H-Diplo | ISSF
**Forum**, No. 2 (2014)

A production of H-Diplo with the journals *Security Studies*, *International Security*, *Journal of Strategic Studies*, and the International Studies Association's Security Studies Section (ISSS).

http://issforum.org

H-Diplo/ISSF Editors: **James McAllister and Diane Labrosse**
H-Diplo/ISSF Web and Production Editor: **George Fujii**
Commissioned for H-Diplo/ISSF by **James McAllister**

Introduction by **Scott D. Sagan**

---

**H-Diplo/ISSF Forum on "What We Talk About When We Talk About Nuclear Weapons."**

Published by H-Diplo/ISSF on **15 June 2014**

Stable URL: http://issforum.org/ISSF/PDF/ISSF-Forum-2.pdf

**Contents**

## "Two Renaissances in Nuclear Security Studies,"
## **Introduction by Scott D. Sagan, Stanford University**

*"Quality is better than quantity, especially in large numbers." Anon.*

Over the past decade, two intellectual renaissances have emerged in the field of nuclear security studies. The first is in political science, where exciting new research has been published about such important subjects as the causes of nuclear weapons proliferation, the linkages between the growth of civilian nuclear power and the spread of nuclear weapons, deterrence and compellence theory and practice, and the consequences of new states acquiring atomic arsenals. A second renaissance is occurring in history, as new archives have opened up and scholars are studying such important subjects as Cold War crises, the evolution of international institutions such as the Treaty on the Non-Proliferation of Nuclear Weapons (NPT) and the International Atomic Energy Agency (IAEA), and the history of medium powers and smaller states that decided to pursue or decided to stop pursing nuclear weapons.

These two scholarly renaissances, however, have largely developed in completely separate spheres, or on parallel tracks at best, with little interchange between historians and political scientists. This is deeply unfortunate, for creative multidisciplinary research can significantly improve our understanding of complex technical, historical, and political phenomena such as the causes and consequences nuclear weapons proliferation. During the golden age of nuclear strategy in the 1950s and 1960s, for example, when many of our theories about nuclear weapons were first developed, the breadth and diversity of scholars engaged in the field was stunning. Political scientist Bernard Brodie, economist Thomas Schelling, mathematician Albert Wohlstetter, physicist Herman Kahn, and historian Roberta Wohlstetter each produced seminal contributions about nuclear weapons and strategic stability, the danger of surprise attacks, and the possibility of arms control that created both important public policy debates in Washington and personal debates in the hallways of RAND.[1] These debates on key security issues both significantly improved the quality each individual's scholarship and the collective policy relevance of academic research for the U.S. government.

In contrast, today, the vigorous debates and intellectual cross-fertilization that enhanced earlier nuclear scholarship are missing. Both political scientists and historians too often publish only in their own disciplinary journals, attend only their own professional conferences, care only about policy implications of their narrow findings, and only engage in debates with members of their own academic tribes. Robert Jervis, James McAllister, and

[1] See Bernard Brodie, *Strategy in the Missile Age* (Princeton, NJ: Princeton University Press, 1959); Thomas C. Schelling, *The Strategy of Conflict* (Harvard University Press, 1960) and *Arms and Influence* (New Haven, CT: Yale University Press, 1966); Albert Wohlstetter, *The Delicate Balance of Terror* (Santa Monica, CA: The RAND Corporation, 1958) and "Nuclear Sharing: NATO and the N+1 Country," *Foreign Affairs* 39 (April 1961): 355-388; Herman Kahn, *On Thermonuclear War* (Princeton, NJ: Princeton University Press, 1961); Roberta Wohlstetter, *Pearl Harbor: Warning and Decision* (Stanford, CA: Stanford University Press, 1962).

Francis Gavin are therefore to be thanked for putting together this H-Diplo forum and for encouraging dialogue across the disciplinary divide.

This H-Diplo forum is a most welcome exchange of views about the strengths and weaknesses of different approaches to nuclear scholarship. My introduction to the forum has three sections. First, I will briefly describe some of the trends I see emerging in the new political science and history scholarship on the effects of nuclear weapons on international politics. Second, I will briefly outline the major points made by the contributors to this lively forum. Third, I will discuss how historians and political scientists can interact and contribute to, rather than simply critique, each other's work more effectively.

In the decades after the end of the Cold War, many political science students and scholars turned their attention to studying civil wars, insurgency, and terrorism, with far less research conducted on nuclear issues. Over the past decade, however – sparked in part by real-world policy concerns about North Korea, Iran, nuclear terrorism, and global disarmament – much new research has been published on nuclear weapons issues. This H-Diplo Forum focuses mostly on new nuclear security literature using large-N statistical methods,[2] but the renaissance in political science work on nuclear issues is much broader in focus and more diverse in terms of methodology than this admittedly important emerging strand of the literature. New nuclear weapons research in political science includes important case-study work examining the domestic political and psychological determinants of proliferation,[3] normative and constructivist analyses of states' and individuals' nuclear identity and ethical taboos,[4] new game theoretic models of proliferation and preventive war decisions,[5] and the use of public opinion survey

---

[2] Important works include Sonali Singh and Christopher R. Way, "The Correlates of Nuclear Proliferation: A Quantitative Test," *Journal of Conflict Resolution* 48, 6 (December 2004): 859-885; Dong-Joon Jo and Erik Gartzke, "Determinants of Nuclear Weapons Proliferation," *Journal of Conflict Resolution* 51, 1 (February 2007): 167-194; Robert Rauchhaus, Matthew Kroenig, and Erik Gartzke, eds., *The Causes and Consequences of Nuclear Proliferation* (London: Routledge, 2011); Matthew Kroenig, "Nuclear Superiority and the Balance of Resolve: Explaining Nuclear Crisis Outcomes," *International Organization* 67, 1 (January 2013): 141-171; and Todd S. Sechser and Matthew Fuhrmann, "Crisis Bargaining and Nuclear Blackmail," *International Organization* 67, 1 (January 2013): 173-195.

[3] Jacques Hymans, *The Psychology of Nuclear Proliferation: Identity, Emotions, and Foreign Policy* (New York: Cambridge University Press, 2006); Etel Solingen, *Nuclear Logics: Contrasting Paths in East Asia and the Middle East* (Princeton, NJ: Princeton University Press, 2007); Jacques Hymans, *Achieving Nuclear Ambitions: Scientists, Politicians, and Proliferation* (New York: Cambridge University Press, 2012).

[4] Alastair Ian Johnston, *Social States: China in International Institutions, 1980-2000* (Princeton, NJ: Princeton University Press, 2007); Nina Tannenwald, *The Nuclear Taboo and the Non-Use of Nuclear Weapons: The United States and the Non-Use of Nuclear Weapons Since 1945* (New York: Cambridge University Press, 2008).

[5] Alexandre Debs and Nuno P. Monteiro, "Known Unknowns: Power Shifts, Uncertainty, and War," *International Organization* 68, 1 (January 2014): 1-31.

experiments.[6] (Some of the contributions to this H-Diplo forum label the large-N statistical work as "the quantitative approach" to nuclear studies, but this is misleading since other commonly used methods – such as game theory and survey experiments – are also clearly quantitative in nature.) Some nuclear scholars – Vipin Narang and Matthew Kroenig, for example – use multiple methods, combining detailed case studies with quantitative tests to determine both broad correlations between variables and the importance of causal mechanisms.[7]

The renaissance in nuclear security studies among historians also displays considerable diversity regarding methods and approach. Important books by Marc Trachtenberg and Francis Gavin, for example, primarily use American archives to illuminate the evolution of U.S. nuclear strategy and high-level international diplomacy during the Cold War.[8] Other scholars have focused on social factors and domestic politics to examine nuclear decision making in different countries, such as Matthew Jones' book about race and nuclear weapons threats in Asia, Sasha Polakow-Suransky's work on Israeli-South African nuclear cooperation, and Tsuyoshi Hasegawa and his collaborators' work on the influence of national perceptions and misperceptions on Cold War crises. [9] Some scholars – Timothy Naftali and Aleksandr Fursenko are a prominent example – formed effective partnerships to conduct joint research in Soviet and American archives.[10] A broad and strong group of diplomatic and international historians have conducted multi-archival research to explore the creation of international treaties such the NPT and international disputes over the spread of civilian nuclear power technology.[11] Matthew Connelly creatively organized a

---

[6] Daryl G. Press, Scott D. Sagan, and Benjamin Valentino, "Atomic Aversion: Experimental Evidence on Taboos, Traditions, and the Non-Use of Nuclear Weapons," *American Political Science Review* 107, 1 (February 2013): 188-206; Michael R. Tomz and Jessica L.P. Weeks, "Public Opinion and the Democratic Peace," *American Political Science Review* 107, 4 (November 203): 849-865.

[7] Matthew Kroenig, *Exporting the Bomb: Technology Transfer and the Spread of Nuclear Weapons* (Ithaca, NY: Cornell University Press, 2010); *Intelligence and National Security* 28, 4 (Vipin Narang, *Nuclear Strategy in the Modern Era: Regional Powers and International Conflict* (Princeton, NJ: Princeton University Press, 2014).

[8] Marc Trachtenberg, *The Craft of International History: A Guide to Method* (Princeton, NJ: Princeton University Press, 2006); Francis J. Gavin, *Nuclear Statecraft: History and Strategy in America's Atomic Age* (Ithaca, NY: Cornell University Press, 2012); Marc Trachtenberg, *The Cold War and After: History, Theory, and the Logic of International Politics* (Princeton, NJ: Princeton University Press, 2012).

[9] Matthew Jones, *After Hiroshima: The United States, Race and Nuclear Weapons* (Cambridge: Cambridge University Press, 2010); Sasha Polakow-Suransky, *The Unspoken Alliance: Israel's Secret Relationship with Apartheid South Africa* (New York: Pantheon Books, 2010); Tsuyoshi Hasegawa, ed., *The Cold War in East Asia, 1945-1991* (Stanford, CA: Stanford University Press, 2011).

[10] Alexander Fursenko and Timothy Naftali, *'One Hell of a Gamble': Khrushchev, Castro, and Kennedy, 1958-1964* (New York: W.W. Norton, 1997); Alexander Fursenko and Timothy Naftali, *Khrushchev's Cold War: The Inside Story of an American Adversary* (New York: W.W. Norton, 2007).

[11] Hal Brands, "Rethinking Nonproliferation: LBJ, the Gilpatric Committee, and U.S. National Security Policy," *Journal of Cold War Studies* 8, 2 (Spring 2006): 83-113; Olav Njolstad, ed., *Nuclear Proliferation and*

large consortium of younger scholars to pool intellectual perspectives and archival data to understand the history of U.S. intelligence estimates regarding nuclear weapons use and proliferation.[12] In a related development, historians, anthropologists, and sociologists using approaches developed in science, technology, and society (STS) studies have produced stunningly innovative work on missile accuracy, the Indian nuclear power and weapons programs, the fire and blast effects of nuclear weapons use, and the links between nuclear technology and development in Africa.[13]

These are exciting developments in political science and history, but it is surprising to see how rarely work in one discipline influences work in the other. This H-Diplo forum is therefore an important and pioneering effort of cross-fertilization. Historian Francis Gavin provides a detailed and critical assessment of some of the recent large-N statistical studies on deterrence and compellence. He argues that misunderstandings about specific historical cases, especially the 1961 Berlin Crisis, have created deep flaws in the specific articles he reviews here, and are also illustrative of fundamental weaknesses in the large-N social science approach to nuclear studies more generally.

The political scientists who are the targets of Gavin's critical review – specifically Matthew Fuhrmann, Matthew Kroenig, and Todd S. Sechser – respond to Gavin's critiques and defend their use of statistical methods both individually and in a joint 'united front' essay. (That Gavin succeeds in creating a unified response from these three scholars is impressive in and of itself since Kroenig, Fuhrmann, and Sechser previously engaged in a lively debate among themselves about the issues in dispute.[14] Once again, the diplomatic adage 'the enemy of my enemy is my friend' appears to ring true.)

Historian Hal Brands then discusses some of the new archival materials that have been made available on nuclear issues, outlines some of his concerns about how political

*International Order: Challenges to the Non-Proliferation Treaty* (New York: Routledge, 2011); William Glenn Gray, "Commercial Liberties and Nuclear Anxieties: The U.S.-German Feud over Brazil 1975-1977," *International History Review* 34, 3 (Summer 2012): 449-474; Shane Maddock, *Nuclear Apartheid: The Quest for American Atomic Supremacy from World War II to the Present* (Chapel Hill, NC: The University of North Carolina Press, 2014).

[12] Matthew Connelly et al., "'General, I Have Fought Just as Many Nuclear Wars as You Have': Forecasts, Future Scenarios, and the Politics of Armageddon," *The American Historical Review* 117, 5 (2012): 1431-1460.

[13] Donald MacKenzie, *Inventing Accuracy: A Historical Sociology of Nuclear Missile Guidance* (Cambridge, MA: MIT Press, 1993); Lynn Eden, *Whole World on Fire* (Ithaca, NY: Cornell University Press, 2004); Gabrielle Hecht, *Being Nuclear: Africans and the Global Uranium Trade* (Cambridge, MA: The MIT Press, 2014); Robert S. Anderson, *Nuclear and Nation: Scientists, International Networks, and Power in India* (Chicago, IL: The University of Chicago Press, 2010); Gabrielle Hecht, ed., *Entangled Geographies: Empire and Technopolitics in the Global Cold War* (Cambridge, MA: The MIT Press, 2011).

[14] "Debating the Benefits of Nuclear Superiority for Crisis Bargaining," *The Duck of Minerva Blog*, March 2013.

scientists code cases, and identifies areas in which historians and political scientists can usefully collaborate. Brands importantly calls for diplomatic and international historians to become more "theory-literate" and to be more willing to do comparative work in order to make accurate generalizations and to contribute more effectively to policy debates. Finally, demonstrating that political scientists themselves have diverse views on the appropriate uses of large-N statistical methods to study nuclear weapons issues, Erik Gartzke provides a spirited defense of his work and that of other scholars using statistical regressions to test theories about the causes and consequences of nuclear proliferation, while Vipin Narang provides a sophisticated criticism of this scholarship for not being transparent about its own limitations and for too often overselling the robustness and magnitude of the results.

I will leave it to the readers to decide for themselves who 'wins' the various debates presented in this forum about what kinds of methods and evidence are best utilized to improve our understanding of nuclear politics. This forum should, however, be considered merely the start of a dialogue across disciplines. For historians and political scientists have much to offer each other. One way to gain the benefits of multidisciplinary research is to form joint research collaborations. Works produced by historians and political scientists working as co-authors on nuclear-related projects include, for example, my article with Jeremi Suri about Richard Nixon's Madman nuclear alert in 1969 and Hal Brands and David Palkki's co-authored work on Saddam Hussein's nuclear ambitions.[15] But even without co-authoring, historians and political scientists could usefully improve each other's scholarship.

Having been the recipient of one of Francis Gavin's critical reviews in the past, I know how valuable it can be for political scientists and historians to debate carefully the accuracy of interpretations, the facts of individual cases, and the influence of hidden assumptions on our findings.[16] Yet I also see an unfortunate tendency, as exemplified by Gavin's approach in this forum, for historians to focus their criticism on interpretations of the specific cases that they know well and accuse political scientists of getting crucial facts wrong and thus miscoding particular cases. This is a useful and fair criticism, of course, but it too easily enables political scientists to wiggle away from broader criticisms by claiming that even if one accepted that a single case was wrong, there are so many other cases represented in their data that one small change in coding does not hurt the robustness of their general findings. The nature of this kind of intellectual critique and the resulting defense is unfortunate, however, for historians can and should play even more important roles in helping to improve the research and findings of political science scholars.

---

[15] Scott D. Sagan and Jeremi Suri, "The Madman Nuclear Alert: Secrecy, Signaling, and Safety in October 1969," *International Security* 27, 4 (Spring 2003): 150-183; Hal Brands and David Palkki, "Saddam, Israel, and the Bomb: Nuclear Alarmism Justified?" *International Security* 36, 1 (Summer 2011): 133-166.

[16] Francis J. Gavin, "Politics, History and the Ivory Tower-Policy Gap in the Nuclear Proliferation Debate," *Journal of Strategic Studies* 35, 4 (August 2012): 573-600; Scott D. Sagan and Kenneth N. Waltz, "Political Scientists and Historians in Search of the Bomb," *The Journal of Strategic Studies* 36, 1 (March 2013): 143-151.

Historians (and political scientists themselves) can most usefully critique political science scholarship in one of three ways. First, they can assess whether the theory or theories developed are logically consistent. Second, they can examine the accuracy of the empirical evidence to determine whether the correlations presented by authors are valid. Third, they can dig into the details of cases – using what Alexander George calls "process tracing"[17] – to determine whether the factors and causal mechanisms posited as being influential in the theory are actually performing that function in the historical record.

It is perhaps understandable that historians like Gavin mostly challenge interpretations of individual historical cases, in this instance the 1961 Berlin Crisis, in their critiques of political science. But because Gavin does not make an effort to demonstrate that Berlin is, as he claims, "the most important and *most representative* case" (see Gavin, 6, emphasis added) of the effects of the nuclear balance between adversaries on crisis outcomes, political scientists can easily claim, as they do repeatedly in this forum, that their findings are robust. This means that even if one accepts Gavin's critique about this one case, the generalizations or findings based on a much larger set of cases in the database remain valid.

It is important that historians be much more than mere 'fact checkers' for political scientists, for they can usefully perform critical roles in contributing to all three of the kinds of critiques outlined above. Let me illustrate the point by focusing briefly on one of the studies that Gavin criticizes in his review: Matthew Kroenig's 2013 *International Organization* article, "Nuclear Superiority and the Balance of Resolve: Explaining International Crisis Outcomes." Kroenig theorizes that states with a larger number of nuclear warheads in their arsenal are more likely to win – that is, to achieve their central objectives – in crises with other nuclear weapons states. He clearly explains his argument – that leaders of superior states will have more resolve because they would suffer less in a nuclear exchange – and provides a formal model demonstrating the logical consistency of the argument. Kroenig even provides a folksy analogy to drive home the argument: "In more colloquial terms, the logic of the argument is that in a game of chicken between two cars on a collision course, one might expect the smaller car to swerve first, even if a crash would be disastrous for both."[18]

At first glance, this does seem logical. But a historical sensibility would encourage a theorist to think differently about the logic in two ways. First, one would want to know if there is any reason to believe that statesmen in many of the cases knew whether their nuclear arsenal was larger or smaller than their competitors' nuclear arsenals during crises. In some Cold War cases, such as the 1961 Berlin crisis or the 1962 Cuban Missile Crisis, the assumption that leaders were aware of the relative sizes of their arsenals seems reasonable. But I see no reason to believe that Indian or Pakistani leaders knew the size of

[17] Alexander L. George and Andrew Bennett, *Case Studies and Theory Development in the Social Sciences* (Cambridge, MA: The MIT Press, 2005).

[18] Kroenig, "Nuclear Superiority and the Balance of Resolve," 150.

their adversary's nuclear arsenal when they tested nuclear weapons in 1998, during the 1999 Kargil war, or during the 2001 crisis after the terrorist attack on the Indian parliament. (Indeed, it is not even clear that Indian and Pakistani civilian leaders knew the size of their own nuclear arsenals during this period.) Second, one would want to examine the whether the variable being measured is logically connected to the factor that the theory claims is of causal importance. Kroenig operationalizes nuclear superiority in a binary manner, coding the state that has the overall larger number of warheads as the superior state regardless of whether the difference in arsenal size is massively large or minutely small and regardless of whether the weapons are strategic (that is could be used against the enemy's homeland) or tactical battlefield weapons. But by what logic would the leaders of a state that has, for example, 4,000 nuclear weapons have more resolve than the leaders of a state that has 3,999 warheads? To extend the game of chicken analogy, by what logic would one expect the driver of a Mercedes E-500 with a larger hood ornament to have more resolve in a game of chicken than the driver of an otherwise identical E-500 with a smaller hood ornament?

