantive effects for Test Superiority and Test Ratio

To determine the substantive effects of both *Test Superiority* and *Test Ratio*, I graph a state's predicted probability of victory in a crisis based on the state's *Test Superiority* and *Test Ratio*, while holding the control variables at their means. Figure 1A illustrates the substantive effects for *Test Superiority,* while Figure 1B shows the substantive effects for *Test Ratio*. With Figure 1A, you can see that moving from the lowest value of *Test Superiority* to the highest value dramatically increases a state's probability of victory in a crisis. When states conduct less nuclear weapons tests than their crisis adversary (*Test Superiority*=0), their probability of victory in that crisis is about 5.9%.[34] As states engage in a strategy of brinkmanship and conduct more nuclear weapons tests than their crisis adversary (*Test Superiority*=1), their probability of victory in that crisis increases to about 64.4%. The difference in the probability of victory between being nuclear weapons testing inferior and nuclear weapons testing superior is 58.5%.

[34] I round all percentages to the nearest tenth.

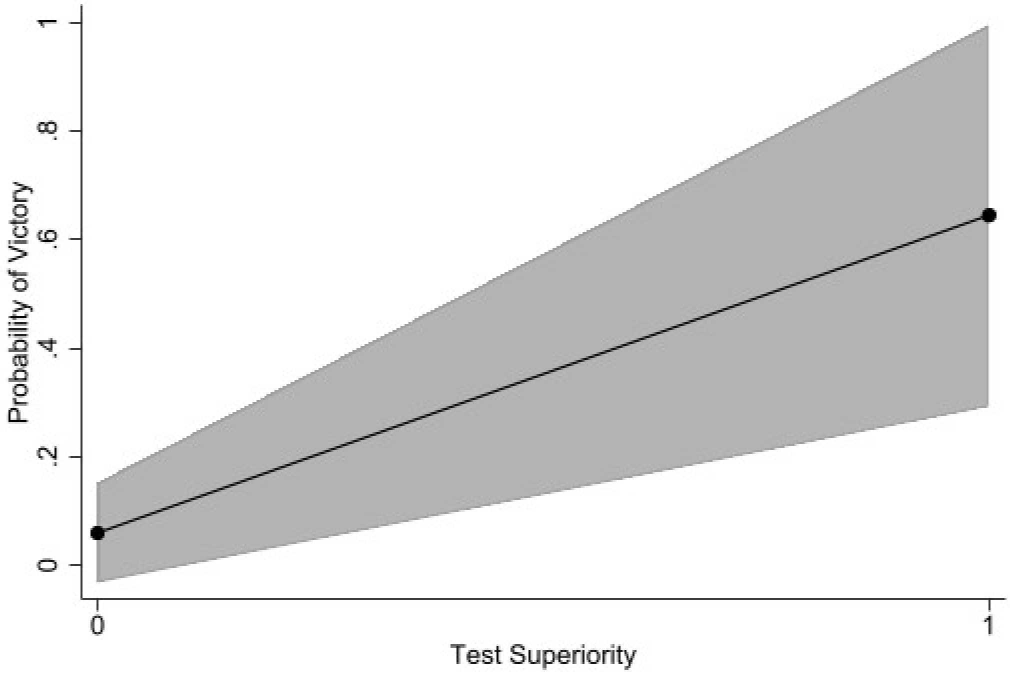

Figure 1A: Substantive effects for Test Superiority

This is a percent change of about 991.5%, a massive substantive effect. It appears that nuclear-armed states that conduct more nuclear weapons tests than their adversaries during a crisis are in a much better position to secure victory, when compared to states who choose not to test their nuclear weapons as much as their opponent during a crisis.