This is not simply an abstract point. Kroenig, for example, notes that the Soviet Union "won" the 1979 "Invasion of Afghanistan" crisis with the United States, and claims that this case supports his theory that the Soviet leadership had more resolve than U.S. leadership because the USSR had nuclear superiority. In 1979, however, the database he uses shows that the Soviet arsenal had 27,935 nuclear warheads (strategic and tactical, deployed and not deployed) while the U.S. arsenal had "only" 24,107 warheads. I leave it to the reader to determine whether there is a logical reason to think that Soviet leaders with such a "superior" nuclear arsenal in 1979 would find such an "advantage" in overall nuclear warhead numbers to be an important source of resolve in the Afghanistan crisis. [19]

Another avenue for critiquing political science is to ask if the evidence is accurate, or in the case of large-N work, if the coding of the variables is done in a manner that is consistent with the known facts. Here, for example, rather than criticizing the coding of one or two cases, a stronger historical critique would examine whether there are systematic biases in the data that produce not one error, but a pattern of errors that push in one direction and would therefore can make the findings less robust.[20] This requires that scholars jump into

[19] Moreover, the Soviet advantage in nuclear weapons numbers was due to its expansion of tactical nuclear weapons in this period. If one looks at deployed strategic nuclear warheads (weapons on ICBMs, SLBMs, and long-range bombers), which seems to me to be a more appropriate way of operationalizing the concept of nuclear superiority, using the same database that Kroenig uses (the Natural Resources Defense Council), one finds that in 1979 the USSR had 7,035 strategic nuclear weapons and the U.S. had 15,156. With this measure of nuclear superiority, therefore, the superior state did not win the 1979 crisis. Natural Resources Defense Council, "Table of Global Nuclear Weapons Stockpiles, 1945-2002," http://www.nrdc.org/nuclear/nudb/datab19.asp.

[20] For examples see Jacques E. C. Hymans, "The Study of Nuclear Proliferation and Nonproliferation: Toward a New Consensus," in *Forecasting Nuclear Proliferation in the 2st Century: The Role of Theory,* eds. William C. Potter and Gaukhar Mukhatzhanova (Stanford, CA: Stanford University Press, 2010) and Alexander H. Montgomery and Scott D. Sagan, "The Perils of Predicting Proliferation," *Journal of Conflict Resolution* 53, 2 (April 2009): 302-328.

the details of the data, studying it the way a good historian studies the materials in an archive. (Here I should note a point of disagreement with Erik Gartzke's claims that quantitative political science studies are far more transparent than historically-oriented "qualitative" work. This claim both exaggerates the extent to which large-N scholars put all relevant data in appendices, which would make replication easier, and underappreciates the extent to which historically-oriented political scientists seek to use "active citations" to enable others to replicate and assess their findings.[21])

Gavin nicely questions the coding used in political science databases that claims that the Soviet Union "won" the 1961 Berlin crisis. But again, that is just one case. A much larger and systematic bias in the widely-used coding contained in political science databases has been suggested by Yevgeniy Kirpichevsky and Phillip Lipscy in their research on regime type and crisis outcomes. Kirpichevsky and Lipscy present a case study of the Cuban Missile Crisis and argue that the secret deal over the Jupiter missiles suggests that the outcome (always coded as a U.S. victory) was not as one-sided as is widely believed.[22] They theorize (and provide some initial statistical tests of the theory) that democratic leaders are less willing than autocratic leaders to accept secret deals that mask hidden compromises as public defeats. If Kirpichevsky and Lipscy's hypothesis is correct, our confidence in the traditional coding of crisis outcomes will be reduced and many findings about crisis management, democratic advantages, and coercive diplomacy will need to be reexamined due to systematic biases in data sets.

A third way to critique political science findings – conducting detailed case studies to assess whether the causal mechanisms deemed to be important are actually producing the outcomes – is what historians excel at. Predictably, Gavin's commentary on Berlin and other crises is at its strongest in this dimension. But I hope that political scientists do not leave the crucial task of 'process tracing' to the historians, for social scientists should always, in my view, be concerned about causation, not just correlations. Here I share Gavin's skepticism about whether Kroenig, Fuhrmann, and Sechser have "proven" their cases, for until they (or others) have studied the details of at least carefully selected cases, many of their "statistically robust" findings could still be utterly spurious.

In addition, historians, with their emphasis on explaining both continuity and change, should be even better than political scientists at identifying discontinuities that can influence the ability to make accurate generalizations across space and time. The degree to which the development of nuclear weapons in 1945 changed the dynamics of international politics, of course, is one common subject of scholarly debate in this regard. But there are

[21] Andrew Moravcsik, "Active Citation: A Precondition for Replicable Qualitative Research," *PS: Political Science & Politics* 43, 1 (January 2010): 29-35; Elizabeth N. Saunders, "Transparency without Tears: A Pragmatic Approach to Transparent Security Studies Research" (paper prepared for symposium, January 2014); Jack L. Snyder, "Active Citation: In Search of Smoking Guns or Meaningful Context?" (unpublished working paper, 2014).

[22] Yevgeniy Kirpichevsky and Phillip Lipscy, "The Dark Side of Democratic Advantage: International Crises and Secret Agreements" (working paper, Stanford University, 2014).

other potential nuclear discontinuities that need to be studied with the historian's sensitivity for seeing change and the political scientist's penchant for discerning patterns. To what degree, for example, did the creation of the NPT in 1968, by instituting international monitoring of nuclear facilities, produce a different dynamic in weapons proliferation behavior? In addition, the number of nuclear weapons states has also changed over time, but did the existence of more potential nuclear rivals alter the dynamics of deterrence or the difficulties of arms control agreements? It is often said that each child is born into a different family. In a similar way, every new nuclear state is born into a different nuclear system, which might alter both the complexity of balancing behavior in deterrent relationships and possible patterns of cooperation. These kinds of questions will best be addressed in the future through a mixture of political science and historical methodologies.

This forum is a lively start to what promises to be an ongoing effort to improve the quality and the policy relevance of nuclear scholarship in the future. Indeed, there is enough heat generated in this exchange of opinions that one can confidently predict that the forum will spark further debates. This is all to the benefit of the nuclear security studies field, provided that historians and political scientists maintain tolerance of intellectual diversity and focus on the shared goals of providing a more accurate understanding of nuclear issues and more policy-relevant scholarship. Certainly, the field of nuclear security studies has sufficient complexity and importance to deserve a big tent.

## "What We Talk About When We Talk About Nuclear Weapons: A Review Essay"[1]

by **Francis J. Gavin**, Frank Stanton Chair in Nuclear Security Policy Studies, MIT

> "Susan, we need to talk. I've been doing a lot of thinking lately. About us. I really like you, but ever since we met in that econ class in college I knew there was something missing from how I felt: quantitative reasoning. We can say we love each other all we want, but I just can't trust it without the data. And after performing an in-depth cost-benefit analysis of our relationship, I just don't think this is working out.
>
> Please know that this decision was not rash. In fact, it was anything but—it was completely devoid of emotion. I just made a series of quantitative calculations, culled from available OECD data on comparable families and conservative estimates of future likelihoods. I then assigned weights to various "feelings" based on importance, as judged by the relevant scholarly literature. From this, it was easy to determine that given all of the options available, the winning decision on both cost-effectiveness and comparative-effectiveness grounds was to see other people. It's not you, it's me. Well, it's not me either: it's just common sense, given the nature of my utility function."
>
> Josh Freedman, "It's Not You, It's Quantitative Cost-Benefit Analysis," *Timothy McSweeney's*[2]

In his memoirs, President Harry S. Truman claimed he issued an ultimatum -- the first of the atomic age -- that forced Stalin to remove Russian troops from Iran in 1946. Years later, however, former U.S. State Department official George V. Allen told political scientist Alexander George that neither he nor other high-level officials from that period – including Averell Harriman, the U.S. Ambassador to the Soviet Union, and James Byrnes, U.S. Secretary of State – knew of any explicit threat issued by the President to the Soviets during the crisis. While Allen acknowledged that sending an aircraft carrier would have sent a powerful message – "it might well have carried an atomic bomb" – he worried how scholars would portray the incident. "The 'ultimatum' story illustrates the problem of pinning down factual information in the so-called social sciences."[3]

---

[1] I would like to thank Mark Bell, Alexandre Debs, Rex Douglas, James Fearon, Rebecca Friedman, Celeste Gventer, Michael Horowitz, Robert Jervis, Colin Kahl, Austin Long, Jessica Mahoney, Mira Rapp-Hooper, Joshua Rovner, Steve Van Evera, Jane Vaynman, Rachel Whitlark, Phil Zelikow and especially Marc Trachtenberg for their helpful thoughts and comments. With apologies to Raymond Carver.

[2] Found at http://www.mcsweeneys.net/articles/its-not-you-its-quantitative-cost-benefit-analysis accessed on May 26th, 2013. See also Jory John, "Nate Silver Offers Up a Statistical Analysis of Your Failing Relationship," found at http://www.mcsweeneys.net/articles/nate-silver-offers-up-a-statistical-analysis-of-your-failing-relationship accessed May 26th, 2013.

[3] "When an incorrect statement appears in presidential memoirs, writers go on repeating it year after year and all the political scientists and historians in the country are unable to prevent its continued currency." See George V. Allen to Alexander George, June 4, 1969, Papers of George V. Allen, Correspondence File, 1945-1969, box 1, Harry S. Truman Presidential Library. I am grateful to Brian Muzas for this reference, which comes from his forthcoming Ph.D. dissertation, ""Sign of Contradiction? Religious Cultural Heritage and the Nuclear Paradox of Truman, Eisenhower, and Reagan."

The 1946 Azerbaijan crisis -- coming less than seven months after the United States became the first and only country to ever drop nuclear weapons on another country -- has both fascinated and confounded scholars. Were the Soviets compelled to leave, or deterred from violating their agreement with their wartime allies, by an ultimatum from a United States armed with atomic weapons? What exactly was communicated to the Soviets, did they understand the message as a threat, and how did it influence their policy? Or was the Soviet withdrawal unrelated to Truman's ultimatum, driven instead by what some scholars contend was their separate, successful negotiations with the new Iranian government, who granted Russia a generous split in oil revenues and promised to protect the Tudeh (communist) party in Iran?[4] More broadly, what role did U.S. atomic weapons play in the outcome, and would the crisis have played out differently in a world without the bomb? And is the ambiguity and uncertainty surrounding the 1946 crisis anomalous, or does it reflect a larger indeterminacy that marks the nuclear age?

This would not be the last time these questions would confound and perplex observers. Only weeks after the Azerbaijan crisis was resolved, the strategist Bernard Brodie published his classic work, *The Absolute Weapon*, which laid out the core challenge policymakers and strategists have wrestled with ever since the American bombings of Hiroshima and Nagasaki: "Thus far the chief purpose of our military establishment has been to win wars. From now on its chief purpose must be to avert them."[5] But how could this be done in a world where states competed ruthlessly, and where war had long been seen as a legitimate and routine policy instrument to achieve their goals? Strategists suggested that these powerful, horrific weapons might serve to prevent other states from attacking those who possessed them. But could atomic and thermonuclear weapons be used for anything more? Were there circumstances -- such as possessing an overwhelming advantage in numbers of atomic weapons vis-à-vis an adversary, or strategies, including a greater willingness to risk situations in which the weapons might go off (purposively or inadvertently) -- where wide space between deterrence and actual use could be exploited by a state to achieve more ambitious goals?

These are questions of fundamental importance that have been fiercely contested for years, and despite scores of articles and books, no consensus has been achieved. Furthermore, the salience of these issues goes beyond mere academic importance: where one sits in this debate influences how one feels about important policy questions. Consider the current discussion over Iran's purported nuclear ambitions. Would Iran use nuclear weapons simply to prevent its adversaries from attacking its homeland? Or might it exploit the bomb to issue threats, engage in brinkmanship, and challenge the United States and its allies in the greater Middle East? It is hard to formulate a consistent, logical position about

[4] For a nice summary of the competing arguments, see Kenneth M. Pollack, *The Persian Puzzle: The Conflict between Iran and America* (New York: Random House, 2004), 44-48, and 440, fn 35. Iran ultimately reneged on its agreement with the Soviet Union.

[5] Bernard Brodie, ed., *The Absolute Weapon* (New York: Harcourt, Brace, 1946), 76.

contemporary challenges without wrestling with deeper theoretical and empirical questions surrounding nuclear dynamics. After seven decades with a nuclear sword of Damocles hanging over our heads, a period which has seen crises but no nuclear use since August 1945, what do we actually know about how nuclear weapons influence and shape world politics? And what is the best way to gain insight into these critical puzzles?

Three bright young scholars believe they have a way to answer these questions. Statistical methods, an increasingly popular and professionally rewarded approach among political scientists, are the basis for the arguments made in two recent papers published in the prestigious journal *International Organization* and reviewed here – "Crisis Bargaining and Nuclear Blackmail" by Todd S. Sechser and Matthew Furhmann, and "Nuclear Superiority and the Balance of Resolve: Explaining Nuclear Crisis Outcomes" by Matthew Kroenig.[6] How does this method work? The goal is to identify and cumulate the "universe" of like cases where important issues such as nuclear blackmail and superiority were engaged, identify and "code" the key variables, and undertake statistical analysis and draw causal inferences about what mattered and what led to certain outcomes. In these articles, coding involves clearly identifying and assigning a numerical value, often binary, to variables including: which state sought to preserve the status quo and which was revisionist, what side had the greater 'stakes' in the outcome of a crisis, which country undertook greater military mobilization to demonstrate resolve, who had more nuclear weapons, and most importantly, which country "won" the standoff?

Sechser and Furhmann want to know whether nuclear weapons give states a greater ability to "compel" their adversaries to change their behavior, as opposed to merely deterring them. In order to answer this question, they've compiled an inventory of over 200 militarized compellent threats between 1918 and 2001. They contend that unlike previous studies, theirs explores nonnuclear as well as nuclear coercion, in order to generate "variation" to identify the real impact of nuclear weapons. Furthermore, they separate out crisis victories that are caused by compellence and those achieved by "brute force," pointing out that the latter are in fact compellence failures. The paper concludes that compellence is only likely to be effective when a challenger credibly seeks to seize an adversary's territory and can enact a threat with few costs to itself. Neither condition is likely to hold in nuclear standoffs: the threat to use nuclear bombs to destroy the sought-after territory hardly makes sense, and few goals are worth risking the international backlash and potential military response that nuclear use would bring. Sechser and Furhmann's statistical analysis suggests that whatever deterrent benefits nuclear weapons may confer, they are poor tools to bring about changes in international relations.

Kroenig asks a related set of questions: what is more likely to determine the outcome of a nuclear standoff, the nuclear balance or the balance of resolve? 'Qualitative' scholars have

[6] Matthew Kroenig, "Nuclear Superiority and the Balance of Resolve: Explaining Nuclear Crisis Outcomes," *International Organization* (Volume 67, Issue 1, January 2013), 141-171; Todd S. Sechser and Matthew Furhmann "Crisis Bargaining and Nuclear Blackmail" by *International Organization* (Volume 67, Issue 1, January 2013), 173-95.

explored this issue through the historical analysis of a few key episodes, but Kroenig sets out to create a systematic analysis of what he considers all the relevant cases. To do so, he also constructs his own dataset, made up of what he has identified as fifty-two nuclear crisis dyads, in order to see what factors determined the outcome. Kroenig concludes that nuclear superiority *does matter*, allowing the state in possession of larger numbers to 'win' more often. Kroenig also finds evidence that the side with greater political stakes is more likely to prevail. Building on the work of Thomas Schelling, Kroenig argues that nuclear crises are "competitions in risk taking," (142) or tests of nerve. Possessing greater numbers of nuclear weapons allows a state to run greater risks and ultimately force the weaker adversary to back down.

These two papers are interesting for a number of reasons. Furhman, Sechser, and Kroenig are part of an exciting scholarly renaissance in nuclear studies. In a field where younger scholars are increasingly incentivized to play 'small ball,' all three deserve high praise for wresting with issues of fundamental importance – both here and elsewhere, they have taken on big questions.[7] Furthermore, both papers build upon but also criticize an earlier generation of what political scientists call qualitative and case-study work on these issues.[8] Neither shy away from bold, certain claims. Sechser and Furhmann claim their findings carry "important theoretical implications," while Kroenig presents "a new theoretical explanation" and "the first comprehensive empirical examination of nuclear crisis outcomes."[9] The murkiness, the contingency, and the contention that mark many of the historical debates over specific crises (why did Soviet Russia leave Iran in 1946? Why did the Soviet Union remove its missiles from Cuba in October 1962? How should we understand the China-Soviet crisis of 1969?) are missing here, which suggests that by quantifying and analyzing these issues 'scientifically,' certainty can be established. Yet while both articles apply statistical analysis to many of the same issues and historical events, they arrive at different, almost opposite conclusions about nuclear dynamics and world politics. Sechser and Furhman contend that nuclear weapons, while good for deterrence, are not useful to compel. Kroenig claims nuclear superiority helps states prevail during standoffs, whether their goal is to deter or compel, in part because it allows them to demonstrate more resolve. These differences have sparked a spirited online

[7] Furhmann has written an important book that shows how countries receiving civilian nuclear assistance are more likely to develop a weapons program: see Matthew Fuhrmann, *Atomic Assistance: How "Atoms for Peace" Programs Cause Nuclear Insecurity*." (Ithaca, NY: Cornell University Press, 2012). Kroenig's monograph smartly recognizes (in a way most defensive realists do not) that superpowers don't like proliferation because it deters *them*. See Matthew Kroenig, *Exporting the Bomb: Technology Transfer and the Spread of Nuclear Weapons* (Ithaca, NY: Cornell University Press, 2010).

[8] While there are many qualitative books and articles that deal with these issues, the most important and impressive work in this tradition on these questions remains Richard K. Betts, *Nuclear Blackmail and Nuclear Balance* (Washington, DC: Brookings, 1987). In addition, an excellent study based on deep archival work and rigorous theory that focuses on the 1958-1962 period Daryl G. Press, *Calculating Credibility: How Leaders Assess Military Threats* (Ithaca, NY: Cornell University Press, 2006), 80-141.

[9] Sechser and Furhmann, 192; Kroenig, 141.

debate amongst the authors, and captured the attention of many younger scholars in the field.[10]

Perhaps most importantly, both pieces draw concrete policy lessons from the authors' research for contemporary decision-makers. Sechser and Furhmann contend that, "from a practical perspective, our findings have important implications for nuclear nonproliferation policy."[11] Because nuclear weapons are not useful to compel, the United States should not be unduly worried if other states get the bomb, and certainly shouldn't use military force to prevent countries like Iran from developing the bomb. Kroenig also argues his "findings are highly relevant to policy debates about arms control, nuclear disarmament, and nuclear force sizing;" his conclusions can be interpreted as providing support for a more aggressive U.S. posture vis-à-vis Iran.[12] These links to contemporary policy are not surprising: all three authors have been awarded prestigious fellowships from the Stanton foundation, which encourages scholars to produce policy-relevant work on nuclear issues, and all have published important opinion pieces in non-academic and foreign policy venues.

In many ways, the three authors are exemplars of policy-relevant scholarship, asking big questions and seeking audiences of influence. What are we to make of their claims? Does either paper resolve the decades-long debate over these long-contested issues in nuclear dynamics, provide a methodological blueprint for how scholarship in this field should move forward in the future, and present us with much needed guidance to navigate the vexing nuclear challenges of the twenty-first century?

---

[10] See http://www.whiteoliphaunt.com/duckofminerva/tag/nuclear-weapons accessed April 10, 2013.

[11] Sechser and Fuhrmann, 192.

[12] Kroenig, 143. Although as Dan Nexon points out, one could argue that Kroenig's scholarship and policy recommendations are in tension; Iran acquiring a small nuclear force that would be vastly inferior to that of the U.S. would support a relative passive policy response, i.e. such inferior forces would be useful for deterrence only. See http://www.whiteoliphaunt.com/duckofminerva/2011/12/scholarship-and-advocacy-bomb-iran.html accessed June 25, 2013. Kroenig's policy recommendations on Iran have generated strong critiques. For the original policy piece, see Matthew Kroenig, "Time to Attack Iran: Why a Strike is the Least Bad Option," *Foreign Affairs,* January/February 2012, available at http://www.foreignaffairs.com/articles/136917/matthew-kroenig/time-to-attack-iran accessed on July 2, 2013. For the critiques, see, for example, Colin H. Kahl, "Not Time to Attack Iran: Why War Should be a Last Resort," *Foreign Affairs,* March/April 2012, available at http://www.foreignaffairs.com/articles/137031/colin-h-kahl/not-time-to-attack-iran accessed on July 2, 2013; Paul R. Pillar, "Worst-casing and Best-casing Iran," *The National Interest*, December 22, 2011, accessed at http://nationalinterest.org/blog/paul-pillar/worst-casing-best-casing-iran-6307 on July 2, 2013; Stephen M. Walt, "The worst case for war with Iran," *Foreign Policy*, December 21st, 2011, accessed at http://walt.foreignpolicy.com/posts/2011/12/21/the_worst_case_for_war_with_iran on July 2nd, 2013. Basing his recommendation on the article reviewed here, Kroenig recently criticized President Obama's recent proposal to further reduce Russian and American strategic nuclear weapons. See "The Case for Overkill – Obama's Wrong: America Does Need Thousands of Nukes," *Foreign Policy*, June 18, 2013, available at http://www.foreignpolicy.com/articles/2013/06/19/the_case_for_overkill accessed on July 2, 2013.