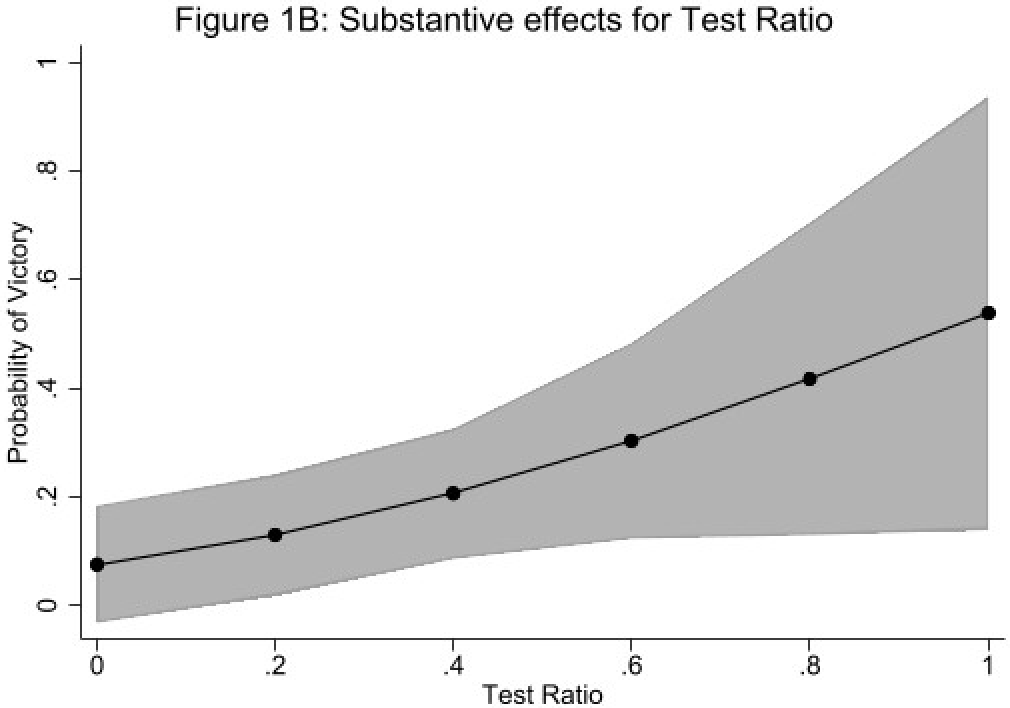

Figure 1B graphs the substantive effects for *Test Ratio*. Graphing the substantive effects of *Test Ratio* allows for a more fine-grained look at how nuclear weapons testing impacts crisis outcomes. Similarly to *Test Superiority*, moving from the lowest value of *Test Ratio* to the highest value increases a state's probability of victory in a crisis. States that do not conduct any nuclear weapons tests during a crisis (*Test Ratio*=0) have a predicted probability of victory in that crisis of about 7.6%. The predicted probability of victory increases to about 53.9% for states that conduct all of the nuclear weapons tests (*Test Ratio*=1) in their dyad during a crisis, a difference of 46.3%. Moving from the lowest level of *Test Ratio* to the highest level results in a percent change of 609.2%.

While these results largely match what we saw with *Test Superiority*, a more interesting picture emerges when we take a deeper dive into the differences among the different values for *Test Ratio*. In general, the differences in predicted probability of victory between the different values of *Test Ratio* grow faster as we move to higher levels of *Test Ratio*. For example, the difference in the probability of victory between states with a *Test Ratio* of 0.2 and 0.0 is about 5.4%, while at the opposite end of spectrum, the difference in the probability of victory between states with a *Test Ratio* of 1.0 and 0.8 is about 12.2%. While the differences in the probability of victory grow faster as *Test Ratio* increases, the substantive effects actually decrease, as the percent change in the predicted probability of victory decreases as *Test Ratio* increases. For example, the percent change in the predicted probability of victory as a state moves from conducting 0% of a dyad's nuclear weapons tests during a crisis to conducting 20% of the tests is about 71.1%. The percent change in the predicted probability of victory as a state moves from conducting 80% of a dyad's nuclear weapons tests during a crisis to conducting 100% of the tests is only about 29.3%. This difference in the percent change in the probability of crisis victory from the lower values of *Test Ratio* to the higher values of *Test Ratio* reflects how nuclear weapons tests reveal the balance of resolve. As one of the states in a crisis dyad reaches higher values of *Test Ratio*, it is clear that they are the the more resolved state, and so moving from conducting 80% of the tests to 100% of the tests does not reveal any extra information to their opponent.

Overall, the results presented above show that states that conduct more nuclear weapons tests than their adversaries during a crisis are more likely to win that crisis and also demonstrate that nuclear weapons testing can have a large substantive impact on nuclear crisis outcomes. The results presented here and the uncertainty of the existing literature beg the question: do states that possess nuclear weapons but are nuclear weapons testing inferior (*Test Superiority*=0) have a greater probability of crisis victory than states that do not possess nuclear weapons at all? In the next section, I conduct additional empirical analysis to answer this question.