As a historian interested in these questions, my way of assessing Sechser, Furhmann, and Kroenig's arguments is straight-forward. I would identify the most important example where these issues are engaged, look at the primary documents, see how the authors 'coded crucial variables and determine how good a job their analysis does in helping us understand both the specific crisis itself and the larger issues driving nuclear dynamics. Political scientists might describe this as running both a 'strong' and a 'critical' test; in other words, if the authors' theories don't fully explain the outcomes and causal mechanisms in the most important and most representative case, how useful are the findings in explaining the broader issues?[13]

Is there such a case? In a speech on November 10th, 1958, Soviet Premier Nikita Khrushchev demanded the Western powers – the United States, Great Britain, and France – remove their military forces from West Berlin within six months.[14] This ultimatum was the start of a tense, four-year period that many believe brought the world closer to thermonuclear war than any time before or since, culminating in the Cuban Missile Crisis of October 1962.[15] According to a leading historian of postwar international politics, "(T)he great Berlin crisis of 1958 to 1962" was "the central episode of the Cold War."[16] And as McGeorge Bundy states, "there were more than four years of political tension over the future of Berlin….Khrushchev's Berlin crisis gives us what is otherwise missing in the nuclear age: a genuine nuclear confrontation in Europe."[17]

Given the stakes and the risks, it is not unreasonable to ask that any theory of nuclear dynamics help us better understand the 1958 to 1962 thermonuclear standoff. I would want to know how (and why) each paper coded the key variables the way it did: who had

---

[13] The idea that there is such a thing as a 'crucial case' is not without controversy among political scientists; some believe that while such a standard might be applicable to deterministic phenomena (like chemical reactions), it is not appropriate to apply to non-deterministic interactions in international relations. For a thoughtful critique of the notion of a "hard or "critical" test, see James Fearon and David Laitin, "Integrating Qualitative and Quantitative Research Methods," in Janet M. Box-Steffensmeier, Henry E. Brady, and David Collier, eds., *The Oxford Handbook of Political Methodology.* New York: Oxford University Press, 2008.

[14] Speech by Soviet Premier Nikita S. Khrushchev at a Soviet-Polish Meeting in Moscow (November 10, 1958), http://germanhistorydocs.ghi-dc.org/sub_document.cfm?document_id=3089 accessed April 11, 2013.

[15] While there are some disagreements over the causes and course of the standoff, most historians see the 1958 to 1962 period as a singular crisis. "The Cuban Missile Crisis was the climax of the great Berlin crisis of 1958-1962." Marc Trachtenberg, "The structure of great power politics, 1963-1975," in Melvyn Leffler and Odd Arne Westad, *The Cambridge History of the Cold War: Volume II, Crises and Détente* (Cambridge: Cambridge University Press, 2010), 483.

[16] Trachtenberg, *A Constructed Peace,* 247.

[17] McGeorge Bundy, *Danger and Survival: Choices about the Bomb in the First Fifty Years* (New York: Random House, 1988), 358-59.

the highest stakes in the crisis, who demonstrated the most resolve, who was the aggressor and who was the 'deterrer,' what role did nuclear superiority play, and who 'won' the crisis? Most importantly, I would want to be convinced that the causal mechanisms identified by the authors did in fact drive the origins, development, and outcome of this crisis.

Perusing these documents, the limitations of the coding in both the Sechser/Furhmann and Kroenig papers – or any effort at coding this complex crisis -- becomes clear right away.[18] Looking at available archival sources, it is clear that *both* the Soviet Union and the United Sates saw the stakes as being higher for themselves, believed the other was the aggressor, and were incredulous that their adversary was willing to risk thermonuclear war to get its way. When Khrushchev told U.S. Ambassador to the Soviet Union Llewellyn Thompson the he could not believe "we would bring on such a catastrophe," the Ambassador retorted, "it was he who would be taking action to change the present situation." Khrushchev disagreed, arguing "we would be ones who would have to cross frontier."[19] To whom did the status quo in Berlin matter more? When Ambassador Thompson told Khrushchev that, "U.S. prestige everywhere in the world was at stake it its commitments to Berliners," Khrushchev scoffed. "Berlin was really of little importance to either America or the Soviet Union, so why should they get so worked up about changing the city's status?"[20]

Which state was the aggressor, and which was the status quo power during this standoff? At first glance, the answer seems obvious - the Soviets initiated the crisis and wanted to change the status of Berlin, going so far as to install medium-range nuclear missiles in Cuba to achieve their ends. Yet the Americans had long recognized that Berlin's odd occupational status was temporary and needed to be fixed. As Eisenhower put it, "we do not seek a perpetuation of the situation in Berlin; clearly, we did not contemplate 50 years in occupation there."[21] Furthermore, there are compelling reasons to believe that Khrushchev's primary goal in the period was to *prevent*, or deter, the Federal Republic of Germany from gaining access to nuclear weapons. The Eisenhower administration's seeming support for a nuclearlized Bundeswehr would have to be coded as profoundly

[18] I deal with these puzzles, dilemmas, and contradictions at greater length in "Nuclear Weapons, Statecraft, and the Berlin Crisis, 1958-1962," in Francis J. Gavin, *Nuclear Statecraft: History and Strategy in America's Atomic Age* (Ithaca, NY: Cornell University Press, 2012), 57-74.

[19] Telegram from the Embassy in the Soviet Union to the Department of State, May 24, 1961, *FRUS, 1961-1963*, Volume XIV, 67.

[20] Frederick Kempe, *Berlin 1961*, (New York: Putnam, 2011), 201. It is important to note that similar conversations occurred secretly among officials in both governments, as they tried to assess both the strategic importance of Berlin to both sides. See Gavin, *Nuclear Statecraft*, 57-59, 62-71.

[21] President's Report of His Private Session with Khrushchev, September 27, 1959, *FRUS, 1958-1960*, Volume IX, 46.

revisionist and deeply inimical to core Soviet interests.[22] As Khrushchev told his colleagues at the start of the crisis, "(A)ll we want to do is to secure the status quo."[23]

And for whom did the outcome matter more? In secret documents and public signaling, both portrayed their own 'stakes' in the crisis in the highest terms.[24] The documents reveal that the Americans held what appeared to be contradictory views: they both recognized and respected Soviet interests during the crisis, complained bitterly about having to defend a city of little strategic importance, yet at the end of the day believed that failing to defend their position could lead to a disastrous collapse of NATO and an unacceptable victory for the Soviet Union. Both Presidents Eisenhower and Kennedy, and Premier Khrushchev, were clearly willing to accept a not insubstantial risk of nuclear war to maintain their position.

Do these studies provide greater insight into what determined the outcome of the crisis, the balance of resolve, the balance of military capabilities, or some combination? There is little doubt that both sides understood the importance of demonstrating resolve: Khrushchev explicitly pursued a brinkmanship strategy, fully aware that America had nuclear superiority but believing it did not matter, while the United States pressed on despite what would seem to be greater stakes for the Soviets.[25] But is it possible to accurately code a subjective factor like interest and resolve during a nuclear crisis?[26] Sechser and Furhmann see military mobilization as an important way of signalling intent. But during the 1958-1962 period, not everyone saw things that way: both Khrushchev and Eisenhower wanted to *reduce* their conventional forces in Central Europe even as the crisis was heating up. In 1961, former Secretary of State Dean Acheson recommended signaling U.S. resolve through extensive military preparations, but National Security Advisor McGeorge Bundy took the Eisenhower/Gaullist line than any conventional mobilization

---

[22] By far the best analysis of the role of nuclear weapons during this period, including the argument that Soviet actions were motivated by fears of a nuclearized Bundeswehr, can be found in Marc Trachtenberg, *A Constructed Peace: The Making of the European Settlement, 1945-1963* (Princeton: Princeton University Press, 1999), especially 251-351.

[23] Alexander Fursenko and Timothy Naftali, *Khrushchev's Cold War*: *The Inside Story of an American Adversary* (New York: W.W. Norton, 2006), 220.

[24] The fact that not only in public but also in private, each side though its stakes were higher than those of the other reveals that the public statements were not simply made for bargaining purposes.

[25] Fursenko and Naftali, *Khrushchev's Cold War*, 243-244.

[26] The Americans, for their part, understood they had to demonstrate resolve, but were also keenly aware that they had what they believed to be a *meaningful* nuclear superiority which would only last a few more years. The evidence pulls in different directions: immersing oneself in the period, one gets the sense that U.S. nuclear superiority *had* to matter, as a conventional defense of West Berlin by NATO was hopeless. On the other hand, American leaders took Khrushchev's threats, made from an inferior military position, very seriously, and it is quite easy to imagine much different outcomes at various points in the crisis.

would undermine the "shield of nuclear deterrence."[27] In other words, calling up reserves or sending additional divisions to Europe made it seem like the West was *less likely* to use nuclear weapons, undermining the adversary's perception of their resolve. Bundy and others believed that the Soviets would only be deterred if they believed any conflict would escalate *almost immediately* to general nuclear war, a condition which would not be affected by the call-up of reserves.[28] More importantly, it may not even be possible to code resolve *ex ante*: the crisis itself reveals that one or both sides misread the other's resolve, which is the very reason for the crisis in the first place. In other words, it is the crisis behavior that reveals resolve, not the other way around.

How about the issues surrounding nuclear superiority – how it is measured and whether it translates into meaningful political power that affects crisis outcomes? These questions lie at the heart of Kroenig's arguments, and there are three flaws to the way he deals with this issue. First, his model does not appear to adequately address the most consequential aspect of the question, the issue that most concerned policymakers during the 1958 to 1962 period: whether either side believed it possessed a robust enough capability to launch a first strike and escape the ensuing response with an acceptable level of damage. While there was no consensus on the question, there is no doubt that key decision-makers in both the Eisenhower and Kennedy administrations saw the issue of 'acceptable damage' as the most important factor when considering the nuclear balance. In fact, the Kennedy administration worked on a sub-SIOP (single integrated operational plan) program after the failed Vienna summit that some within the administration believed could knock out the Soviet Union's ability to respond. As Carl Kaysen, the author of the study (and no hawk), suggested, "there are numerous reasons for believing that the assumptions are reasonable, that we have the wherewithal to execute the raid, and that, while a wide range of outcomes is possible, we have a fair probability of achieving a substantial measure of success."[29] The plan was debated and discussed throughout the fall of 1961, in language that made it clear that the key issue was not a simplistic measure of superiority but whether a strike could incapacitate the Soviet Union's ability to retaliate.[30]

[27] McGeorge Bundy, cited in Craig, *Destroying the Village*,132.

[28] It is not only quantitative models that have difficulty recognizing these subtleties. For a recent formal model that fails to recognize the Kennedy administration was worried conventional mobilization might signal a *weakening* of its resolve to use nuclear weapons during a crisis, see Scott Wolford, "Show Restraint, Signaling Resolve: Coalitions, Cooperation, and Crisis Bargaining," *American Journal of Political Science,* (2013), doi: 10.1111/ajps.12049, especially 9.

[29] For a discussion, based on primary documents and interviews, of the first-strike plan and the ensuing debate, see Fred Kaplan, "JFK's First Strike Plan," *The Atlantic*, October 2001. Accessed at http://www.theatlantic.com/past/docs/issues/2001/10/kaplan.htm on June 11th, 2013. Look also at the documents here http://www.gwu.edu/~nsarchiv/NSAEBB/NSAEBB56/ accessed June 28th, 2013.

[30] Although Robert Jervis makes the fascinating point that the United States willingness to protect an exposed Western Europe may have created mutual vulnerability far earlier than expected. "The problem of extended deterrence has another aspect that is little remarked upon. The basic logic is that the American threat was credible to the extent that the Soviets believed that the U.S. would believe that an attack on Europe was a prelude to an attack on it or that the Americans saw the Europeans as so much like themselves that

Furthermore, many in the administration – including the President himself – were aware that this potential first-strike capability was a rapidly wasting asset, even if the United States maintained a massive numerical superiority, a fact that must have driven crisis calculations. The end of this window – and its consequences for how President Kennedy viewed future nuclear confrontations – is quite clear from a September 1963 briefing:

> The President asked whether, even if we attack the USSR first, the loss to the U.S. would be unacceptable to political leaders. General Johnson replied that it would be, i.e. even if we preempt, surviving Soviet capability is sufficient to produce an unacceptable loss in the U.S.
>
> The President asked whether then in fact we are in a period of nuclear stalemate. General Johnson replied that we are.
>
> Referring to a statement of the Air Force Association which appeared in this morning's *Washington Post*, the President asked how we could obtain nuclear superiority as recommended by the Air Force Association. General Johnson said this was a very difficult question to answer. He acknowledged that there is no way, no matter what we do, to avoid unacceptable damage in the U.S. if nuclear war breaks out. He later acknowledged that it would be impossible for us to achieve nuclear superiority.[31]

---

they would respond to an attack on the former as though it were directed against the U.S. homeland. The latter argument was the one that was repeated most often-and the repetition was partly intended to make it a self-fulfilling prophesy. But neither observers nor policy-makers paid much attention to the opposite side of this coin. If the Americans really valued the Europeans to this extent, then the Soviets' ability to destroy West Europe gave them great leverage over the U.S., and they gained the equivalent of a second-strike capability long before they could retaliate against the American homeland. It is perhaps not surprising that those in office did not want to dwell on this implication of their policy because it is not an entirely comfortable one, but the silence of scholars is more puzzling." Robert Jervis, H-Diplo Roundtable Review, Francis J. Gavin. *Nuclear Statecraft: History and Strategy in America's Atomic Age*. Ithaca: Cornell University Press, 2012, found at http://h-net.msu.edu/cgi-bin/logbrowse.pl?trx=vx&list=H-Diplo&month=1309&week=a&msg=zprswBReP9o15lbZ7oUKkQ&user=&pw=, accessed September 7, 2013.

[31] Summary Record of the 517th Meeting of the National Security Council1, Washington, September 12, 1963, 11 a.m *Foreign Relations Of The United States*, 1961–1963 Volume Viii, National Security Policy, available online at http://history.state.gov/historicaldocuments/frus1961-63v08/d141 You can listen to a taped recording of the meeting as well http://www.jfklibrary.org/Asset-Viewer/Archives/JFKPOF-MTG-110-004.aspx accessed June 11th, 2013. Two years earlier, Kennedy had asked military officials for explicit answers to questions surrounding the mechanics and consequences of a U.S. first strike on the Soviets. "Berlin developments may confront us with a situation where we may desire to take the initiative in the escalation of conflict from the local to the general war level." Memorandum from General Maxwell Taylor to General Lemnitzer, 19 September 1961, enclosing memorandum on "Strategic Air Planning," accessed at http://www.gwu.edu/~nsarchiv/NSAEBB/NSAEBB56/BerlinC3.pdf on June 28th, 2013. This idea that the United States faced a closing "window of opportunity" where it could launch a first-strike was not a new idea. Secretary of State John Foster Dulles, recognizing how poor Soviet retaliatory capabilities were, lamented in 1958 that the U.S. had a first-strike capability that would not last forever. "We would probably not have another such chance. But probably we did not have the nerve to take advantage of the probabilities…. Our

Obviously, there is a profound difference between numerical superiority, no matter how large, and a nuclear balance that allows a country to launch a first-strike that allows it to absorb an acceptable level of retaliatory damage. [32]

What about Kroenig's effort to code nuclear superiority, even in the absence of a first-strike capability? Kroenig is correct that some countries – the United States and the Soviet Union – may have wanted more strategic weapons because such an advantage "limits the expected damage that a country would incur in the event of a nuclear exchange," even in the absence of a first-strike capability.[33] It is important to note, however, that the ability to increase 'damage limitation' capabilities has been driven as much if not more by *qualitative* improvements than the increase in raw numbers. SALT I and SALT II kept the strategic balance between the Soviet Union and the United States relatively stable in terms of *numbers* of strategic weapons, and even in terms of the mix among missiles, submarines launched, and bombers delivered. Yet beginning in the early 1970s under Secretary of Defense James Schlesinger and accelerating during the Ford, Carter, and Reagan administrations, the United States spent hundreds of billions of dollars on technological changes in the nuclear area that may have had a profound effect on how each side viewed the strategic balance. A wide range of initiatives to improve damage limitation capabilities, including stealth technologies, cruise missiles, increased accuracy, better targeting, improved command, control, communications, and intelligence, missile defense, target hardening, submarine silencing, and anti-submarine warfare were pursued. When new platforms, such as the B-1 bomber, MX missile, and Trident D-5, replaced old ones, the overall numbers of strategic weapons did not increase – in fact, *megatonnage* decreased -- but the damage limitation capabilities were much improved. Numerical parity in 1972 was not the same thing as numerical parity in 1985, a fact that influenced behavior on both sides.[34] The bottom line is that numerical superiority *does not always* convey the dynamics of damage limitation.

---

successors, a decade from now, might pay the price." Fursenko and Naftali, 177.

[32] It is important to note that Kroenig mischaracterizes Trachtenberg's argument. It is very clear that Trachtenberg is not focusing on simplistic numerical calculations of nuclear superiority but the more meaningful notion of whether a state can go first in a crisis and suffer acceptable damage in a response. Kroenig argues that the United States was not in such a position in this period, but as we see, it is not clear that he is correct. The September 12, 1963 briefing makes it clear that for President Kennedy, meaningful superiority meant an advantage that could be translated into better political outcomes, meant a first-strike capability, a capacity *he may have possessed earlier* but no longer had. When the Soviets could respond after being hit with enough force to make any U.S. nuclear attack nonsensical, the calculations changed dramatically.

[33] Kroenig, 149.

[34] For a good summary of this situation, see Austin Long, H-Diplo Roundtable Review, Francis J. Gavin. *Nuclear Statecraft: History and Strategy in America's Atomic Age*. Ithaca: Cornell University Press, 2012, found at [http://h-net.msu.edu/cgi-bin/logbrowse.pl?trx=vx&list=H-Diplo&month=1309&week=a&msg=zprswBReP9o15lbZ7oUKkQ&user=&pw](http://h-net.msu.edu/cgi-bin/logbrowse.pl?trx=vx&list=H-Diplo&month=1309&week=a&msg=zprswBReP9o15lbZ7oUKkQ&user=&pw)=, accessed September 7, 2013. As a high-level Soviet military official said, the U.S. superiority in qualitative factors like command, control,

Finally, why, out of nine nuclear weapons states, have the Soviet Union and the United States been the most aggressive in seeking damage limitation capabilities? And why has one country -- the United States -- pursued damage limitation deployments far more assertively than any other nuclear country, even in the decades after its primary nuclear rival collapsed? While it may seem logical that any state would want to have more weapons than its adversaries, there is another dynamic at work: credible damage limitation strategies were needed to *reassure allies* as much, and at times more than, to deter or compel adversaries. This was not done by U.S. policymakers out of the goodness of their hearts – they knew that without the assurances provided by a robust nuclear umbrella, countries such as Germany, Japan, South Korea, and a host of others might deploy their own nuclear weapons, a development that would be inimical to America's *strategic* interest, for reasons Kroenig understands all too well.[35] This underlying and powerful geopolitical logic -- that dampening nuclear proliferation amongst *friends* required security assurances which were much more credible when the United States pursued damage limitation – does not appear to be captured in either model. Yet the complex story of extended deterrence and nuclear nonproliferation was one of the key drivers of U.S. nuclear strategy (and at the heart of several of the key nuclear crises) throughout the postwar period.[36]

---

communications and intelligence meant that, "Soviet superiority in the number of launchers did not give them any real advantage. The numerical superiority reflected a mechanistic, wasteful approach to force building." Interview with A.S. Kalashnikov, April 1993, John Hines, et. al., *Soviet Intentions 1965-1985, Volume II: Soviet Post-Cold War Testimonial Evidence* (Mclean, VA: BDM Federal, 1995), 90, accessed at http://www.gwu.edu/~nsarchiv/nukevault/ebb285/vol%20II%20Kalashnikov.PDF on June 11th, 2013. As another high-level Russian defense planner pointed out, "several factors, especially accuracy, for example, increased the power of the U.S. arsenal by a factor of three." Interview with Vitalii Leonidovich Kataev, June 23rd, 1993, John Hines, et. al., *Soviet Intentions 1965-1985, Volume II: Soviet Post-Cold War Testimonial Evidence* (Mclean, VA: BDM Federal, 1995), 100, accessed at http://www.gwu.edu/~nsarchiv/nukevault/ebb285/vol%20II%20Kataev.PDF on June 13th, 2013.