Additional Analysis

To answer the question posed above I expand my sample and bring in crisis dyads where at least one state posseses nuclear weapons or neither state possess nuclear weapons, increasing the sample size to 810. I also create 3 new independent variables based on a state's nuclear status during a crisis and whether or not they conduct more nuclear weapons tests than their crisis opponent. The first variable, *Test Superior*, is the exact same as the main independent variable from the main analysis above. The second variable, *Test Inferior*, equals 1 when a state possesses nuclear weapons but conducts less tests than their opponent during a crisis. Finally, the third variable is *No Nukes*, which equals 1 when a state does not possess nuclear weapons. I run 3 additional probit models and change the baseline category in each model, while keeping all control variables the same. The results are presented in Table 3.

Model 1 includes *Test Superior* and *No Nukes* with *Test Inferior* as the baseline category. The results from Model 1 lend additional support to the *Nuclear Testing Superiority hypothesis*, as states that are nuclear weapons testing superior have a greater probability of victory in a crisis than states that are nuclear weapons testing inferior. Interestingly, the results for the *No Nukes* variable indicates that states that do not possess nuclear weapons have a greater probability of victory in a crisis than states that possess nuclear weapons but are testing inferior, although this fails to reach significance. In model 2, both *Test Superior* and *Test Inferior* fail to reach signficance with *No Nukes* as the baseline category. Finally, the results from Model 3, where *Test Superior* is the baseline category, offers more

Table 3: **Additional analysis**

	Model 1	Model 2	Model 3
Test Superior	0.44**	0.12	
	(0.21)	(0.29)	
Test Inferior		-0.31	-0.44**
		(0.24)	(0.21)
No nukes	0.31		-0.12
	(0.24)		(0.29)
Warhead Superior	0.39	0.39	0.39
	(0.26)	(0.26)	(0.26)
Violence	0.04	0.04	0.04
	(0.04)	(0.04)	(0.04)
Gravity	-0.05	-0.05	-0.05
	(0.11)	(0.11)	(0.11)
Proximity	-0.06	-0.06	-0.06
	(0.14)	(0.14)	(0.14)
Capabilities	0.66**	0.66**	0.66**
	(0.21)	(0.21)	(0.21)
Population	-0.04	-0.04	-0.04
	(0.04)	(0.04)	(0.04)
Democracy	0.24**	0.24**	0.24**
	(0.11)	(0.11)	(0.11)
Security	-0.07	-0.07	-0.07
	(0.08)	(0.08)	(0.08)
Second Strike	-0.12	-0.12	-0.12
	(0.20)	(0.20)	(0.20)
Constant	-0.92*	-0.61*	-0.59
	(0.48)	(0.36)	(0.50)
Observations	810	810	810

Standard errors in parentheses
* p<.10, ** p<.05

support for the *Nuclear Testing Superiority hypothesis*, as *Test Inferior* is negative and statistically significant. Once again, this demonstrates that when compared to states that possess nuclear weapons and conduct more nuclear weapons tests than their adversaries during a crisis, states that possess nuclear weapons but are nuclear weapons testing inferior have a lower probability of victory in a crisis.

Taken altogether, the results from this additional analysis demonstrate support for the *Nuclear Testing Superiority hypothesis* and the notion that nuclear-armed states that conduct more nuclear weapons tests than their adversary during a crisis are more likely to win that crisis. The results do not definitively answer the question that motivated this additional analysis, as it appears that there is no statistical difference in the likelihood of crisis victory between testing inferior states and states that do not possess nuclear weapons. This has implications for policymaking still, as it shows that states that possess nuclear weapons must undertake provocative actions with those weapons, in order to demonstrate their resolve during a crisis. Just possessing nuclear weapons is not enough to improve your chances of victory. I now turn to a brief case study of the 1969 Sino-Soviet Border Conflict to illustrate how nuclear testing superiority can translate to crisis victory.

Sino-Soviet Border Conflict (1969)

The most immediate causes of the 1969 Sino-Soviet Border Conflict began in 1963, when China first raised the issue of 'unequal treaties' that were imposed on the weak Qing Dynasty in the 19th century and looked to remedy some of these historic complaints (Segal 1985). Throughout the 1960s, both the Soviets and the Chinese began the process of reinforcing their military presence along their shared border. Direct tensions began to arise again in 1968 with the Soviet invasion of Czechoslovakia and the issuing of the Brezhnev doctrine, which scared Chinese leaders into believing the Soviets may interfere with Chinese domestic politics under the guise of "strengthening international Communism." The crisis officially began on 2 March 1969, when Chinese forces ambushed Soviet troops on Zhenbao Island in the Ussuri River, which formed the border between the two states (Gerson 2010). The island had seen numerous clashes between Chinese and Soviet border forces in the past, but the fighting usually devolved into fistfights, with the occasional use of clubs, sticks, and firehoses (Gerson 2010). The aim of the Chinese ambush was to deter the Soviets from interfering in China (Segal 1985).