Kroenig does have a variable for "second strike" forces, but it seems clear that these forces (and any damage limitation capability) only make sense if used first, a fact the Soviets fully understood. "We assumed that the U.S. would launch first and, given your focus on accuracy and relatively smaller yields per warhead, that you intended to strike our weapons and control systems in an attempt to disarm us." Interview with Vitalii Leonidovich Kataev, 100. According to Kalashnikov, the Soviet Union's "Achilles heel" was its inability to "create a sophisticated, survivable, integrated command, control and communication system" on par with the United States, which meant after an "all-out nuclear strike" the Soviets would only be able to launch 2% of their missiles. Interview with Kalashnikov, 90.

[35] "Power-projecting states, states with the ability to project conventional military power over a particular target, have a lot to lose when that target state acquires nuclear weapons.…Once that state acquires nuclear weapons, however, this strategic advantage is certainly placed at risk and may be fully lost. For these reasons, power-projecting states fear nuclear proliferation to both allied and enemy states." Matthew Kroenig *Exporting the Bomb: Technology Transfer and the Spread of Nuclear Weapons* Ithaca: Cornell University Press, 2010, 3.

[36] See Gavin, *Nuclear Statecraft* and Francis J. Gavin, "Politics, History and the Ivory Tower-Policy Gap in the Nuclear Proliferation Debate," *Journal of Strategic Studies*, August 2012, 573-600.

Perhaps most importantly in studies concerned with 'outcomes, do the papers help us better understand who 'won' the 1958-1962 standoff and why? While most historians believe you can only understand what drove both the origins and outcomes of the crisis by seeing the 1958 to 1962 period as continuous, these data sets break them down into three separate, distinct crises. What does their coding tell us? The International Crisis Behavior (ICB) project codes the 1961 construction of the Berlin Wall as a Soviet victory. But many within the Kennedy administration recognized that by walling off the *eastern* part of the city to stem the flow of refugees, the Soviets were more likely to allow the status quo in the *western* part to remain, which would make this an American victory.[37] Also, is it really clear the Soviets 'lost' the Cuban Missile Crisis – as ICB codes it -- if Castro's regime was preserved, U.S. missiles were removed from Turkey, and, most importantly, West Germany remained non-nuclear? Of course, that does not mean it was a 'loss' for the United States, as the Soviets removed the missiles from Cuba, the status quo in Berlin was maintained, and perhaps more importantly, the Americans had come to recognize that perhaps German nuclearization was not in their interests either.[38] Arguably, both the United States and the Soviet Union got what they wanted, highlighting how the zero-sum 'win-lose' approach of large-N studies is ill-suited to this case and international politics more broadly.

All of this highlights how limited any effort to code complex historical events will be. Under Sechser and Furhmann's own coding, compellence *did* work in 1962, and under Kroenig's, a vastly outgunned Soviet Union pursued nuclear brinkmanship over a long period of time – what Khrushchev called the "meniscus strategy"[39] -- and may have gotten everything it really wanted. Perhaps, one might say, that it is not fair to focus on the 1958-1962 superpower standoff: it might be unique, an outlier, too complex to code effectively.[40] But then what is 1958-1962 a 'case' of, precisely? And if these models can't tell us anything about arguably the most important and consequential nuclear standoff in history, should I take comfort that it apparently can explain why the U.S. successfully restored Jean-Bertrand Aristide in 1994 or "won" in Nicaragua in 1984? Or look at the 1983 'Able Archer' affair, perhaps the most recent case where the risk of thermonuclear was possible (if highly unlikely). The 'crisis' emerged because the Soviets were worried that a NATO war game may have been a preparation for an attack on the Soviet Union. If accurate, then what precisely was America's basic goal, if any, during this crisis? If this crisis was something one or both sides simply stumbled into by accident, with no code-able goal, how is it

[37] Fursenko and Naftali, *Khrushchev's Cold War*, 384.

[38] Trachtenberg, *A Constructed Peace*, 379-380.

[39] Fursenko and Naftali, *Khrushchev's Cold War*, 5.

[40] Kroenig implies that changing the coding on any one case will not challenge the robustness of his findings. I leave it to others to decide whether my analysis of how these models explain 1958 to 1962 affects their confidence in his overall findings, theories, and policy recommendations. See Kroenig, "Nuclear Superiority and the Balance of Resolve," fn. 69, 154-155.

relevant?[41] Unfortunately, these kinds of crucial subtleties are inevitably lost in the efforts to code and quantify complex events.

The problems in this approach, however, go well beyond the difficulty of coding. To make meaningful insights from statistical analysis, we need a certain number of like and comparable observations. But nuclear crises are rare precisely because they are so serious. Should every disagreement involving a country with the bomb be coded as a nuclear standoff? Why would we automatically assume that the nuclear balance is front and center when leaders of two nuclear states clash in some form? And shouldn't the model convey some sense of the *level* of nuclear danger, and distinguish between the apocalyptic fears produced by the Cuban Missile Crisis and the more mundane worries generated by the 1964 Congo crisis? A common-sense approach quickly tells you that for most of the cases in both datasets there was no danger that nuclear weapons would be used.[42] While the focus in both papers is on outcomes, there really should be a fuller discussion of how we even *know* something is a nuclear crisis (and how much of a crisis it is) before we start compiling and comparing them.

There are other difficulties. Consider the distinction – crucial to both sets of arguments – between deterrence and compellence, a difference made famous by Thomas Schelling.[43] There are several problems here. First, deterrence is supposed to be easier and compellence harder, a proposition that has rarely been tested in either the qualitative or quantitative literature.[44] This implies that it should have been relatively straight-forward to prevent the Soviets from taking over West Berlin but much harder to compel them to leave. But presumably the Russians would have understood that before moving in, so how could they have been deterred from taking the city in the first place? More fundamentally, as the 1958-1962 period makes clear, defining the status quo – and coding who is the

[41] There is no consensus on how threatening the 1983 crisis actually was. For an argument that the crisis was quite dangerous, see Dmitry Dima Adamsky" "The 1983 Nuclear Crisis – Lessons for Deterrence Theory and Practice," *Journal of Strategic Studies*, Volume 36, Issue 1, 2013, pp. 4-41. For a convincing argument, based on Soviet documents, that the fears of a nuclear war in 1983 have been overblown, see Mark Kramer, "The Able Archer 83 Non-Crisis: Did Soviet Leaders Really Fear an Imminent Nuclear Attack in 1983?", unpublished paper.

[42] In addition to the 1958 to 1962 period, the winter of 1950-51 – after the People's Republic of China (PRC) intervened in the Korean War and nearly threw the Americans off the peninsula – was extraordinarily dangerous. The United States had lost the atomic monopoly, had not yet implemented the defense build-up called for in NSC-68, and Western Europe and Japan were still economically feeble, militarily impotent, and largely unprotected. American policymakers not only worried about a war, but felt that if one came they might lose it. These two superpower standoffs strike one as being in categories of their own, significantly more dangerous than any other listed in either data set.

[43] Thomas C. Schelling, *Arms and Influence,* (New Haven, CT: Yale University Press, 1966), p, 69-91.

[44] One of the only efforts to test the difference can be found in Walter J. Petersen "Deterrence and Compellence: A Critical Assessment of Conventional Wisdom," *International Studies Quarterly* (1986) 30, 269-294. Petersen's study suggests that compellence is *not* harder than immediate deterrence.

'compeller' and who is the deterrer – is often in the eye of the beholder. Was Khrushchev trying to compel the Western powers to leave Berlin or deter the United States from supporting West Germany's nuclearization? Or both? Did Stalin order the blockade of Berlin in 1948 to deter the creation of the Federal Republic of Germany, or was he trying to compel the West to abandon a policy – the 'western strategy' for Germany – which they had been pursuing since 1946? Or to look at a pre-nuclear case: in 1914 was Austria trying to *compel* Serbia to abandon its policy of creating a greater Serbian state at Austrian expense (and to compel Russia to abandon its policy of supporting Serbia in that area), or was it trying to *deter* Serbia from challenging the status quo?

There is a more important issue. James Fearon and others have pointed out that countries *select into* crises like the ones studied here, meaning their pre-existing beliefs about the balance of military power and resolve have already come into play in their decision to initiate and respond to a crisis.[45] To truly understand how important military balances or resolve are, you would not just analyze crisis *outcomes*; you would also need to include *the crises that never happened,* because a state calculated that either it was outgunned or didn't possess the requisite resolve to prevail. In other words, the question of resolve and the military balance has already come into play before a crisis is even initiated, so studying nuclear crises does not reveal the full story of whether and how military power plays affect world politics.[46] It is hard to imagine how you can effectively "control" for such a thing --

[45] James D. Fearon, "Domestic Political Audiences and the Escalation of International Disputes,"*The American Political Science Review* Vol. 88, No. 3 (Se, 1994), 577-592.

[46] See Fearon, 1994, 578. "Surprisingly, in the model, neither the balance of forces nor the balance of interests has any direct effect on the probability that one side rather than the other will back down once both sides have escalated. The reason is that in choosing initially whether to threaten or to resist a threat, rational leaders will take into account observable indices of relative power and interest in a way that tends to neutralize their impact if a crisis ensues. For example, a militarily weak state will choose to resist the demands of a stronger one only if it happens to be quite resolved on the issues in dispute and so is relatively willing to escalate despite its military inferiority. The argument implies that observable aspects of capabilities and interests should strongly influence *who gets what in international politics but that their impact should be seen more in uncontested positions and faits accomplis than in crises* (Italics mine). Which side backs down in a crisis should be determined by relative audience costs and by unobservable, privately known elements of states' capabilities and resolve." See also Fearon, 1994, 586. "Two of the most common informal claims about state behavior in international crises are that (1) the militarily weaker state is more likely to back down and (2) the side with fewer "intrinsic interests" at stake is more likely to back down. These arguments are problematic. If relative capabilities or interests can be assessed by leaders prior to a crisis and if they also determine the outcome, then we should not observe crises between rational opponents: if rational, the weaker or observably less interested state should simply concede the issues without offering public, costly resistance…. A second striking result from the equilibrium analysis is that observable measures of the balance of capabilities and balance of interests should be unrelated to the relative likelihood that one state or the other backs down in crises where both sides choose to escalate. Less formally, the result suggests that rational states will "select themselves" into crises on the basis of observable measures of relative capabilities and interests and will do so in a way that neutralizes any subsequent impact of these measures. Possessing military strength or a manifestly strong foreign policy interest does deter challenges, in the model. But if a challenge occurs nonetheless, the challenger has signaled that it is more strongly resolved than initially expected and so is no more or less likely to back down for the fact that it is militarily weaker or was initially thought less interested." I am grateful to Marc Trachtenberg for identifying and explaining the

who knows how many nuclear crises never happened because one side or the other was deterred from either initiating or responding to a provocation. It may very well have been the Soviet Union's sense of military inferiority or lack of resolve that caused it to stand by as the United States re-armed West Germany between 1952-1954, for example, a policy that was deeply threatening to Soviet Russia. It is very hard to undertake statistical analysis on events that have never happened, and perhaps the best we can to do is attempt to reconstruct the decision-making on both sides.[47]

The authors might respond that they are interested in the far more narrow explanation of crisis that outcomes have already been selected into. If so, then their causal explanations behind their theory must work; in other words, they cannot explain just the outcome, but must also explain *why* the outcome happened the way it did. Even if their coding were perfect, does either theory convincingly identify the causal mechanism that drives the origin, dynamics, and outcome of the standoff? As Marc Trachtenberg reminds us, "the world at the end of 1962 was very different from the world of November 1958."[48] Do the theories or causal mechanisms identified in either paper -- nuclear superiority or the taking of territory – help us understand the reasons for these important changes?

Looking at the documents, not so much. Damage limitation, the fear of unacceptable damage, and the role of territory may have played a role, pulling in different directions, but none of these factors was decisive or really tells us very much, in the same way that identifying the "aggressor: and the "defender" or coding who "won" the standoff provides little more than the most superficial view of these complex, consequential events. In the end, the 1958 to 1962 crisis was resolved through a *political* settlement that emerged by 1963 and that reflected the core interests of both superpowers.[49] This relates to the final point about this approach: the whole focus on military factors, divorced from their political contexts – not just in these papers but in most treatments of nuclear dynamics in security studies – can be misleading. There is little doubt that nuclear weapons have transformed world politics, and that military factors played a crucial role in the 1963 settlement. But it is a state's political goals and preferences that shape its military strategies, and not the other way around.

To better understand this point, think about a nuclear crisis that many expected but never happened -- the years leading up to the 1971 agreement on Berlin. The documents from the late 1960s make it clear that the Nixon administration feared a renewal of the Berlin crisis when it took office. None of the explicit, public issues Khrushchev demanded in 1958

---

meaning and importance of Fearon's arguments to understanding these issues.

[47] Eugene Gholz and Daryl Press make this recommendation to deal with selection effects in their paper, "Untangling Selection Effects in Studies of Coercion," unpublished working paper, 2006.

[48] Marc Trachtenberg, *History and Policy* (Princeton: Princeton University Press, 1991, 231.

[49] This is the core theme of Trachtenberg's *A Constructed Peace*.

had been resolved – U.S. British, and French troops were still in West Berlin – but the one thing that had changed was the military situation. The nuclear balance shifted from overwhelming U.S. superiority to near parity with the Soviets. Richard Nixon and National Security Advisor Henry Kissinger expected the Russians to exploit their improved military status, but to their great surprise, the Soviets were far more reasonable than anyone had anticipated and no crisis occurred.[50] In the end, the Soviets offered a settlement which West German Chancellor Willy Brandt pointed out was more favorable than what "was discussed in Geneva in 1959" or which Dean Acheson's Berlin report to President Kennedy hoped to achieve in 1961, periods when the U.S. possessed nuclear superiority.[51] Kissinger agreed: "I feel that we're doing better than, than I thought possible."[52] This success was possible, in spite of the fact, as Brandt pointed out, "that we all know the military position rather is more favorable for the Soviet Union than it was then."[53]

Why was there no nuclear crisis over Berlin in the late 1960s and early 1970s? One would have to look through both the Soviet and American (as well as various Western and Eastern European) documents to come up with a definitive answer, but a decent guess is that the fundamental geopolitical issue that drove the crisis in the first place – West Germany's nuclear status – had been resolved in ways pleasing to Soviet Russia's (and it terms out, everyone but West Germany's) leadership. In other words, neither the nuclear balance nor the actual territory was the key variables in determining whether there was a superpower crisis or not. Instead, it was one of the core geopolitical issues that dominated international politics in the decades after World War II, and which is not coded in either study – the nuclearization of the Federal Republic of Germany (FRG). Focusing exclusively on military factors can obscure the issues that really matter, the forces that drove the origins and outcomes of these clashes. We know this intuitively – we don't analyze the U.S. first strike advantages over Canada or Brazil, because they are not enemies. But these types models tell us very little about why this is so, or why Russia was an enemy and is one no longer, or why China was an enemy, then a friend, and now something in between, or why we worry more about an Iran with the bomb than a nuclear Sweden.

It turns that this whole set of issues surrounding both nuclear dynamics and methodology is a sequel of sorts; we've seen this movie before. In 1984, Paul Huth and Bruce Russet compiled a comprehensive list of what they called extended-immediate military deterrence situations between 1900 and 1980 to assess when deterrence worked,

[50] Gavin, *Nuclear Statecraft,* 72-73. Also see David C. Geyer, *Foreign Relations of the United States:* 1969–1976, Volume XL, Germany and Berlin, 1969–1972 (Washington, DC: US Government Printing Office, 2008).

[51] Conversation among President Nixon, German Chancellor Brandt, the President's Assistant for National Security Affairs, and the German State Secretary for Foreign, Defense, and German Policy, June 15, 1971, *FRUS, 1969-1976*, Volume XL, 742.

[52] Ibid., 744.

[53] Ibid., 743.

when it failed, and why. Using a statistical analysis, they determined this type of deterrence only worked in 31 cases (57% of the time), and that the keys to success were factors like close economic and political-military ties between the defender and its protégé, higher stakes by the defender in the protégé, and "local" (as opposed to overall) military capabilities. Surprisingly, nuclear possession was shown to be of only marginal importance.[54] Huth and Russett updated their data set in 1988, dropping some cases and adding others, and increasing the time frame to 1885 to 1984.[55]

These articles inspired a sharp critique from Richard Ned Lebow and Janice Gross Stein. Lebow and Stein questioned much of Huth and Russett's coding, arguing that in many cases they got the aggressor and the defender wrong, made mistakes about what was a success and what was a failure, and confused deterrence with compellence. They contended that the overwhelmingly majority of Huth and Russett's cases should not have even been in the data set, identifying only nine cases that met the appropriate criteria, a number too small to generate significant findings from statistical analysis. The authors went back and forth, questioning definitions, research design, scope conditions, and coding decisions.[56] Others eventually weighed in, including James Fearon, who focused less on the research design and empirical dispute and instead highlighted what he saw as "the inadequacy of rational deterrence theory," especially what he saw as the false/unhelpful distinction between general and immediate deterrence.[57] This helped popularize Fearon's important insights about selection effects.

The substantive and methodological problems that surfaced in this early debate are remarkably similar to the problems in Sechser, Furhmann, and Kroenig papers. This brings up a larger set of questions: because of the selection effects, the disputes over coding, the challenge of determining the unit of observation, the rarity of nuclear crises, and the difficulty of comparing across cases, is statistical analysis the most appropriate tool to understand nuclear dynamics? Both papers make claims to 'control' for a variety of factors, such as the interdependence of cases and the survivability of strategic forces. But can even the most sophisticated statistical controls provide greater insight to these questions than

---

[54] Paul Huth and Bruce Russett. "What Makes Deterrence Work? Cases from 1900 to 1980," *World Politics* 36, no. 4 (July 1984).

[55] Paul Huth and Bruce Russett. "Deterrence Failure and Crisis Escalation," *International Studies Quarterly* 32, no. 1 (March 1988).

[56] Richard Ned Lebow and Janice Gross Stein. "Deterrence: The Elusive Dependent Variable," *World Politics* 42, no. 3 (April 1990; Paul Huth and Bruce Russett, "Testing Deterrence Theory: Rigor Makes a Difference," *World Politics* 42, no. 4 (July 1990).

[57] James D. Fearon, "Selection Effects and Deterrence," *International Interaction* 28 (2002). See also Harvey, Frank "Rational Deterrence Theory Revisited: A Progress Report," *Canadian Journal of Political Science* 28, no. 2 (September 1995), and James D. Fearon "Signaling versus the Balance of Power and Interests: An Empirical Test of a Crisis Bargaining Model," *The Journal of Conflict Resolution* 38, no. 2 (June 1994).

other methods? These types of issues are precisely why studies of this type are so vexing to historians and policymakers.

This is the part of the review where I, the historian, should be making claims on behalf of qualitative work based on archival research to better understand nuclear dynamics. And there is a powerful case to be made. It may be that the only way to get *real* insight, to develop causal inferences about these kinds of critical issues, is to reconstruct the thoughts, decisions, interactions between, and reconsiderations of top decision-makers as they wrestled with these extraordinarily important questions when trying to make policy. There is, however, another important argument in favor this kind of historical work – opportunity costs. When Huth, Russett, Lebow and Stein first worked on these issues, archival research was much more difficult and there were far fewer documents available. Nuclear decision-making was a black box, and coding often involved, at best, educated guesses. Today, we are in the midst of a declassified document revolution, with archives around the world and organizations like the National Security Archive, the Cold War International History Project, and the Nuclear Proliferation International History Project providing access to millions of pages of previously unavailable material. The dirty secret is that with much of this material being made available online and in published volumes, an international relations scholar needn't leave her living room to see reams of extraordinary evidence that bear on the questions engaged in these articles, at far less effort and expense than anyone could have dreamed of when I started my Ph.D. in the early 1990s.