Soviet leadership, believing the 2 March 1969 ambush was emblematic of an increasingly aggressive and revisionist China, initiated[35] another skirmish on 15 March 1969 (Gerson 2010). This skirmish was much larger and more deadly than the 2 March incident, with the Soviets using tanks, armored personnel carriers, artillery, and other heavy weapons in order to teach the Chinese a 'lesson'. The nuclear aspect of the crisis also began around the same time as the 15 March incident, with Soviet radio broadcasts and newspapers beginning to make nuclear threats toward China (Gerson 2010). Also, the Soviets conducted 2 nuclear weapons tests less than a month after the second skirmish, on 4 April and 13 April 1969, making it clear nuclear weapons would play a role in the crisis. These initial threats were originally not seen as credible by Chinese leadership but by August 1969, Chinese leaders had become increasingly concerned about the prospect of a Soviet first strike (Gerson 2010).

Once convinced that a first strike was imminent, Chinese leaders immediately began to prepare for war, including testing nuclear weapons on 22 September and 29 September (Johnston 2009; Segal 1985). The crisis finally ended on 20 October 1969 when the Chinese finally agreed to sit down for negotiations over the border (Gerson 2010). Most scholars and analysts consider this crisis a Soviet victory[36], as the crisis was not over what a final border deal between the Soviets and Chinese would look like, but instead whether they would hold talks over the border

---

[35] While both sides accused the other side of initiating the 15 March 1969 skirmish most scholars and analysts, including a CIA assessment, believe the Soviets initiated this battle (Gerson 2010).

[36] Two notable exceptions: the ICB Project (Brecher et al. 2021) and Sechser and Fuhrmann (2017) consider this crisis to be a stalemate or inconclusive.

at all (Kroenig 2018). This was clearly settled in the Soviet Union's favor as the Chinese leadership unwillingly submitted to negotiations over the border issues.

Why did the Soviet Union prevail in the 1969 Sino-Soviet Border Conflict?

The main reason why the Soviets won was that once the crisis started, they revealed themselves to be more resolved and risk-accepting than the Chinese. Kroenig (2018) argues that the Soviets won the crisis because their larger nuclear arsenal allowed them to be more risk-accepting, but this cannot be the entire story as the Soviet nuclear arsenal was superior to China's before the Chinese initiated the crisis and during the entire length of the crisis. Chinese leaders were not persuaded by their relative nuclear inferiority to not initiate the crisis or to quickly give in once the crisis began, as they held off on Soviet demands for negotiations for over 7 months. It was only after the Soviets undertook provocative, dynamic actions, such as nuclear weapons tests, that the Chinese backed down.

The Soviets undertook a few actions during the crisis that demonstrated their superior resolve and risk-acceptance to the Chinese. First, the Soviets conducted 6 nuclear weapons tests between the beginning of the crisis and August 1969, with 6 more to come while the crisis was still ongoing in September and October 1969 (Johnston 2009). This was compared to 0 tests conducted by the Chinese during that same March to August period. Second, the Soviets continued to initiate border skirmishes, including a 13 August 1969 attack that killed 38 Chinese soldiers (Gerson 2010). Not only did they continue to initiate these skirmishes, but the Soviets also escalated the skirmishes by being the side that introduced heavy weapons, like tanks and artillery, into the skirmishes. Finally, in the summer of 1969, Soviet officials began approaching foreign government officials and asking how their governments would respond to a Soviet first strike against China. This was publicly revealed to Chinese leadership and the world on 27

August 1969, when CIA Director Richard Helms announced that the Soviets had been approaching governments in Eastern Europe about an attack on China's nuclear program[37] (Gerson 2010).