Let's say you wanted to really understand how nuclear dynamics worked in the critical 1958 to 1962 period, to explore the role that deterrence, compellence, brinksmanship and nuclear superiority played in the origins and outcomes of the standoff. What sort of evidence would be available to you? On the U.S. side, there are literally tens of thousands of pages of declassified documents available online about the Berlin and Cuban Missile Crises. Simply by using your laptop computer, you could access a treasure trove from the *Foreign*

*Relations of the United States* series,[58] the National Security Archive,[59] secretly taped Presidential Recordings,[60] and the Declassified Document Reference System.[61] You could gain great insight on the thinking of America's closest ally, Great Britain, by examining Prime Minister Harold McMillan's papers.[62] Records on the Soviet side are not as open, but there are important materials available, including secret Presidium meetings,[63] participant

---

[58] The best source for U.S. nuclear decision-making during the 1958 to 1962 period is the United States Department of State's *Foreign Relations of the United States*, many (including all cited here) of which are available online. http://history.state.gov/historicaldocuments See, among other volumes, Edward C. Keefer and David W. Mabon, eds, *Foreign Relations Of The United States, 1958–1960, National Security Policy; Arms Control And Disarmament, Volume III* (Washington, DC: United States Government Printing Office, 1996); *Foreign Relations Of The United States, 1958–1960, National Security Policy; Arms Control And Disarmament, Volume III, Microfiche Supplement; Foreign Relations Of The United States, 1958–1960, Berlin Crisis, 1958–1959, Volume VIII* (Washington, DC: United States Government Printing Office, 1993); David M. Baehler and Charles S. Sampson, eds. *Foreign Relations Of The United States, 1958–1960, Berlin Crisis, 1959–1960, Germany, Austria, Volume IX* (Washington, DC: United States Government Printing Office, 1993); Charles S. Sampson and John Michael Joyce, eds, *Foreign Relations Of The United States, 1961–1963, Volume V, Soviet Union* (Washington, DC: United States Government Printing Office, 1998); Charles S. Sampson, ed, *Foreign Relations Of The United States, 1961–1963, Volume VI, Kennedy-Khrushchev Exchanges* (Washington, DC: United States Government Printing Office, 1996); David W. Mabon and David S. Patterson, eds., *Foreign Relations Of The United States, 1961–1963, Volume VII, Arms Control And Disarmament* (Washington, DC: United States Government Printing Office, 1995); Edward C. Keefer, Charles S. Sampson, and Louis J. Smith, *Foreign Relations Of The United States, 1961–1963, Volume XI, Cuban Missile Crisis And Aftermath* (Washington, DC: United States Government Printing Office, 1996); Charles S. Sampson, *Foreign Relations Of The United States, 1961–1963, Volume XIV, Berlin Crisis, 1961–1962* (Washington, DC: United States Government Printing Office, 1993); Charles S. Sampson*, Foreign Relations Of The United States, 1961–1963, Volume XV, Berlin Crisis, 1962–1963* (Washington, DC: United States Government Printing Office Washington 1994).

[59] The National Security Archive has a microfiche collection devoted to this very subject: *The Berlin Crisis, 1958-1962,* described at http://www.gwu.edu/~nsarchiv/nsa/publications/berlin_crisis/berlin.html and available at most research libraries. They also have a searchable digital archive and a long list of online "briefing books", including almost seventy related to nuclear history http://www.gwu.edu/~nsarchiv/NSAEBB/index.html

[60] http://millercenter.org/academic/presidentialrecordings and especially http://millercenter.org/scripps/archive/presidentialrecordings/kennedy/1962/index

[61] http://gdc.gale.com/products/declassified-documents-reference-system/

[62] http://www.amdigital.co.uk/m-collections/collection/macmillan-cabinet-papers-1957-1963/

[63] Aleksandr Furksenko, ed. *Prezidium TsK KPSS, 1954-1964: Chernovye protokol'nye zapisi zasedanii stenogrammy [The Presidium of the Central Committee of the Communist Party of the Soviet Union, 1954-1964: Rough Stenographic Meeting Minutes]*. Moscow: Rosspen, 2003. Some of those transcripts, pertaining to the CMC, can be found in The Cold War International History Project, CWIHP *Bulletin*, Issue 17/18, "The Global Cuban Missile Crisis at 50," especially pp. 303-15, accessed at http://www.wilsoncenter.org/sites/default/files/CWHIP_Bulletin_17-18_Cuban_Missile_Crisis_v2_COMPLETE.pdf on September 8, 2013 My sincere thanks to Simon Miles for identifying these and other important Russian source materials.

recollections,[64] and secondary sources that used primary source materials.[65] Nor is pursuing this kind of work difficult or as overwhelming as it might appear for a political scientist: as Marc Trachtenberg points out "there is a method for reaching relatively solid conclusions about major historical issues in a reasonable amount of time – say, in about three or four months of uninterrupted work."[66] The raises an important question – should young scholars spend their time constructing a problematic data set and running regressions, or will our collective knowledge be better advanced through in-depth analyses of the truly relevant cases based on primary documents?

Of course, there are powerful disciplinary incentives that shape the research strategies of younger scholars and prevent them from making this choice. I am part of a discipline that has largely abandoned studying important issues such as international security and nuclear weapons and is in the midst of a four-decade, slow motion act of collective suicide. There simply is not, nor will there be anytime soon, a critical mass of diplomatic and military historians available to research these important questions or make use of these amazing materials. This is a national tragedy for which the field of history and our institutions of higher education should be ashamed and for which I fear that the United States will pay a price. The field of political science deserves great praise for taking up some of the slack; it is one of the few places within higher education where there is serious, sustained, collective interest and debate over crucial issues of national and international security. It is not, however, without its own pathologies. I can't imagine the flagship journals in the field such as *International Organization*, the *Journal of Conflict Resolution*, or the *American Political Science Revie*w would be interested in a deeply researched, multi-archival piece on the nuclear dynamics of the 1958-1962 period, though I would love to be proven wrong. From what I gather, brave – and from a career perspective, sadly, unwise -- is the Ph.D. student in international relations who undertakes a dissertation that does not include formal models, data sets, and multiple regressions. Employing quantitative tools, especially statistics, is highly rewarded within political science and often seen as being more 'rigorous' and scientific. Whether this is a good or a bad thing for political science more

[64] To give three important examples, two of which have been translated to English: Anatoly Dobrynin. *In Confidence: Moscow's Ambassador to America's Six Cold War Presidents*. (New York: Times Books, 1995); Nikita Khrushchev, Nikita. *Memoirs of Nikita Khrushchev*, vol. 3, *Statesman*. (University Park: Pennsylvania State University Press, 2007), and Georgii Kornienko, *Kholodnaia voina: Svidetel'stvo ee uchastnika [The Cold War: Testimony of a Participant]*. (Moscow: Olma, 2001).

[65] There are quite a few, but one can get very far with the following three excellent books: Alexander Fursenko and Timothy Naftali, *Khrushchev's Cold War*: *The Inside Story of an American Adversary* (New York: W.W. Norton, 2006); Hope Harrison, *Driving the Soviets up the Wall: Soviet-East German Relations, 1953-1961*. (Princeton, Princeton University Press, 2003); William Taubman, *Khrushchev: The Man, His Era*. (New York: Free Press, 2005).

[66] "I think political scientists, or at least those studying international relations, need to know how to do historical work. They may need to know other things as well, but it is hard to see how they can hope to understand international politics if they do not know how to do historical analysis in a fairly serious way." Marc Trachtenberg, *The Craft of International History: A Guide to Method*, (Princeton: Princeton University Press, 2006), xiii-ix.

broadly is not really for me to say. Statistics can be a powerful but blunt instrument.[67] Turning complex historical processes into quantifiable variables risks losing both information and even accuracy, which should affect our confidence in the findings.[68] This is especially true in the area of nuclear statecraft, where statistical analysis does not strike me as the best method for understanding complex, interactive political decision-making about issues of life and death where the most important 'N's' are 9, 2, and 0.[69]

In the end, however, the real issue is not about methodology per se. There are bound to be legitimate disputes about coding, just as there are arguments about historical interpretations. The authors could understandably disagree with my coding of 1958-1962, or argue that this case was exceptional and that other cases in the sample were more indicative of the underlying causal mechanisms that may operate in future cases. Their models could be refined to better deal with selection effects, more accurately handle the issue of resolve, or more effectively measure nuclear superiority. There is no shortage of problematic research designs among qualitative and formal scholars. Perhaps most importantly, the authors could point out that for all the availability of new primary materials, pursuing a 'historical' strategy appears no more likely to produce certain results than their quantitative efforts.

They would have a point. In theory, access to more primary materials should provide a more accurate picture of events, but we know that policymakers can misrepresent the past, mislead, or even provide contradictory views of the same meeting in written documents.[70]

---

[67] For an excellent "inside baseball" review of many of the problems that plague quantitative work in political science, see Christopher H. Achen, "Toward A New Political Methodology: Microfoundations and ART," *Annual Review of Political Science,* Vol. 5, pp. 42350. "Even at the most quantitative end of the profession, much contemporary empirical work has little long-term scientific value …. (t)he present state of the field is troubling. For all our hard work, we have yet to give most of our new statistical procedures legitimate theoretical micro- foundations, and we have had difficulty with the real task of quantitative work—the discovery of reliable empirical generalizations." Like my critique of the articles reviewed here, Achen fears many statistical models do not think hard enough about what is being compared in the analysis, or correctly identifying the universe of "like" events, and worries that the absence of any logical theory of how the variables interact prevents the correct econometric tools from being chosen, as opposed to canned models with lots of commonly used control variables.

[68] For an excellent guide to when it is appropriate to use qualitative or quantitative methods in international relations research, see Stephen Van Evera, "Director's Statement: Trends in Political Science and the Future of Security Studies," *Security Studies Program, Annual Report 2009-10,* 6-9, accessed at http://web.mit.edu/ssp/publications/annual/ssp_annual_report2010.pdf on September 8th, 2013.

[69] Furthermore, if I am right about the state of the field of history (and I truly hope I am not), it is hard to see how political scientists can accurately code these events if there is not a good historical literature to rely upon.

[70] For an excellent caution on the accuracy of participant memoirs, see Robert Jervis's H-Diplo review of Melvyn Leffler, "The Foreign Policies of the George W. Bush Administration: Memoirs, History and Legacy," *Diplomatic History* 37:2, accessed at http://h-diplo.org/reviews/PDF/AR405.pdf on July 8th, 2014. For a classic example of this, see Fred I. Greenstein and Richard H. Immerman, "What Did Eisenhower Tell Kennedy about Indochina? The Politics of Misperception," *The Journal of American History*, Vol. 79, No. 2 (Se, 1992),

Even if a deep immersion in documents produces historical accuracy, it often comes at the cost of the generalizations and policy insights about nuclear dynamics we all crave. Everywhere you look in the historical record there are puzzles, riddles and anomalies that seem to elude our best theories. Why did the Soviets initiate the Berlin and Cuban Missile Crises at a time of great nuclear inferiority, and prove so pliable over the same set of issues when they achieved parity? Why has China, with enormous economic resources, built rather modest nuclear forces, whereas a much poorer Pakistan seems hell-bent on deploying a massive force? Can anyone really speak with great confidence as to how a nuclear Iran would behave? Or consider the challenge of understanding President Nixon, a President who ratified the Nuclear Nonproliferation Treaty and negotiated the Anti-Ballistic Missile (ABM) and Strategic Arms Limitation (SALT) treaties with the Soviet Union, yet did not believe in arms control and yearned to reclaim the overwhelming, first-strike capability he believed President Kennedy had possessed.[71] "(I)n 1962, at the time of the Cuban missile crisis, it had been 'no contest,' because we had a ten to one superiority."[72] He also regularly engaged in what he believed was nuclear brinkmanship; as he told Kissinger, "we've got to play it recklessly. That's the safest course."[73] Yet it is not clear the Soviets ever picked up, understood, or reacted to any of his nuclear threats. How we generalize from this? How can anyone 'code' Richard Nixon's perplexing, sophisticated, and at times frightening thinking and policies on nuclear weapons?[74]

Perhaps the real issue is not what methodology we use to explore these issues, but rather, our comfort with uncertainty, our natural reluctance to embrace epistemological modesty on questions of such great importance. For a long time, we believed that the answer surrounding nuclear dynamics was relatively simple – once the superpowers achieved secure, second-strike capabilities, the possibility of thermonuclear war dissipated, and deterrence would prevail. Some even suggested that the stabilizing features of nuclear weapons were a positive feature, and that proliferation should not be seen as the end of the world.[75] While the reckless attitudes like Nixon's may have been anomalous, as more and

568-587.

[71] Francis J. Gavin, *Nuclear Statecraft*, chapter 5.

[72] Meeting Between President Nixon and Committee on Arms Control and Disarmament, March 21, 1972. Editorial Note. FRUS XIV, (Washington, DC: U.S. Government Printing Office, 2006 218.

[73] Ibid., Conversation between President Nixon and his Assistant for National Security (hereafter Kissinger), April 4, 1972, 258.

[74] In the context of selection effects, it would be interesting to know how and whether the military balance affects peaceful, if coerced, bargaining. Nixon was aware of and spoke about how the nuclear balance had changed dramatically and put him in a much worse situation than Kennedy. Does this mean that Nixon backed off more than Kennedy? I am grateful to Robert Jervis for this excellent insight.

[75] For my critique of this view, and the whole way the nuclear proliferation question has been handled in much of the strategy literature, see Francis J. Gavin, "Politics, History and the Ivory Tower-Policy Gap in the Nuclear Proliferation Debate," *Journal of Strategic Studies*, August 2012, 573-600. For Sagan and

more documents become available we began to fully recognize how dangerous and often unstable the nuclear age has been. Top policymakers never shared the comfort felt by the strategists over Mutual Assured Destruction (MAD). Consider Secretary of State Dean Rusk's reactions during the meeting where President Kennedy was told that long-predicted era of mutual vulnerability had finally arrived:

> General Johnson agreed, adding that nuclear war is impossible if rational men control governments.
>
> Secretary Rusk said he agreed, but he did not get much comfort from this fact because, if both sides believed that neither side would use nuclear weapons, one side or the other would be tempted to act in a way which would push the other side beyond its tolerance level. He added that a response to pressure might be suicidal, being prompted by a desire to get it over with. He referred to the current situation as "This God Damn poker game."[76]

Sechser, Furhmann, and Kroenig deserve credit for wrestling with these fundamental issues in a serious way. Both papers have important if obvious insights: at some intuitive level, it should not surprise us that deterrence is easier than compellence (even if it is difficult to code), or that a state may want more nuclear weapons than its adversaries. But getting beyond the obvious is difficult, and in the end, both papers overstate their theoretical claims and their policy relevance, and leave this historian no more confident on these questions than he was before. But can the historian provide anything more satisfying, more generalizable? While no doubt disappointing, at this point we can do little better than the assessment recently offered by Philip Zelikow, while reviewing my own work on the subject:

> "U.S. nuclear superiority mattered. And, at some level, it also didn't. At times both of these propositions were, at one and the same time, true. It is not easy to generalize much from this story about nuclear weapons except that they do matter, to those who have them and to those who don't."[77]

For those that would complain that such indeterminacy undermines the idea of a political "science," I would respond, guilty as charged. These two articles, it should be pointed out, are not the first occurrence of rival quantitative studies coming to starkly different

---

Waltz's response, see Scott D. Sagan and Kenneth N. Waltz, "Political Scientists and Historians in Search of the Bomb," *Journal of Strategic Studies*, Volume 36, Issue 1, 2013, 143-151.

[76] Summary Record of the 517th Meeting of the National Security Council1, Washington, September 12, 1963.

[77] Philip Zelikow, H-Diplo Roundtable Review, Francis J. Gavin. *Nuclear Statecraft: History and Strategy in America's Atomic Age*. Ithaca: Cornell University Press, 2012, http://h-diplo.org/roundtables/PDF/Roundtable-XV-1.pdf (Page 29), accessed September 8 2013.

conclusions about the dynamics of nuclear statecraft.[78]

The news is not all bad, of course. In the end, despite great fears, expectations, and apocalyptic predictions, we have never had a thermonuclear war, nor does it look like we will have one anytime soon. We *think* deterrence works. The problem is that, notwithstanding the confident claims of countless theorists, including those reviewed here, we don't really know why nuclear bombs have not been dropped since 1945, or at the very least, cannot prove our theories and instincts. Was it good statesmanship? Was Kenneth Waltz right, and nuclear weapons really are the great stabilizers?[79] Or perhaps it was just luck? We have a difficult enough time making sense of things that have actually occurred. Despite the labors of countless scholars from around the world, assessing millions of pages of documents, there is still no consensus on the causes of the First World War. The methodological challenge of trying to understand something that *never* happened, even something as important as thermonuclear war, is far, far more difficult. As is often said about the inadvisability of testing nuclear deterrence failures, we've never run the experiment, and hopefully never will.

This matters even more for policymakers than scholars: while we have theories, in the end it is close to impossible to predict how an Iran with nuclear weapons would behave, or whether it is worth the enormous consequences that would ensue from a military effort to prevent Iran's nuclearization, without trying to come to terms with these larger questions

[78] As Jacques Hymans notes, statistical efforts by Dong-Joon Jo and Erik Gartze to understand nuclear proliferation dynamics came produced different answers than the quantitative model of Sonali Singh and Christopher Way. While Hymans suggests both quantitative efforts are "sophisticated," the "differences in their results may reflect a more basic problem for quantitative attacks on this question: the lack of a reliable data set on which to base worldwide statistical tests ….(I)n light of this confusion about the basic historical facts, it may be premature to attempt large-N analyses of the proliferation phenomenon." Jacques E.C. Hymans, "The Study of Nuclear Proliferation and Nonproliferation: Toward a New Consensus?" in William C. Potter with Gaukhar Mukhatzhanova, *Nuclear Proliferation in the 21st Century: The Role of Theory* Vol. 1, (Stanford: Stanford University Press, 2010), 21-2. For an excellent critique of what has been called the second wave of quantitative studies on nuclear proliferation, see Alex H. Montgomery and Scott Sagan, "The Perils of Predicting Proliferation," *Journal of Conflict Resolution* April 2009 53: 302-28. "First, nuclear programs' initiation and completion dates are ambiguous and difficult to code, but findings are rarely subjected to sufficient robustness tests using alternative codings. Second, independent variables overlook important factors such as prestige and bureaucratic power and often use poor proxies for concepts such as the nonproliferation regime. Third, methodologies and data sets should be tightly coupled to empirical questions but are instead often chosen for convenience. Fourth, some findings provide insights already known or believed to be true. Fifth, findings can ignore or gloss over data crucial for policy making and wider debates." 302.

[79] For the first time he laid out this argument, see Kenneth Waltz, "The Spread of Nuclear Weapons: More May Better," *Adelphi Papers*, Number 171 (London: International Institute for Strategic Studies, 1981). For the last time, see Kenneth Waltz, "Why Iran Should Get the Bomb: Nuclear Balancing Would Mean Stability," *Foreign Affairs,* July/August 2012, accessed at http://www.foreignaffairs.com/articles/137731/kenneth-n-waltz/why-iran-should-get-the-bomb on September 8th, 2013.

surrounding nuclear dynamics.[80] This is why the questions that both papers engage are so important, even if the papers' theories are unconvincing, their methods problematic, and the policy implications unclear. We are understandably eager to have an explanation for the most important non-event in human history, if only to see if there are lessons that can be applied today to keep the streak going as long as possible. It is a daunting task, far more challenging that we like to acknowledge, and many of the research questions that we do focus on are merely proxies for this larger concern. For make no mistake about it: when we talk about any number of subjects surrounding nuclear weapons, such as why states do or not build the bomb, how they behave when they get it, or how and when deterrence works, the core question animating our curiosity is a powerful desire to understand why there has never been, and hopefully will never be, a thermonuclear war. Despite the limitations of these and other studies, despite the methodological difficulties and the near impossibility of being certain about our claims, it is hard to imagine a more important question, one that is worthy of the most vigorous research, discussion, and debate.