I argue that the Chinese backed down in October 1969 because by then the balance of resolve between them and the Soviet Union had been clearly revealed to be in the Soviets' favor. Soviet actions clearly demonstrated they were more resolved than the Chinese during the crisis and more willing to run the risk of nuclear war. Consistent with my theory, the Soviet Union conducted 12 nuclear weapons tests during the crisis compared to only 2 for the Chinese. More importantly, 6 Soviet nuclear weapons tests were conducted during the crucial period between March 1969 and August 1969 (compared to 0 for the Chinese) in which Chinese leaders became concerned about a Soviet first strike. While we lack direct evidence of the effects of Soviet nuclear testing superiority on Chinese decision-making, it is not hard to see that these tests, in combination with other actions like continued border skirmishes and the approaching of foreign governments about a first strike, demonstrated superior Soviet resolve. Additionally, these same actions also had the effect of raising the risk of nuclear war. This caused the Chinese to back down and begin negotiations over the border, a clear victory for the Soviet Union.

Discussion and Conclusion

The findings from the analyses conducted above lead to a few policy recommendations. First, the results presented above demonstrate that states can benefit from conducting nuclear weapons tests during crises. One policy recommendation then is for nuclear-armed states to resume their nuclear weapons testing programs and be prepared to test in case a crisis breaks out. While states may not choose to use nuclear weapons tests during every crisis because of the costs

[37] Helms left out that the Soviets had also approached American officials for their thoughts about a Soviet first-strike against Chinese nuclear facilities (Gerson 2010).

associated with the tests, being prepared to conduct some tests may prove helpful in securing a crisis victory. This policy comes with some dangerous consequences though, including increasing the risk of nuclear war breaking out and increasing the amount of harm done to the physical environment and the health of human populations.

Due to these dangerous consequences, a much safer policy recommendation would be to convince states that have not signed and ratified the Comprehensive Test Ban Treaty to take steps to sign and ratify the agreement. Getting holdout states to sign and ratify the CTBT presents a significant hurdle with current politics, as competition between Israel-Iran and the United States-China make it unlikely that any of these states would agree to the terms of the CTBT without their main competitor moving first. Additionally, even absent international competition, it is unlikely the United States would be able to sign and ratify this treaty because of likely opposition from Republicans in the United States' Senate (Hansen 2006). Even with these hurdles, pushing for the CTBT to enter into force is a worthwhile goal as evidence from other international security institutions, such as the Nuclear Nonproliferation Treaty (NPT), demonstrates that increased systemic complicance with the institution decreases the likelihood of individual states violating the agreement (Kaplow 2022).

Additionally, the findings from the additional analysis can be used to strengthen and support the international nuclear non-proliferation regime because the results show that nuclear weapons only help you win crises if you are willing to incur additional costs by undertaking provocative actions with those weapons, like conducting nuclear weapons tests. States considering acquiring or developing nuclear weapons because of the benefits it may bring them in coercive diplomacy and interstate disputes should reconsider if they are not willing to take provocative and risky actions with those weapons.

This project is the first to theoretically and empircally examine nuclear weapons tests in the international relations literature. Furthermore, I argue that the empirical analyses carried out in this project offer a more direct test of the the brinkmanship model by capturing dynamic, provocative actions taken during a crisis that can reveal the balance of resolve. The main finding from this paper is as follows: as predicted in the *Nuclear Testing Superiority hypothesis*, states that conduct more nuclear weapons tests than their adversaries during a crisis are more likely to win that crisis. Future extensions of this project could empirically test other provocative actions that could be linked to brinkmanship, such as missile tests or the deployment of new weapons systems, or examine the relationship between nuclear weapons tests and other portions of the crisis process such as onset or escalation.

Works Cited

Beardsley, Kyle and Victor Asal. (2009) Winning with the Bomb. *Journal of Conflict Resolution* 53(2): 278-301.

Betts, Richard K. (1987) *Nuclear Blackmail and Nuclear Balance.* Washington D.C.: Brookings Institution Press.

Brecher, Michael and Johnathan Wilkenfield. (2000) *A Study of Crisis.* Ann Arbor, MI: University of Michigan Press.

Brecher, Michael, Johnathan Wilkenfield, Kyle Beardsley, Patrick James and David Quinn. (2021) *International Crisis Behavior Codebook, Version 14.* <*https://sites.duke.edu/icbdata/data-collections/*> (Accessed 1 September 2021).

Carr, Alan B. (2020) Of Clouds and Craters: The History of U.S. Nuclear Weapons Testing. *Los Alamos National Laboratory* and *U.S. Department of Energy*. <https://www.osti.gov/biblio/1615651-clouds-craters-history-nuclear-weapons-testing> (Accessed 2 March 2022).

Cha, Victor. (2013) North Korea's Third Nuclear Test. *Center for Strategic and International Studies.* <https://www.csis.org/analysis/north-korea%E2%80%99s-third-nuclear-test> (Accessed on 9 September 2021).