[80] Probabilistic models may good for getting a sense of the possible trends over the next fifty or one-hundred cases, but are of far less use to the decision makers trying to craft policies for the specific, highly consequential and likely complex issues in the n+1 case, especially if is unclear whether the theory/causal mechanism is correct or applicable to the case at hand. For a good exploration of this issue, see Philip Zelikow, "The Nature of History's Lessons," unpublished paper presented at Duke University's "History and Policy" conference, May 16th, 2013; for the challenges facing decision-makers facing radical uncertainty, with a particular focus on Iran's nuclear ambitions, see Francis J. Gavin and James B. Steinberg, "The Unknown Unknowns," *Foreign Policy*, February 14th, 2012; Francis J. Gavin and James B. Steinberg, "Mind the Gap: Why Policymakers and Scholars Ignore Each other, and What Can be Done About it?," *Carnegie Reporter,* Spring 2012, http://carnegie.org/publications/carnegie-reporter/single/view/article/item/308/ For an important effort to improve expert judgment to make better policy predictions, see Michael C. Horowitz and Philip E. Tetlock, "Trending Upwards: How the intelligence community can better see into the future," *Foreign Policy,* September 6, 2012, accessed at http://www.foreignpolicy.com/articles/2012/09/06/trending_upward on July 8th, 2013. For an excellent example of research that exploits international relations theory and history in order to provide policymakers with real-world policy guidance on nuclear questions, see Colin H. Kahl, Melissa Dalton, Matthew Irvine "Risk and Rivalry: Iran, Israel and the Bomb," Report, *Center for a New American Security*, June 2012, http://www.cnas.org/riskandrivalry; and Colin H. Kahl, Melissa Dalton, Matthew Irvine, "Atomic Kingdom: If Iran Builds the Bomb, Will Saudi Arabia be Next?" Report, *Center for a New American Security* February 2013, http://www.cnas.org/atomickingdom.

## Response: "The Case for Using Statistics to Study Nuclear Security,"[1]

by **Matthew Fuhrmann,** Texas A&M University, **Matthew Kroenig,** Georgetown University, and **Todd S. Sechser,** University of Virginia

In "What We Talk About When We Talk About Nuclear Weapons: A Review Essay," Francis J. Gavin reviews and evaluates the recent quantitative turn in nuclear security studies, singling out for special attention our recent articles in *International Organization*, "Nuclear Superiority and the Balance of Resolve: Explaining Nuclear Crisis Outcomes" and "Crisis Bargaining and Nuclear Blackmail."[2] The scope of Gavin's critique extends well beyond our articles, however, encompassing the broader scholarly literature employing quantitative analysis to understand questions of nuclear security. "Statistical analysis," he writes, "does not strike me as the best method for understanding complex, interactive political decision-making about issues of life and death." Instead, Gavin argues for a research agenda dominated by qualitative methods and analysis of archival evidence to better understand the effects of nuclear weapons on international politics.

Gavin's critique arrives at a pivotal time for the field of nuclear security. Scholars are increasingly turning to quantitative methods to answer questions about the political dynamics of nuclear weapons.[3] Given the importance of the subject matter, it is crucial that nuclear security scholars carefully scrutinize their research methods. If Gavin is correct, much intellectual firepower is being wasted – with potentially serious consequences.

[1] The authors' names are listed alphabetically; equal authorship is implied. The authors would like to thank Jason Gusdorf for helpful research assistance.

[2] Todd S. Sechser and Matthew Fuhrmann, "Crisis Bargaining and Nuclear Blackmail," *International Organization* Vol. 67, No. 1 (2013): 173-95, and Matthew Kroenig, Nuclear Superiority and the Balance of Resolve: Explaining Nuclear Crisis Outcomes," *International Organization* Vol. 67, No. 1 (2013): 141-71.

[3] Much of our own work falls along these lines. See, for example, Matthew Fuhrmann, *Atomic Assistance: How 'Atoms for Peace' Programs Cause Nuclear Insecurity* (Ithaca, NY: Cornell University Press, 2012); Matthew Kroenig, *Exporting the Bomb*, (Ithaca, NY: Cornell University Press, 2010); Matthew Fuhrmann and Todd S. Sechser, "Signaling Alliance Commitments: Hand-Tying and Sunk Costs in Extended Nuclear Deterrence," *American Journal of Political Science* (forthcoming); Matthew Kroenig, "Exporting the Bomb: Why States Provide Sensitive Nuclear Assistance," *American Political Science Review*, Vol. 103, No. 1 (2009): 113-33; Erik Gartzke and Matthew Kroenig, "A Strategic Approach to Nuclear Proliferation," *Journal of Conflict Resolution,* Vol. 53, No. 2 (2009): 151-160; Erik Gartzke and Matthew Kroenig, "Nuclear Posture, Nonproliferation Policy, and the Spread of Nuclear Weapons," *Journal of Conflict Resolution,* Vol. 58, No. 3 (2014); and Matthew Fuhrmann and Todd S. Sechser, "Nuclear Strategy, Nonproliferation, and the Causes of Foreign Nuclear Deployments," *Journal of Conflict Resolution,* Vol. 58, No. 3 (2014). Just a few of the many examples of quantitative methods in nuclear security studies include Vipin Narang, "What Does it Take to Deter? Regional Power Nuclear Postures and International Conflict," *Journal of Conflict Resolution*, Vol. 57, No. 3 (2013): 478-508; Mark S. Bell and Nicholas L. Miller, "Questioning the Effect of Nuclear Weapons on Conflict," *Journal of Conflict Resolution* (forthcoming); Erik Gartzke and Dong Joon Jo, "Bargaining, Nuclear Proliferation, and Interstate Disputes," *Journal of Conflict Resolution,* Vol. 53, No. 2 (2009): 209-33; Kyle Beardsley and Victory Asal, "Winning With the Bomb," *Journal of Conflict Resolution*, Vol. 53, No. 2 (2009): 278-301; and Michael Horowitz, "The Spread of Nuclear Weapons and International Conflict: Does Experience Matter?" *Journal of Conflict Resolution*, Vol. 53, No. 2 (2009): 234-57.

The issues Gavin raises are not unique to the field of nuclear security. The value of statistical analysis has been a topic of heated discussion for years in political science, sociology, anthropology, and other disciplines traditionally dominated by qualitative approaches. Beyond academia, quantitative analysis is being used today for corporate employee evaluation, the targeting of digital advertisements, political campaign management, online product recommendations, and many other areas of our lives. These debates have even seeped into professional sports: baseball and other sports have undergone an upheaval in recent years as quantitative methods have revolutionized the way teams think about strategy, scouting, and compensation.[4] Gavin's criticisms of these methods therefore have far-reaching implications, and deserve to be taken seriously.

We appreciate Gavin's engagement with our work, and it is a privilege for us to participate in this important roundtable. As regular readers of this roundtable series already know, Gavin's work on nuclear proliferation has helped reshape our understanding of the ways in which nuclear weapons shape world politics.[5] In our view, however, Gavin's critiques are badly misguided. In this essay, we defend quantitative analysis as an important tool for illuminating the complex effects of nuclear weapons. Contrary to Gavin's claims, statistical analysis has several useful attributes for studying nuclear security. Instead of revealing the flaws of quantitative analysis, Gavin's critiques suggest a misunderstanding of the nature and purpose of quantitative research, particularly as applied to nuclear security studies. Moreover, the alternative Gavin proposes would not solve the problems he highlights, and in some cases would exacerbate them. While the authors of this essay disagree among ourselves about many important substantive issues, we are united in the belief that statistical analysis has an important role to play in the field of nuclear security studies.

The rest of the essay will continue in four parts. First, we briefly describe our approach to studying nuclear issues, and contrast it with Gavin's method. Second, we explain the unique advantages offered by statistical analysis. Next, we discuss the limitations of this approach and the complementary strengths of alternative methods. Finally, we offer concluding remarks.

[4] See, for example, Michael Lewis, *Moneyball: The Art of Winning an Unfair Game* (New York: W.W. Norton, 2004); Tobias Moskowitz and L. Jon Wertheim, *Scorecasting: The Hidden Influences Behind How Sports Are Played and Games Are Won* (New York: Crown Archetype, 2011); Benjamin Baumer and Andrew Zimbalist, *The Sabermetric Revolution: Assessing the Growth of Analytics in Baseball* (Philadelphia: University of Pennsylvania Press, 2014).

[5] See, for example, Francis J. Gavin, *Nuclear Statecraft: History and Strategy in America's Atomic Age* (Ithaca, NY: Cornell University Press, 2012); Francis J. Gavin, "Blasts from the Past: Proliferation Lessons from the 1960s," International Security, Vol. 29, No. 3 (Winter 2004/2005): 100-35; and Francis J. Gavin, "Same As It Ever Was: Nuclear Alarmism, Proliferation, and the Cold War," *International Security*, Vol. 34, No. 3 (2010): 7-37.

We agree with Gavin on an important point: the study of nuclear weapons in world politics is important. Indeed, there are few issues of greater policy significance than the causes and effects of nuclear proliferation. All of the participants in this roundtable, therefore, want to better understand how nuclear weapons influence deterrence and coercion, and whether nuclear superiority provides countries with advantages in crisis bargaining. Yet we approach these issues from fundamentally different methodological perspectives.

Gavin's proposed approach to studying these issues is straightforward: he suggests that we first "identify the most important example where these issues are engaged, look at the primary documents," and then "reconstruct the thoughts, decisions, interactions between, and reconsiderations of top decision-makers" in order to determine whether nuclear weapons played a role in the crisis.[6] In other words, Gavin argues that the best way to understand the political effects of nuclear weapons is to probe a single case (or a small number of cases) deeply. Specifically, he points to the 1958–1962 superpower standoff over Berlin and Cuba as the "most important and representative case" for studying nuclear crisis behavior,[7] and returns to this episode repeatedly throughout the essay to support his arguments about nuclear deterrence and compellence.[8]

Our respective articles in *International Organization* adopt a considerably different approach. Whereas Gavin focuses his attention on a single episode, our articles compare many episodes: Kroenig's study evaluates 52 nuclear crisis participants, and Sechser and Fuhrmann analyze 210 coercive threats. For each study, we devise quantitative indicators for several key concepts – including nuclear capabilities, crisis actions, and several other factors. We then estimate statistical models to determine whether nuclear capabilities are reliably correlated with crisis outcomes, while controlling for other variables that could influence both a state's nuclear status and whether it prevails in a crisis.

To illuminate the differences between our approaches, consider how we might approach an important question that is unrelated to nuclear security: does smoking cause cancer? While smoking and nuclear crises are vastly different phenomena, they share important similarities from a research standpoint. Both cigarettes and nuclear weapons have potentially significant effects on large numbers of people, but the ways in which they operate are not always visible to the naked eye. Further, conducting direct experiments to assess their effects is infeasible, whether due to ethical or practical reasons. Both questions therefore require us to think carefully about how to distill causal effects from imperfect data.

[6] Gavin, "What We Talk About When We Talk About Nuclear Weapons," pp. 5, 18.

[7] Ibid., p. 6.

[8] This case is also a central subject of Gavin's recent book, *Nuclear Statecraft*.

Using the "most important example" approach discussed by Gavin,[9] the first step would be to locate the most important individual example of smoking and evaluate this person's life. One problem with this approach emerges immediately: it is not clear who the most important smoker might be – perhaps an individual with a high public profile, one with a great deal of wealth, or someone who smoked excessively. As we will discuss below, identifying the most important nuclear crisis is likewise fraught with difficulty. Assuming we could identify this "most important" case, we might then ask, for example, how much this person smoked each day, what diseases or health problems he acquired, his opinion about how smoking affected his life and health, and at what age he died. We would also scour medical records and the results of medical tests for information about this person's behavior and health over his lifetime. What did his doctors say about the effects of his smoking? Did they believe that smoking put him at a higher risk of getting cancer? What did this person himself believe? Do we have transcripts in which this person discusses his views on the health effects of his habit? With this trove of information, we could then formulate an understanding about smoking and cancer based on the written and verbal record of this person's health.

By contrast, our approach would begin by collecting data on a large number of individuals, including smokers and nonsmokers as well as people with and without cancer. Across this large group of subjects, we would analyze whether there was a strong correlation between the "independent variable" (smoking) and the "dependent variable" (cancer). Are there patterns between smoking and cancer that cannot be explained by chance alone? If we found that smoking was statistically associated with a higher incidence of cancer – after controlling for other factors that affect whether one smokes and whether they contract cancer, such as family history, occupation, and other behaviors – we would then conclude that smoking is correlated with, and may therefore be a cause of, cancer.

Which approach described above is more useful for understanding the causes and effects of smoking? What about the causes and effects of nuclear proliferation? Gavin's method is not without merit, but it suffers from major drawbacks that impede one from making reliable inferences about how nuclear weapons affect world politics. Our approach has important limitations as well, but we believe that it is a more powerful tool for providing answers to the questions we raise in our *International Organization* articles.

Using statistics to study nuclear security offers many advantages. The central goal in any social scientific study is *inference*: to observe a small slice of the world in order to learn how it works. Specifically, both of our studies in *International Organization* endeavor toward causal inference, aiming to learn about the causal effects of nuclear weapons in crisis situations.[10] How do nuclear weapons shape the way crises play out?

[9] We do not know for certain how Gavin would approach this question, of course, since he does not address the relationship between smoking and cancer in his essay. Our goal here is simply to discuss how his approach to studying nuclear security would apply in another context.

[10] King, Keohane, and Verba distinguish between descriptive inference, which uses observable information to learn about unobserved facts (such as inferring one's favorite sports team from one's city of residence), and causal inference, in which the goal is to ascertain causal effects. See Gary King, Robert O.

Are nuclear states more successful when employing deterrence and coercion? Are states with a nuclear advantage more likely to achieve their basic political goals in a crisis?

Which method offers clearer insights into the questions driving our studies? Below we describe four key advantages of employing a quantitative approach to evaluating the political effects of nuclear weapons.

First, quantitative approaches allow us to compare large numbers of observations. Undoubtedly, Gavin's approach to studying nuclear weapons is the better method for learning about the events of the 1958–1962 period. But does this single episode tell us all we need to know about nuclear weapons? Our *International Organization* articles ask broader questions about how nuclear weapons have impacted world politics in the half-century before and after this period – and how they might do so in the future. Claiming that the Berlin/Cuba episode is sufficient for answering these questions presumes that the decades before and since the Cuban missile crisis have little to teach about the consequences of nuclear weapons – a claim we find specious at best.

The main limitation of Gavin's approach to assessing the effects of nuclear weapons is that it provides us with few, if any, generalizable inferences – the central aim of social science. In other words, his approach tells us much about the Cuba and Berlin crises, and rather little about nuclear weapons more broadly.[11] Did the patterns that were evident from 1958–1962 continue throughout the Cold War, or were they anomalous? Should we expect that these patterns will hold into the future? Gavin's approach offers no answer because it unduly limits its scope to a single historical episode, rather than examining how that episode fits into broader trends.

The questions we ask in our articles require a more comprehensive approach to data collection. By collecting information about dozens (or hundreds) of cases rather than just one or two, we can gain insights into whether the patterns we observe in any individual case are representative of broader trends. The implicit question in our research is always 'what would have happened if conditions had been different?' Of course, it is impossible to answer this counterfactual with certainty since history happens only once, and we cannot repeat the 'experiment' in a laboratory. But that does not mean we should shrug our shoulders and abandon the enterprise.

Instead, we can gain insight by looking at cases in which conditions were, in fact, different. To illustrate, let's return to the smoking example above. Studying a single smoker in depth might give us an accurate and textured understanding of the role of smoking in this

Keohane, and Sidney Verba, *Designing Social Inquiry: Scientific Inference in Qualitative Research* (Princeton: Princeton University Press, 1994).

[11] Indeed, Gavin admits as much when he concedes that "Even if a deep immersion in documents produces historical accuracy, it often comes at the cost of the generalizations and policy insights about nuclear dynamics we all crave." See Gavin, "What We Talk About When We Talk About Nuclear Weapons," 21.

person's life, but it would be a poor way to learn about the broader health effects of smoking, because we could not make an informed guess about what would have happened had he *not* smoked. Our approach described earlier, in contrast, allows us to generalize about the effects of smoking on health. For precisely this reason, large-scale quantitative analysis is the primary method by which medical researchers have tackled the health effects of tobacco smoke. To be sure, some of the data in our hypothetical study would surely be inaccurate, and we would know comparatively little about the lives of each individual subject. But the loss in individual case knowledge would be more than compensated by the increase in information about the variables we hope to study.

So it is with nuclear weapons. To understand how nuclear weapons impact international crises, we must examine crises in which nuclear 'conditions' were different. For Kroenig, this means comparing the fortunes of crisis participants that enjoyed nuclear superiority to those that did not. For Sechser and Fuhrmann, it means comparing the effectiveness of coercive threats made by nuclear states to those made by nonnuclear states. By making these comparisons, we can begin to engage in informed and evidence-based speculation about how nuclear weapons change (or do not change) crisis dynamics. Indeed, the statistical models we employ require this comparison – they will return no results if all of our cases look the same.

Gavin argues that the Berlin/Cuba episode is sufficient for understanding the dynamics of nuclear weapons because it is the "most important and representative" case of nuclear deterrence and coercion.[12] There are two distinct (and contradictory) claims here: that the case is the most important crisis episode for studying nuclear weapons, and that it is representative of the broader universe of such episodes. With respect to the first claim, Gavin offers no criteria for evaluating what an "important" case might be. What makes a case important – its profile among the general public? Its consequences? The availability of information about it? The countries involved? Moreover, for whom must the case be important? Gavin may view the 1958–1962 case as critical for understanding nuclear dynamics, but it is by no means clear that policymakers today look to this example for guidance about dealing with Iran or North Korea. This is not to say that we disagree with Gavin's assessment – undoubtedly the 1958–1962 episode is important in many respects. But importance, like beauty, is in the eye of the beholder.

The second claim is equally dubious: that the 1958–1962 episode is somehow representative of the ways in which nuclear weapons typically shape international politics. Without first examining other cases, Gavin simply has no grounds on which to base this claim. Moreover, there is tension between this claim and his previous assertion that the case is important: one key reason the Cuba/Berlin episode is often seen as important is because it was *not* like other Cold War crises: nuclear weapons were brandished more explicitly, and stoked more public anxiety about nuclear war, than any other crisis before or since. In the broader universe of crises, this episode actually may be quite anomalous. If

[12] Ibid., 6.

so, then studying it to the exclusion of other cases would yield misleading conclusions about the role of nuclear weapons in world politics.

A key advantage of quantitative methods is that the researcher need not make questionable judgments about which cases are more or less important: unless explicitly instructed otherwise, statistical models assign equal weight to each case. Likewise, statistical models provide ways to identify – and exclude – anomalous cases that deviate markedly from dominant trends. Indeed, a quantitative analysis can be a useful precursor to the selection of individual cases for in-depth analysis, precisely because it allows us to locate cases that either represent or deviate from the overall pattern. These selections, however, are based on careful comparisons with other cases, not opaque judgments.

A second advantage is that quantitative analyses provide greater transparence about methods, judgments, and conclusions. One of Gavin's central critiques is that various cases in our quantitative analyses have been miscoded. In other words, he argues, we have mismeasured important factors.[13] This criticism – irrespective of its validity[14] – is possible only because our coding decisions are unambiguous and easily ascertained from our datasets. Moreover, each of our studies sets forth clear rules for how each variable in our datasets was coded. This does not mean that our coding decisions are all correct and beyond dispute, but it does mean that they are clearly stated for outside scholars to evaluate. This degree of transparency is a key strength of quantitative research. Because each case in a quantitative analysis necessarily must be clearly coded,[15] there is no ambiguity about how the researcher has classified each case. If other researchers believe a case should be coded differently, they can make that change and rerun the analysis.

By extension, quantitative research designs permit scholars to easily evaluate how much a study's findings depend on individual coding decisions. Simply noting a few coding errors or differences of interpretation in a large quantitative dataset is of little consequence unless one can demonstrate that those differences are responsible for generating incorrect inferences. In a quantitative study, this typically amounts to recoding disputed cases and repeating the core statistical models to determine whether the results change substantially.[16] Not only are the original coding decisions laid bare, but it is also

[13] This is also a central pillar of Sagan's critique of the quantitative turn in nuclear security studies. See Scott D. Sagan, "The Causes of Nuclear Weapons Proliferation," *Annual Review of Political Science*, Vol. 14 (2011): 225-44.

[14] In our individual responses to Gavin's essay, we address the particulars of these coding disagreements.

[15] In the quantitative models we use, cases are dropped from the statistical analysis if even one variable is not coded.

[16] For example of a case where such a reassessment yields substantially different findings, see Alexander B. Downes and Todd S. Sechser, "The Illusion of Democratic Credibility," *International Organization*, Vol. 66, No. 3 (2012): 457-89.

straightforward to determine whether the study's inferences depend on them. This high level of transparency – and the external quality-control it enables – is one of the most attractive features of quantitative research designs. Transparency is useful not because it produces scholarly consensus, but because it allows opposing sides to identify the precise nature and implications of their disagreements.