CNN. (2001) U.S. lifts final sanctions on Pakistan. *Cable News Network*. <https://web.archive.org/web/20121016201831/http://articles.cnn.com/2001-10-29/us/gen.us.pak.sanctions_1_pakistan-and-india-sanctions-nuclear-weapons?_s=PM:US> (Accessed 18 November 2021).

Comprehensive Nuclear-Test-Ban-Treaty Organization Preparatory Commission. (2012) Overview of the Verification Regime. *Comprehensive Nuclear-Test-Ban-Treaty Organization*. <https://www.ctbto.org/verification-regime/background/overview-of-the-verification-regime/#:~:text=The%20verification%20regime%20of%20the,underwater%20or%20in%20the%20atmosphere> (Accessed 7 March 2021).

de Villiers, J.W., Roger Jardine, and Mitchell Reiss. (1993) Why South Africa Gave Up the Bomb. *Foreign Affairs* 72(5): 98-109.

Fearon, James D. (1997) Signaling Foreign Policy Interests: Tying Hands versus Sinking Costs. *Journal of Conflict Resolution* 41(1): 68-90.

Gartzke, Erik A., Shannon Carcelli, J. Andres Gannon, and Jiakun Jack Zhang. (2017) Signaling in Foreign Policy. In *Oxford Research Encyclopedia of Politics*.

Gelpi, Christopher F. and Michael Griesdorf. (2001) Winners or Losers? Democracies in International Crises, 1918-1994. *American Political Science Review* 95(3): 633-647.

Gerson, Michael S. (2010) The Sino-Soviet Border Conflict: Deterrence, Escalation, and the Threat of Nuclear War in 1969. *Center for Naval Analyses*. <https://www.cna.org/archive/CNA_Files/pdf/d0022974.a2.pdf> (Accessed 22 June 2022).

Hansen, Keith A. (2006) *The Comprehensive Nuclear Test Ban Treaty: An Insider's Perspective*. Stanford, CA: Stanford University Press.

Hewitt, J. Joseph. (2003) Dyadic Processes and International Crises. *Journal of Conflict Resolution* 47(5): 669-692.

Hoffman, Stanley. (1965) *The State of War: Essays on the Theory and Practice of International Politics.* New York: Praeger.

Horschig, Doreen. (2022) Israeli Public Opinion on the Use of Nuclear Weapons: Lessons from Terror Management Theory. *Journal of Global Security Studies* 7(2).

International Physicians for the Prevention of Nuclear War. (1991) *Radioactive Heaven and Earth: The health and environmental effects of nuclear weapons testing in, on, and above the earth.* New York: The Apex Press.

Jervis, Robert. (1989) *The Meaning of the Nuclear Revolution: Statecraft and the Prospect of Armageddon*. Ithaca, NY: Cornell University Press.

Johnston, Robert. (2009) Nuclear Tests—Databases and Other Material. *Johnston's Archive.* <https://www.johnstonsarchive.net/nuclear/tests/> (Accessed on 2 March 2021).

Kaplow, Jeffrey M. (2022) State Compliance and the Track Record of International Security Institutions: Evidence from the Nuclear Nonproliferation Regime. *Journal of Global Security Studies* 7(1).

Kimball, Daryl. (2020) The Nuclear Testing Tally. *Arms Control Association*. <https://www.armscontrol.org/factsheets/nucleartesttally> (Accessed 2 March 2022).

Kroenig, Matthew. (2013) Nuclear Superiority and the Balance of Resolve: Explaining Nuclear Crisis Outcomes. *International Organization* 67: 141-171.

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

Krzyzaniak, John. (2019) Double Flash: Forty years ago, the Carter administration covered up a presumed Israeli nuclear test. *Bulletin of Atomic Scientists*. <https://thebulletin.org/2019/09/double-flash-vela-incident/> (Accessed 13 November 2021).

Nikitin, Mary Beth D. (2021) "North Korea's Nuclear Weapons and Missile Programs." *Congressional Research Service*. <https://crsreports.congress.gov/product/pdf/IF/IF10472> (Accessed 13 November 2021).

Nuclear Threat Initiative. (2014) Israel Nuclear Overview. *Nuclear Threat Initiative.* <https://www.nti.org/analysis/articles/israel-nuclear/> (Accessed on 9 March 2022).

Powell, Robert. (1988) Nuclear Brinkmanship with Two-Sided Incomplete Information. *American Political Science Review* 82(1): 155-178.

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

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

Roser, Max and Mohamed Nagdy. (2013) Nuclear Weapons. *Our World in Data.* <https://ourworldindata.org/nuclear-weapons> (Accessed on 2 March 2021).

Rhodes, Richard. (1995) *Dark Sun: The Making of the Hydrogen Bomb.* New York, NY: Simon & Schuster.

Sagan, Scott D. and Benjamin A. Valentino. (2017) Revisiting Hiroshima in Iran: What Americans Really Think about Using Nuclear Weapons and Killing Noncombatants. *International Security* 42(1): 41-79.

Sample, Susan. (2016) Anticipating War? War Preparations and the Steps-to-War Thesis. *British Journal of Political Science* 48: 489-511.

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

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

Schwartz, Stephen I. (1998) *Atomic Audit: The Costs and Consequences of U.S. Nuclear Weapons Since 1940.* Washington, D.C.: Brookings Institution Press.

Scott, Sam. (2017) "What Bikini Atoll Looks Like Today." *Stanford Magazine*. <https://medium.com/stanford-magazine/stanford-research-on-effects-of-radioactivity-from-bikini-atoll-nuclear-tests-on-coral-and-crab-dna-48459144020c> (Accessed 12 March 2021).

Sechser, Todd S. and Matthew Fuhrmann. (2013) Crisis bargaining and nuclear blackmail. *International Organization* 67(1): 173-195.

Sechser, Todd S. and Matthew Fuhrmann. (2017) *Nuclear Weapons and Coercive Diplomacy*. Cambridge, UK: Cambridge University Press. Kindle Edition.

Segal, Gerald. (1985) *Defending China.* New York, NY: Oxford University Press.

Senese, Paul D. and John A. Vasquez. (2005) Assessing the Steps to War. *British Journal of Political Science* 35(4): 607-633.

Singer, J. David, Stuart Bremer, and John Stuckey. (1972) Capability Distribution, Uncertainty, and Major Power War, 1820-1965. In *Peace, War, and Numbers*, ed. Bruce Russett, 19-48. Beverly Hills, CA: Sage Publishing.

Tannenwald, Nina. (1999) The Nuclear Taboo: The United States and the Normative Basis of Nuclear Non-Use. *International Organization* 53(3): 433-468.

Treaty Between The United States of America and The Union of Soviet Socialist Republics on the Limitation of Underground Nuclear Weapon Tests (Threshold Test Ban Treaty), U.S.-U.S.S.R, July 3, 1974. < https://2009-2017.state.gov/t/isn/5204.htm> (Accessed 3 November 2021).

Vasquez, John A. (1993) *The war puzzle.* Cambridge, UK: Cambridge University Press.

United Press International. (1961) Senator Jackson Says U.S. Must Resume Nuclear Tests in Atmosphere: RUSSIANS' BLAST HELD 'BLACKMAIL'. *The New York Times*. <https://www.nytimes.com/1961/10/31/archives/senator-jackson-says-us-must-resume-nuclear-tests-in-atmosphere.html?searchResultPosition=9> (Accessed 10 March 2021).

United States Department of Energy. (2015) United States Nuclear Tests: July 1945 through September 1992. *U.S. Department of Energy*. < https://www.nnss.gov/docs/docs_LibraryPublications/DOE_NV-209_Rev16.pdf> (Accessed 9 September 2021).

Wellerstein, Alex. (2021) An Unearthly Spectacle: The untold story of the world's biggest nuclear bomb. *Bulletin of Atomic Scientists*. <https://thebulletin.org/2021/11/the-untold-story-of-the-worlds-biggest-nuclear-bomb/> (Accessed 30 October 2021).

Supplementary files

Supplementary table 1: **Nuclear Crises, 1945-2016**

Crisis Name	Year	Nuclear Participants
Korean War	1950	Soviet Union*, United States
Suez Crisis	1956	Soviet Union*, United Kingdom, United States*
Berlin Deadline	1958	Soviet Union, United Kingdom, United States
Berlin Wall	1961	France, Soviet Union*, United Kingdom, United States
Cuban Missile Crisis	1962	Soviet Union, United States*
Congo Crisis	1964	Soviet Union, United States*
Six Day War	1967	Soviet Union, United States*
Sino-Soviet Border War	1969	China, Soviet Union
War of Attrition	1970	Israel, Soviet Union
Yom Kippur War	1973	Israel, Soviet Union, United States*
War in Angola	1975	Soviet Union*, United States
Afghanistan Invasion	1979	Soviet Union*, United States
Kashmir	1990	India, Pakistan
India/Pakistan Nuclear Tests	1998	India, Pakistan
Kargil Crisis	1999	India*, Pakistan
India Parliament Attack	2001	India*, Pakistan
Kaluchak	2002	India, Pakistan
N. Korea Nuclear Crisis III	2006	North Korea, United States
N. Korea Nuclear Crisis IV	2009	North Korea, United States
N. Korea Nuclear Crisis V	2013	North Korea, United States
India-Pakistan Border Firing	2014	India, Pakistan