Consider, for example, the 1990 exchange in *World Politics* between Paul Huth and Bruce Russett on one hand, and Richard Ned Lebow and Janice Gross Stein on the other.[17] Gavin highlights the similarities between this debate and the present exchange, separated by almost twenty-five years, as evidence that quantitative analysis has made little progress in understanding nuclear issues. We see the issue differently. Both debates, in fact, illustrate a key strength of quantitative analysis: the ability to assess the importance of individual coding decisions. In the *World Politics* debate, Lebow and Stein objected that Huth and Russett had improperly coded many cases in their deterrence dataset, much as Gavin has disputed some of our classifications But Huth and Russett responded by noting that "even if Lebow and Stein's recodings of our cases are accepted, the statistical and substantive findings of our past research remain fundamentally unchanged."[18] Similarly, as we report in our articles, our central findings do not change even if we accept Gavin's arguments. In a quantitative study, simply showing that certain coding decisions can be contested is insufficient: one must also demonstrate that the core results depend on those decisions. While Gavin is correct to argue that coding cases is a tricky exercise, quantitative approaches allow us to evaluate the substantive importance of questionable coding decisions.

Qualitative research, by contrast, is not always so amenable to external oversight. Whereas quantitative models demand clear coding decisions, qualitative research designs can be much more forgiving of ambiguous classifications. Gavin's critique of our coding decisions illustrates this problem: while he criticizes the way we have coded particular cases in our datasets, he offers no clear alternative coding scheme. He raises questions about our coding decisions, but then declines to answer them. This ambiguity allows him to have his cake and eat it too: he can criticize our classifications without being liable for his own.

Uncertainty, of course, is inherent to any scientific enterprise, and quantification is sometimes criticized for presenting a false illusion of certainty. To be clear, quantitative research cannot create certainty where the evidence is ambiguous. Just because a case is coded a certain way does not mean that the broader scholarly community (or even the researcher) has reached a consensus about that case. Likewise, the problem of ambiguity is not inherent to qualitative research: nothing intrinsic to historical research precludes

[17] Richard Ned Lebow and Janice Gross Stein, "Deterrence: The Elusive Dependent Variable," *World Politics*, Vol. 42, No. 3 (1990): 336-69; and Paul Huth and Bruce Russett, "Testing Deterrence Theory: Rigor Makes a Difference," *World Politics*, Vol. 42, No. 4 (1990): 466-501.

[18] Thanks to Erik Gartzke for pointing this out. See Huth and Russett, "Testing Deterrence Theory," 468.

scholars from laying their assumptions bare. But by compelling scholars to take a clear initial position on coding cases, the process of quantification allows scholars to debate each decision and evaluate whether potentially questionable choices are decisive in generating a study's core results. This transparency is central to peer evaluation and, ultimately, scientific advancement.

A third advantage of statistical analysis is that it is designed to cope with probabilistic events. In the physical world, causal relationships are often deterministic: a certain amount of force imparted to an object will cause that object to move a certain distance. So long as conditions are kept constant, this result will obtain again and again, no matter how many times the experiment is repeated. In the social world, however, we are not blessed with such ironclad reliability. No two individual people are exactly identical, and even in carefully controlled environments it is rare to find a "force" that begets exactly the same effect on all people with perfect regularity. The causal relationships we observe are not deterministic – they are *probabilistic*, occurring with imperfect regularity.[19]

The 'force' of interest to us in our articles is, broadly, the possession of nuclear weapons. When this force is applied to crisis bargaining situations, what happens? Implicit in this question, however, is a question about probability: when nuclear weapons are inserted into a crisis bargaining situation, what is the *likelihood* of a particular outcome? Kroenig's study, for example, asks: in a nuclear crisis, what is the likelihood that the nuclear-superior side will achieve its basic goals? Likewise, Sechser and Fuhrmann seek to discover the likelihood that a coercive demand made by a nuclear-armed state will be met.

The central difficulty with posing our research questions in this way is that we cannot actually see the thing we care about: probability is inherently unobservable. We cannot examine a crisis and directly observe the probability of one side capitulating; we can only observe whether it actually capitulated.[20] How, then, can we begin to answer our original research question?

Quantitative research is designed for precisely this sort of situation. If we cannot directly

[19] King, Keohane, and Verba note that there are two possibilities here: either the social world truly is probabilistic, or it is actually deterministic and simply appears probabilistic to us because our explanations are imperfect. These two possible worlds, they note, are indistinguishable to our eyes – in other words, probabilistic theories are equally useful for describing phenomena in either one. See King, Keohane, and Verba, *Designing Social Inquiry*, pp. 59-60.

[20] Participants in crises sometimes offer their own probability assessments. President John F. Kennedy famously estimated the odds of nuclear war during the Cuban missile crisis as "somewhere between one in three and even." See Theodore C. Sorensen, *Kennedy* (New York: Harper & Row, 1965), 795. Yet there is no reason to believe that participants, operating in the heat of the moment and with limited information, would have any better access to underlying probabilities than a disinterested analyst conducting an assessment after the fact. Further, participants often disagree sharply: Kennedy's National Security Adviser, McGeorge Bundy, estimated the risk of war to be just one in 100 (though "still much too large for comfort"). See McGeorge Bundy, *Danger and Survival: Choices About the Bomb in the First Fifty Years* (New York: Random House, 1988), 461.

observe whether we are holding a loaded six-sided die, for example, we can throw it many times, observe the result, and infer the underlying probability from the results. Throwing the die just one time would tell us little, since all six numbers are theoretically possible even if the die were loaded. Only after observing the pattern of results across many events can we determine the underlying probabilities of each number turning up.

The single-case approach Gavin proposes cannot cope with probabilistic events as effectively. Knowing that one smoker happened to die of cancer does not tell us much about the broader health effects of tobacco. Based on this single data point, we might conclude that smoking leads to cancer 100 percent of the time. Yet we know this to be false: there are heavy smokers who remain cancer-free, just as there are nonsmokers who still get cancer. The true relationship between smoking and cancer emerges only after looking at a large number of cases. Similarly, even if we determine that nuclear weapons appeared to "matter" from 1958-1962, we cannot safely infer from this observation that nuclear weapons influence crisis outcomes *in general*. Any relationships observed during this particular period could have been due to any number of chance events that might be unlikely to recur. Studying just one episode allows us to say much about that episode but little about the underlying relationships.

Fourth, statistical analysis allows researchers to uncover causal relationships in social phenomena even if the participants themselves do not record, record accurately, or understand these relationships. Gavin's approach, in contrast, requires finding primary source documents and learning what participants themselves believed to be the relevant causal factors at play. His essay conveys an exceptionally narrow conception of how one should gather knowledge about the effect of nuclear weapons on international politics. Gavin believes that if one wants to "really understand" the effect of nuclear weapons on international politics,[21] archival research is "the only way to get *real* insight."[22] While we agree that studying primary documents has great value, we believe that there are many other ways to generate useful knowledge, and that a narrow focus on primary documents can often lead a scholar astray.

A focus on primary documents alone has several problems. First, the historical record is often incomplete, but the absence of evidence is not evidence of absence. In any historical nuclear crisis, a leader might have made an important statement to his advisers about how nuclear weapons fundamentally affected a crucial decision, but the statement might have never been recorded or could have been lost in the sands of time. Without that crucial piece of information, one might conclude, incorrectly, that nuclear weapons were not part of the calculation. Second, and related, something might not appear in the historical record because it was taken for granted. Researchers might then conclude that this factor was not salient, when in fact it might have been so important that it was well understood by everyone in the room and did not need to be mentioned. For example, nuclear weapons

[21] Gavin, "What We Talk About When We Talk About Nuclear Weapons," 18.

[22] Ibid., 16. Italics in original.

may have been so irrelevant (or central) to a particular crisis that leaders didn't even see the need to raise the issue. Third, individuals sometimes intend to mislead. Leaders may make inaccurate or incomplete statements – even in private – in order to influence the outcome of an internal debate, to improve domestic political fortunes, to shape how they will be viewed by posterity, or a variety of other reasons. Fourth, something might not appear in historical documents because participants themselves were unaware of how important it was. A case study of a smoker in the 1920s might not turn up any evidence from the smoker's medical records that smoking was damaging to his health. But, it was damaging nonetheless. Similarly, nuclear weapons may have had an influence on crisis dynamics even if leaders themselves didn't fully appreciate it.

Statistical analysis, on the other hand, does not depend on the participants themselves to understand and record accurately the causal forces at work. Rather, the researcher can independently identify and measure the variables of interest and search for correlations in large samples of data.

In short, factors may matter regardless of whether the participants record them, record them accurately, or even understand their existence. Statistical analysis can reveal these connections. The documentary record might not.

Statistical analysis can help us understand world politics, but it is not without limitations. Like any method, large-*N* analysis has some potential pitfalls with which scholars must grapple. In this section we review some of the key limitations of this approach and describe how we cope with them in our studies.

Our goal as social scientists is to understand causal relationships between variables of interest. However, it is difficult to determine causality in studies that use observational data, like ours. Why is this the case? When we find a strong correlation between variables A and B in our observations of the real world, this does not necessarily mean that A "causes" B. If a researcher found, for example, a positive and statistically significant correlation between ice cream consumption and drowning, this would not necessarily imply that eating ice cream causes people to be more susceptible to drowning. Rather, it is more likely that a third variable – temperature – was responsible, since in warm weather people are more likely both to go swimming – which increases the risk of drowning – and eat ice cream.

There is a similar risk of drawing spurious relationships when studying the political effects of nuclear weapons. Imagine that one discovered a positive correlation between nuclear weapons and military conflict. This would not necessarily imply that nuclear weapons cause states to behave more belligerently. Nuclear proliferation occurs as a result of a strategic process: countries possess atomic bombs only after governments decide that it is in their interest to build them. The factors causing states to build the bomb – rather than the weapons themselves – might account for the observed relationship between bomb possession and conflict. In particular, states in threatening security environments may be more likely to seek atomic bombs to protect themselves. In this case, the dangerous security environment is causing the state to both build nuclear weapons and become embroiled in a large number of disputes.[23] What can be done to address this issue?

The best way to deal with it, from the perspective of causal inference, would be to create an experimental world in which no state possessed nuclear weapons, artificially manufacture dozens of crises, and then randomly assign nuclear weapons to some states. We could then compare whether states that received bombs (the treatment group) were more conflict-prone than countries that did not get weapons (the control group). If they were, we could reasonably conclude that nuclear weapons caused countries to become more belligerent, since a state's nuclear status emerged by chance – not as a result of a strategic process. Of course, this approach is not only impossible, but also dangerous. We therefore must rely on other solutions to address barriers to inference that arise in observational studies.[24]

The most straightforward way to reduce the risk that an observed relationship is spurious is to control for the factors that might be correlated with both the independent and dependent variable. Our studies account for several covariates – including conventional military power – that could affect crisis bargaining and the nuclear status of the crisis participants. However, in some cases, the relevant factors for which one needs to control may not be observable or measurable. Political scientists are increasingly turning to 'high-tech' tactics for addressing concerns about causality.[25] Yet there are 'low-tech' solutions too that may be equally (and sometimes more) effective. Developing sound theory can go a long way in dealing with this issue. If an analyst has a logically consistent argument that can plausibly explain an observed correlation, she might have more liberty to make causal claims about that relationship. This would be especially true if she could use theory to rule out alternative explanations that could account for the observed pattern. In addition, supplementing statistical findings with case studies that unpack causal mechanisms can be an especially fruitful strategy.[26]

---

[23] See Bell and Miller, "Questioning the Effect of Nuclear Weapons on Conflict."

[24] Scholars may, however, conduct experiments in which they manipulate the treatment. For example, Daryl Press, Scott Sagan, and Benjamin Valentino recently conducted a survey experiment to assess why people oppose the first use of nuclear weapons. See Daryl Press, Scott D. Sagan, and Benjamin Valentino, "Atomic Aversion: Experimental Evidence on Taboos, Traditions, and the Non-Use of Nuclear Weapons," *American Political Science Review,* Vol. 107, No. 1 (2013): 187-206. A natural drawback of this method, of course, is limited external validity; insights from the laboratory do not always translate to the real world.

[25] For example, matching analysis is increasingly used in international relations research. Through this analysis, scholars construct samples in which the 'treatment' and 'control' groups are as similar as possible before estimating their statistical models. See, for example, Yonatan Lupu, "The Informative Power of Treaty Commitment: Using the Spatial Model to Address Selection Effects," *American Journal of Political Science,* Vol. 57, No. 4 (2013), 912-925.

[26] This is a strategy that we have pursued in our own work. See Kroenig, *Exporting the Bomb;* Fuhrmann, *Atomic Assistance;* and Todd S. Sechser, "Goliath's Curse: Coercive Threats and Asymmetric Power," *International Organization,* Vol. 64, No. 4 (2010), 627-660.

It is important to note that challenges related to causal inference are hardly unique to our methodological approach. Establishing causation in the social sciences is notoriously difficult, and it is a problem that plagues not just statistics, but many research methods. Gavin's preferred method is particularly ill equipped for assessing causality. It is exceedingly difficult to know whether nuclear superiority *caused* the United States to win crises from 1958 to 1962 by following his approach. What we really want to know, if our goal is to make causal claims, is whether the outcomes of the crises during that period would have been different had the Soviets possessed strategic superiority instead. Of course, there is no way to rerun history, as we noted earlier in this essay. Yet one could get at this indirectly by comparing the 1958–1962 period to later crises in which the United States had nuclear inferiority but which were similar to the Berlin/Cuba cases in other respects. Gavin does not explicitly recommend this type of comparative analysis (although his discussion of the 1971 agreement on Berlin implicitly underscores the importance of controlled comparisons). If he had, however, our central point here would remain the same: conducting qualitative analysis does not exonerate scholars from wrestling with the issue of causality.

Readers might respond that Gavin's approach is not designed to make causal claims, and that one's ability to make such inferences using any method is exceedingly limited. At least one statement in Gavin's essay implies that he accepts this view. As he writes, "For those that would complain that such indeterminacy undermines the idea of a political 'science,' I would respond, guilty as charged."[27] This is a perspective that we do not share, and neither do the vast majority of political scientists – including many who rely exclusively on qualitative methods.[28]

A second limitation of statistical analysis has to do with measurement: it is sometimes difficult to quantify concepts of interest. Prestige, for example, is widely believed to be important in world politics. Indeed, according to the prevailing wisdom, one reason states seek nuclear weapons is to increase their standing in the international system. Yet it is tough to assign a numerical value to prestige. We can devise variables to proxy for this concept – for instance, one recent study codes status seeking behavior based on performance at the Olympics[29] – but we cannot measure it directly. It is therefore difficult to test theories about prestige using statistical analysis. Measurement challenges likewise accompany other concepts in international relations: reputation, influence, soft power, etc.

Gavin zeroes in on this limitation when critiquing our research. He argues that seemingly straightforward phenomena, like who won a crisis, can be difficult to code. Consider the

[27] Gavin, "What We Talk About When We Talk About Nuclear Weapons," 25.

[28] See, for example, Stephen Van Evera, *Guide to Methods for Students of Political Science* (Ithaca, NY: Cornell University Press, 1997); and Alexander L. George and Andrew Bennett, *Case Studies and Theory Development in the Social Sciences* (Cambridge, Mass.: MIT Press, 2005).

[29] Bryan R. Early, "Exploring the Final Frontier: An Empirical Analysis of Global Civil Space Proliferation, *International Studies Quarterly* (forthcoming).

Cuban Missile Crisis. Kroenig codes this case as a crisis victory for the United States and a defeat for the Soviet Union. Sechser and Fuhrmann similarly code the U.S. demand to withdraw the missiles as successful. However, Gavin argues that this coding is overly simplistic. Although the Soviets withdrew the nuclear weapons from Cuba in response to U.S. pressure, he suggests that both sides ultimately got what they wanted. The Soviets "won" because they preserved the Castro regime and forced the United States to remove Jupiter missiles from Turkey. Yet this nuance is lost in our quantitative analysis, according to Gavin.

We agree with Gavin that measurement issues can create challenges for scholars employing quantitative analysis. Yet qualitative research has to deal with these problems too. Let's say, for the sake of illustration, that one wanted to design a qualitative study to assess whether nuclear superiority mattered in the Cuban Missile Crisis. Reliance on qualitative methods does not free an analyst from dealing with potentially thorny measurement issues. One would still need to know a minimum of two things: (1) which side won the Cuban missile crisis and (2) which side had nuclear superiority. It is impossible to conduct a proper test of the argument in the absence of this information. Any social scientific inquiry requires analysts to measure their key variables, whether they employ qualitative or quantitative analysis. The measurement problems that Gavin identifies undoubtedly exist to some degree in our studies, but they are by no means unique to our studies or to statistical analysis more generally.

Gavin also overstates the magnitude of these problems for our research when he implies that they impede our ability to glean any meaningful insights. To be sure, in some cases, who won a crisis may be ambiguous. But in many instances, which state was victorious is clearer than Gavin suggests. In 1979, for example, a crisis erupted after supporters of the Iranian revolution took 52 hostages at the American embassy in Tehran. The Iranian government refused to return the hostages even after the United States threatened it with military force, making this case an unambiguous failure for U.S. coercive diplomacy.

We also find Gavin's interpretation of the outcome of the Cuban Missile Crisis difficult to sustain. We would not conclude that the Denver Broncos "won" Super Bowl XLVIII – a game they lost badly to the Seattle Seahawks – because their appearance in the game increased Broncos merchandise sales. The Broncos undoubtedly benefited in many ways from their Super Bowl appearance, but this hardly means that they won the game. Similarly, even if the Soviets got something out of the Cuban missile crisis, they failed to obtain their main objective – namely, keeping the missiles in Cuba. Our respective codings of this case reflect the most direct interpretation of the crisis' outcome.

In some cases scholars who employ statistical analysis can cope with the measurement issues that Gavin identifies. If, for example, there is disagreement about the way a certain case is coded, scholars can simply recode it and see if the results change. To the extent that the findings are consistent regardless of how certain cases are coded, one can have greater confidence in the robustness of the results. As we explain above and below, our findings in these studies are robust to such changes.

A third potential concern is that statistical findings may be sensitive, especially when studying rare events. Students of international relations who employ statistical analysis face a key challenge: many of the events in which they are interested occur infrequently. This is particularly true in the area of nuclear proliferation, given that only 10 countries have built nuclear weapons. The rarity of nuclear proliferation does not mean that statistical analysis is useless for understanding how nuclear weapons affect world politics, as Gavin implies. It does mean, however, that scholars should exercise appropriate caution when using this tool. When dealing with rare events, there is always the possibility that statistical findings are driven by a small number of cases.

It is worth noting that the phenomena of interest in our studies – crises involving nuclear states – are not particularly rare. Kroenig's study evaluates the performance of 52 nuclear crisis participants, and Sechser and Fuhrmann analyze 210 coercive threat episodes. Crisis participants achieved their basic goals in 35 percent of Kroenig's cases, and in 30 percent of the cases evaluated by Sechser and Fuhrmann. Half of all crisis participants, by Kroenig's definition, enjoyed nuclear superiority, while nuclear-armed challengers issued 20 percent of the coercive threats in Sechser and Fuhrmann's study. In neither of our studies were the critical numbers "9, 2, and 0," as Gavin suggests.[30] In short, the rare-events problem is far less pernicious in our articles than Gavin implies.

Still, given that our samples are relatively small, the patterns we identified may be sensitive to minor changes in the statistical models. One way to deal with to this issue is to conduct additional tests that are designed to evaluate the strength of a given finding, a solution similar to the one proposed above for addressing measurement problems. We employ this strategy in our articles – we rerun our models after recoding the dependent variables, modifying how we measure nuclear weapons and nuclear superiority, dropping potentially influential cases, and excluding particular control variables that could potentially bias our results. As we report in our respective studies, our main results survive these further tests. This does not guarantee that our findings are bulletproof, but it should inspire greater confidence that the patterns we identified are reasonably robust.

A fourth limitation is that statistical analysis is not designed to explain outliers. It is instead intended to identify average effects. Statistical analysis can tell us, on average, how an independent variable relates to a dependent variable. This is useful because it helps scholars to determine whether their theories are generalizable to a broad set of cases. Yet broad patterns may not apply to any single case. Most statistical models in political science produce outliers – cases that are not explained by the theory being tested. The 1999 Kargil War, for instance, is an outlier for the democratic peace theory. Although democracies are in general unlikely to fight militarized conflicts with one another, India and Pakistan nevertheless fought this particular war when they were both democratic states. But because most international relations theories are based on probabilistic logic (as opposed to deterministic logic), the presence of a few outliers does not necessarily disprove them.