Turkey-Russia Jet Incident	2015	France, Soviet Union, United Kingdom, United States
N. Korea Nuclear Crisis VI	2016	North Korea, United States
Uri Base Attack	2016	India, Pakistan

Source: Hewitt (2003)
Note: The victor in each crisis is denoted by an asterisk. Not all crises have victors, and some have multiple victors.

Supplementary files

Supplementary table 2: **Cross-Tabulations of Nuclear Crisis Outcomes, 1945-2016**

		Outcome		
		Win	*Lose*	*Total*
Testing Superiority	Yes	7 (50%)	7 (50%)	14 (100%)
	No	10 (21%)	38 (79%)	48 (100%)
	Total χ^2= 4.6331 (p=0.031)	17 (27%)	45 (73%)	62 (100%)

Supplementary table 2 shows that states have only won a clear victory in 27% of nuclear crises between 1945 and 2016. The table also shows that being nuclear testing superior during a crisis greatly enhances a state's probability of crisis victory. Nuclear testing superior states have won 50% of the nuclear crises in which they were involved, compared to only 21% for states that were nuclear testing inferior, and 27% for all crisis participants. The results of a chi-square test allow me to reject the null hypothesis that there is no relationship between nuclear testing superiority and nuclear crisis outcomes. The cross-tabulations presented here show that states who test more nuclear weapons during a crisis than their crisis adversaries are more likely to win nuclear crises.

Supplementary files

Supplementary table 3: **Robustness check controlling for whether State A initiates the crisis**

	Model 1
Test Superiority	0.37* (0.21)
State A challenger	-0.65** (0.12)
Warhead Superior	0.19 (0.19)
Violence	0.02 (0.04)
Gravity	-0.04 (0.11)
Proximity	0.02 (0.14)
Capabilities	0.75** (0.21)
Population	-0.07* (0.04)
Democracy	0.16 (0.12)
Security	-0.10 (0.12)
Second Strike	-0.19 (0.22)
Constant	-0.12 (0.36)
Observations	810

Standard errors in parentheses
* p<.10, ** p<.05

Supplementary table 3 re-runs the main model from table 2 on all crisis dyads. The key variable here is *State A challenger* which controls for whether State A in the dyad initiated the crisis. Since deterrence is easier to achieve than compellence (Schelling 1966), it may be the case that the results from table 2 occur because we are observing only cases of deterrence. The results from this analysis show this is not the case, and even when we account for whether a state initiates a crisis or not, *Test Superiority* is still significant and have a positive coefficient.

Supplementary files

Supplementary table 4: **Removing the U.S. from the analysis**

	Model 1	Model 2
Test Superiority	8.64**	0.56*
	(2.19)	(0.29)
Warhead Superior	8.59	0.06
	(7.40)	(0.21)
Violence	-1.16	0.01
	(0.84)	(0.04)
Gravity	3.87	0.003
	(2.39)	(0.12)
Proximity	-1.93	-0.10
	(1.47)	(0.14)
Capabilities	30.59*	0.275**
	(15.77)	(0.22)
Population	-4.54**	-0.08**
	(2.26)	(0.04)
Democracy	12.91	0.26*
	(9.28)	(0.12)
Security	5.88	-0.03
	(3.64)	(0.09)
Second Strike	-13.07**	-0.13
	(5.33)	(0.28)
Constant	23.58**	-0.17
	(10.36)	(0.37)
Observations	32	674

Standard errors in parentheses
* p<.10, ** p<.05

Supplementary table 4 removes the United States from all observations. Since the United States conducted the most nuclear weapons tests of any state, it may be the case that we were only observing U.S.-influenced effects in table 2. Model 1 in supplementary table 4 includes crisis dyads where both states possess nuclear weapons, while model 2 includes all crisis dyads. *Test Superiority* is positive and significant in both models.