[30] Gavin, "What We Talk About When We Talk About Nuclear Weapons," 21.

It is often helpful to explicitly identify outliers, something that Kroenig (Table 1) and Sechser and Fuhrmann (Table 3) do in their articles. Scholars can then study these cases to refine their theories or, at the very least, identify the conditions under which their arguments hold. This type of 'nested analysis' – selecting cases for study based on results from a statistical model – is increasingly employed in international relations research to improve the explanatory power of theories.[31] Identifying when outliers occur can be important for policy purposes. If we understand when nuclear powers were able to make successful threats in the past, for example, we can better understand whether Iran's ability to blackmail its adversaries might change if it builds the bomb.

In an attempt to illustrate the folly of studying nuclear weapons with statistics, Gavin begins his article with an excerpt from a farcical story by Josh Freedman about a man breaking up with his girlfriend, Susan, as the result of "a series of quantitative calculations." As the heartbreaker explains, "We can say we love each other all we want, but I just can't trust it without the data." The implication seems to be that certain things, like love and nuclear security, are simply too mysterious to understand using numbers.

Yet, in recent years, psychologists and sociologists interested in patterns of human courtship and mating have made major advances in their understanding of love and relationships through the employment of statistical analysis.[32] Newly-available data from online dating sites (which themselves use quantitative data to power their pairing algorithms) have allowed scholars to test and refine their theories, and to develop new ones, in ways never before possible. For instance, the *New York Times* reports that scholars are now able to put a quantitative estimate on the value women put on a man's professional success: "men reporting incomes in excess of $250,000 received 156 percent more e-mail messages than those with incomes below $50,000."[33] And this line of research has uncovered startling new findings, such as that white Americans, despite stated beliefs in racial equality, are especially unlikely to marry outside of their race.[34] These are preferences that few people would admit to, or even be consciously aware of, if subjected to direct interviews by researchers, yet they become glaringly obvious in large patterns of data.

There are, then, similarities between love and nuclear security: statistical analysis can contribute to a fuller understanding of both. Like the study of human courtship, nuclear

---

[31] See, for example, Fuhrmann, *Atomic Assistance*, pp. 110-42. For more on nested analysis see Evan Lieberman, "Nested Analysis as a Mixed-Method Strategy for Comparative Research," *American Political Science Review,* Vol. 99, No. 3 (2005): 435-52.

[32] See, for example, the recent book by Ty Tashiro, *The Science of Happily Ever After*: *What Really Matters in the Quest for Enduring Love* (Ontario: Harlequin, 2014).

[33] Hal R. Varian, "Online Dating? Thin and Rich Works Here, Too," *New York Times* (June 30, 2005).

[34] Stephanie Rosenbloom, "Love, Lies and What They Learned," *New York Times* (November 12, 2011), ST1.

security has undergone a renaissance in recent years, which has only been made possible by newly available datasets and a new generation of scholars applying statistical tools to address age-old questions. As we argue, these methods are advantageous for their ability to compare large numbers of observations, transparently reveal the researcher's methods and judgments, deal with probabilistic phenomena, and uncover relationships about which participants themselves might not be cognizant.

Nevertheless, while we believe statistical analysis has important strengths, it cannot be the only instrument in our toolkit. Even with statistics, it is difficult to ascertain causation, key concepts may defy easy measurement, available data may be limited, and outliers will demand explanation. Yet, these drawbacks do not mean that statistical analysis should be abandoned altogether. Qualitative methods suffer from some of the same problems, and many others as well.

For decades, scholars have been employing the methods proposed by Gavin to study the role of nuclear weapons in international politics. Scholars have written dozens, if not hundreds, of studies assessing the consequences of nuclear weapons for deterrence and coercion. And yet, this approach has not lived up to its promise. Despite the availability of millions of pages of archival documents, our understanding of these issues remains inadequate. Indeed, Gavin's own research (as he acknowledges) yields no conclusive answers about the coercive effects of nuclear weapons, and little useful policy advice, boiling down to the underwhelming finding that "U.S. nuclear superiority mattered. And, at some level, it also didn't."[35] Faced with this situation, it would be unwise to respond by continuing to tackle questions of nuclear security in exactly the same way as before, hoping that the next trove of documents will settle the debate once and for all. Rather, we should consider the possibility that an incomplete methodological toolkit has been obscuring our view.

**Matthew Fuhrmann** is an Assistant Professor of Political Science at Texas A&M University. He has previously held research fellowships at Harvard University (2007-08) and the Council on Foreign Relations (2010-11). He is the author of *Atomic Assistance: How "Atoms for Peace" Programs Cause Nuclear Insecurity* (Cornell University Press, 2012) and the co-editor of *The Nuclear Renaissance and International Security* (Stanford University Press, 2013). His work has been published or conditionally accepted in peer reviewed journals such as *American Journal of Political Science*, *British Journal of Political Science*, *International Organization*, *International Security*, *Journal of Conflict Resolution*, *Journal of Peace Research,* and *Journal of Politics*.

**Matthew Kroenig** is an Associate Professor and International Relations Field Chair in the Department of Government at Georgetown University and a Nonresident Senior Fellow in

[35] Philip Zelikow, *H-Diplo Roundtable Reviews 15(1)*: 2013, 29. https://www.h-diplo.org/roundtables/PDF/Roundtable-XV-1.pdf. Gavin approvingly quotes this passage in his essay, leading us to wonder why he asks whether our work "resolve[s] the decades-long debate over these long-contested issues in nuclear dynamics" when he does not appear to hold his own to the same standard.

the Brent Scowcroft Center on International Security at The Atlantic Council. He is the author or editor of several books, including *A Time to Attack: The Looming Iranian Nuclear Threat* (forthcoming 2014) and *Exporting the Bomb: Technology Transfer and the Spread of Nuclear Weapons* (2010), which won the International Studies Association Best Book Award, Honorable Mention. His articles on international politics have appeared in such publications as *American Political Science Review*, *Foreign Affairs*, *Foreign Policy*, *International Organization*, *The New York Times*, *The Wall Street Journal*, *The Washington Post*, and *USA Today*. From May 2010 to May 2011, he served as a Special Advisor in the Office of the Secretary of Defense on a Council on Foreign Relations International Affairs Fellowship, where he worked on defense policy and strategy for Iran.

**Todd S. Sechser** is an Assistant Professor of Politics at the University of Virginia. His research interests include coercive diplomacy, reputations in international relations, the strategic effects of nuclear weapons, and the sources and consequences of military doctrine. His work has been published in peer-reviewed journals such as the *American Journal of Political Science*, *International Organization*, *International Studies Quarterly*, and *the Journal of Conflict Resolution*, and his commentary on international affairs has appeared in the *Wall Street Journal*, *Christian Science Monitor*, *Boston Globe*, and other outlets. He has held research fellowships at Stanford University and Harvard University, and from 2011-12 was a Stanton Nuclear Security Fellow at the Council on Foreign Relations. He is also a member of the Program on Strategic Stability Evaluation.

## Response: "Nuclear Weapons Are (Still) Poor Instruments of Blackmail: A Reply to Francis J. Gavin's Critique"

by **Todd S. Sechser,** University of Virginia and **Matthew Fuhrmann**, Texas A&M University

We thank Francis J. Gavin for his insightful review of our article, "Crisis Bargaining and Nuclear Blackmail."[1] Gavin's essay offers several thoughtful criticisms of our study, and of the use of quantitative methods in nuclear security studies more broadly. In our joint response with Matthew Kroenig ("The Case for Using Statistics to Study Nuclear Security"), we explained why Gavin's critiques of our methodological approach are well off the mark. Here we turn to his specific criticisms of our article and its findings. Gavin's arguments are interesting and worthy of serious consideration, but they fail to undermine our paper's central conclusions.

Our article explores how nuclear weapons influence international crisis bargaining. When states attempt to coerce (or 'compel') their adversaries into changing their behavior or relinquishing possessions, do nuclear weapons provide them with an advantage? Or are nuclear weapons useful primarily as tools of deterrence?[2] We found a surprising answer: nuclear states generally are no more successful at coercing their adversaries than other states, despite the enormous destructive potential of their nuclear capabilities. Why is this the case?

Military technologies provide coercive leverage to the extent that they: (1) increase a state's ability to seize disputed items militarily, or (2) bolster a state's capacity to punish its adversary at an acceptable cost. Most of the time, nuclear weapons meet neither of these criteria.

On the first count, using nuclear weapons to seize territory or other disputed possessions would likely be counterproductive: doing so could damage or destroy the very item that the challenger hoped to obtain. It would make little sense, for example, for China to wrest the Senkaku Islands from Japan by launching a nuclear attack against the disputed territory.

Alternatively (and more likely), China might attempt to gain control over the islands by threatening to destroy Tokyo. Yet this brings us to a second point: nuclear weapons are high-cost tools of punishment. Carrying out a coercive nuclear threat would be exceedingly costly for the challenger. It would result in international blowback and establish a dangerous precedent, among other costs. Challengers might pay these costs if their vital interests were on the line. But our review of the historical record indicates that coercive threats rarely involve stakes that are high enough for the challenger to justify paying such a stiff price. Coercive nuclear threats therefore will usually be dismissed as incredible.

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

[2] The distinction between deterrence and compellence was articulated in Thomas C. Schelling's *Arms and Influence* (New Haven: Yale University Press, 1966).

Deterrence, however, is a different story. The costs of implementing nuclear deterrent threats would be comparatively low, because such threats are carried out only after an attack against a country or its ally. And the stakes are typically higher – indeed, the defender's national survival may be at stake. Deterrent nuclear threats, then, are more likely to be credible. Many observers found it believable, for instance, that the United States would have used nuclear weapons in the event of a Soviet invasion of Western Europe during the Cold War. By contrast, it would have been far more difficult to have persuaded anyone that the United States was willing to launch a nuclear attack for the central purpose of forcing the Soviets out of East Berlin in the 1950s.

Our argument stands in stark contrast to the emerging view that nuclear weapons (and nuclear superiority) are useful for military blackmail.[3] Nuclear weapons allow states to bully and intimidate their adversaries, according to this line of thinking, because they are so destructive. Nonnuclear (or nuclear-inferior) countries are therefore less resolved during crises with nuclear-armed (or nuclear-superior) opponents: weaker countries would rather bow to the wishes of their superior opponents than face the possibility of a catastrophic nuclear attack. A key tenet of this view is that nuclear weapons loom in the background, even when their use is not explicitly threatened. In this view, nuclear powers need not play dangerous games of brinkmanship to get their way in world politics; simply possessing the bomb is often sufficient to do the trick.

Does the historical record support our argument? To find out, we analyzed a database of 210 militarized compellent threats.[4] The dataset contains well-known episodes, like the 1962 Cuban Missile Crisis, in addition to less studied (but not necessarily less important) cases of coercive diplomacy, such as the 1969 Sino-Soviet border crisis, the 1975 USS *Mayaguez* incident, the 2001-2002 standoff following the terrorist attacks on India's parliament, and many others. Our central objective was to compare whether nuclear challengers enjoy a higher rate of coercive success than their nonnuclear counterparts. We found that there was very little difference in the success rates of nuclear and nonnuclear challengers. Nuclear weapons, in other words, appear to provide states with little coercive leverage.

Gavin is not persuaded by our study: he calls our theory "unconvincing." Yet it is not clear that he disagrees with us. Indeed, he provides no evidence suggesting that our

[3] For instance, Bruce Bueno de Mesquita and William H. Riker, "An Assessment of the Merits of Selective Nuclear Proliferation," *Journal of Conflict Resolution* 26 (2): 1982, 283-306; Robert A. Pape, *Bombing to Win: Air Power and Coercion in War* (Ithaca, NY: Cornell University Press, 1996); Kyle Beardsley and Victor Asal, "Winning with the Bomb," *Journal of Conflict Resolution* 53 (2): 2009, 278-301; and Matthew Kroenig, "Nuclear Superiority and the Balance of Resolve: Explaining Nuclear Crisis Outcomes," *International Organization* 67 (1): 2013, 141-71.

[4] Todd S. Sechser, "Militarized Compellent Threats, 1918-2001," *Conflict Management and Peace Science* 28 (4): 2011, 377-401.

argument or evidence is incorrect.[5] However, he offers several critiques of the scope and assumptions of our study. Here we respond to four specific criticisms: (1) our evidence does not help explain the 1958-1962 superpower standoff over Berlin and Cuba; (2) we study minor cases that are unimportant; (3) the distinction between deterrence and coercion is sometimes blurry; and (4) we do not properly deal with selection effects. We consider each of these points in turn, and explain why none of them pose a major challenge to the findings we presented in our paper.

Gavin's main complaint about our article is that it does not help him better understand the superpower standoff from 1958 to 1962. This is a puzzling criticism, since this was not the purpose of our article. Our objective was to investigate whether nuclear weapons help states make more effective coercive threats. The crises that occurred from 1958 to 1962 are undoubtedly relevant for understanding this question, but they are by no means the only cases that offer insights into it. Indeed, our article documents numerous cases in which the coercive diplomacy of nuclear states failed: the 1979 Sino-Vietnamese War, the 1982 Falklands crisis, the 1990-1991 Persian Gulf crisis, and others. Gavin does not challenge us on those cases, nor does he explain why the 1958-1962 episode is a more important lens through which to view our question. In short, he offers no reason to doubt our argument about the coercive limitations of nuclear weapons.

Gavin's criticism is equally puzzling because he offers no reason to believe that our theory cannot account for the 1958-1962 period. As Gavin's own work has shown, there is considerable doubt whether nuclear superiority was responsible for the American coercive success during the Cuban Missile Crisis.[6] The outcomes of the Berlin crises that preceded the Soviet missile deployment to Cuba are also consistent with our theory: Khrushchev was unable to use the (implicit) threat of nuclear war to force American troops out of West Berlin, despite repeated efforts to do so.[7] Contrary to Gavin's suggestions, the 1958-1962 episode actually squares quite well with our argument.

Gavin's second criticism of our article is that we examine the wrong cases. In his view, it is

[5] Instead, Gavin's most forceful critiques address our methodological approach, which we discuss in our joint response with Kroenig.

[6] Francis J. Gavin, *Nuclear Statecraft*: *History and Strategy in America's Atomic Age* (Ithaca, NY: Cornell University Press, 2012), 119. Gavin also notes in his essay that the outcome of the Cuban missile crisis could also be described as a Soviet victory. If so, then the crisis is even more consistent with our theory, since it would constitute a lesser coercive success for nuclear states than it might initially seem.

[7] Khrushchev faced a difficult challenge: as the coercer, he had to credibly threaten the first-use of nuclear weapons. For reasons explained in our article (and above), this is exceedingly difficult to do. By contrast, the United States merely had to convey that it would retaliate if it were attacked first. Khrushchev therefore faced an inherent bargaining disadvantage, irrespective of the nuclear balance. See Alexandr Fursenko and Timothy Naftali, *Khrushchev's Cold War: The Inside Story of an American Adversary* (New York: W.W. Norton & Company, 2006), 209. In the end, many officials in Washington dismissed Khrushchev's threats as bluster. One might surmise that Khrushchev failed because he was up against a nuclear-superior United States. In our view, however, this was not a decisive factor.

uninformative to study episodes in which nuclear war was not an obvious danger. He wonders, "Should every disagreement involving a country with the bomb be coded as a nuclear standoff? Why would we automatically assume that the nuclear balance is front and center when leaders of two nuclear states clash in some form?"[8]

In fact, we assume no such thing. Quite the contrary: the central claim of our study is that most coercive threats – even those made by nuclear states – do *not* raise the specter of nuclear war. But Gavin is wrong to assume that these episodes have nothing to teach about whether nuclear weapons convey coercive leverage. Indeed, expanding our scope beyond nuclear standoffs is crucial to understanding the coercive effects of nuclear weapons.[9]

First, nonnuclear crises provide a critical basis for comparison. We cannot know whether nuclear states achieve better coercive outcomes than their nonnuclear counterparts unless we study both groups. Imagine, for example, that we wish to find out whether a particular sports drink improves athletic performance. Studying only people who consume the drink would take us only halfway. We would also need to know how people fare *without* the sports drink so that we can compare the two groups. A similar logic holds in our study: only by comparing nuclear and nonnuclear coercers can we determine whether one group enjoys greater bargaining leverage. Evaluating crises in which there was seemingly little danger of nuclear war is central to this task.

Second, studying lower-profile crises helps avoid stacking the deck in favor of our argument. High-profile nuclear confrontations like the Berlin crisis tend to be considered historically important precisely because they nearly led to war. These crises, however, also tend to be crises in which coercive threats were not terribly successful: had threats been more effective in these crises, they would have been resolved long before they escalated to dangerous levels. Studying only high-risk cases therefore might overlook the most successful coercive threats, and bias the evidence in our favor.

Third, nuclear weapons may carry coercive leverage without ever being invoked. Indeed, some scholars have argued that nuclear weapons cast a shadow even in crises that we might consider nonnuclear.[10] Although we disagree with this claim, we cannot fairly evaluate it without looking at a broad spectrum of crises in order to determine whether nuclear-armed states achieve systematically higher rates of success. If our theoretical argument is wrong and nuclear weapons are, in fact, useful for coercion, examining only dangerous nuclear standoffs might not give us a complete picture of their utility.

[8] Gavin, "What We Talk About When We Talk About Nuclear Weapons," 13.

[9] We elaborate on this issue in Sechser and Fuhrmann, "Crisis Bargaining and Nuclear Blackmail," 179-80.

[10] Gavin himself hints at this possibility. See Gavin, "What We Talk About When We Talk About Nuclear Weapons," 22, note 74.

In short, lower-risk crises are a central part of the story: not only do they help us place the most dangerous nuclear crises in their proper context, but they also help ensure that we have not inadvertently skewed the data in favor of our own theoretical argument.

Our main claim in "Crisis Bargaining and Nuclear Blackmail" is that nuclear weapons may be useful for deterrence, but they are poor instruments of compellence. Gavin, however, questions the distinction between these two concepts, arguing that in the 1958-1962 standoff between the United States and the Soviet Union, both sides saw themselves as protecting the status quo, while casting the other as the aggressor. "Defining the status quo," he argues, "is often in the eye of the beholder."[11]

The 'status quo,' of course, is the core of the distinction between deterrence and compellence: deterrence aims to maintain the status quo, whereas compellence seeks to modify it.[12] We agree with Gavin that states rarely see themselves as aggressors: actions that may appear aggressive to outsiders are often viewed as defensive by the leaders who order them. But it does not follow that we cannot objectively identify the status quo in historical cases.

Our study employs a database of compellent threats that defines the status quo as objectively as possible. The rules of the dataset define a compellent threat as "an explicit demand by one state (the challenger) that another state (the target) alter the status quo in some material way."[13] In other words, our operational definition of "compellence" does not require that the two sides agree on who was the aggressor and who was the defender; the question is whether the challenger's demand required the target to modify the material status quo in some way – even if that status quo was only recently established. If so, then the case is included in the dataset; if not, it is excluded. This is faithful to Thomas C. Schelling's original conceptualization of compellence, which held that compellent threats are unique because they require the target to act – to "do something" – in order to comply.[14] Gavin may disagree with using a material baseline to identify compellent attempts, but if so, he does not explain why.

Gavin further argues that leaders often have both deterrent and compellent objectives during crises: in the Berlin episode, for example, he wonders: "Was Khrushchev trying to compel the Western powers to leave Berlin or deter the United States from supporting West Germany's nuclearization?"[15] Our answer would be: both. Khrushchev failed in his

[11] Ibid., pp. 7, 14. For an argument along similar lines, see Richard Ned Lebow and Janice Gross Stein, "Deterrence: The Elusive Dependent Variable," *World Politics* 42 (3): 1990, 336-69.

[12] See Schelling, *Arms and Influence*, 69-91.

[13] Sechser, "Militarized Compellent Threats," 380.

[14] Schelling, *Arms and Influence*, 69.

[15] Gavin, "What We Talk About When We Talk About Nuclear Weapons," 14.

first objective, and succeeded in the second, just as our theory expects. Deterrence and compellence are logical opposites, but they are not necessarily mutually exclusive.

Coding issues aside, one might nevertheless argue that the distinction between deterrence and compellence is unhelpful for understanding how nuclear weapons affect crisis bargaining. It might be preferable, according to this line of thinking, to analyze all crises together, regardless of whether a state is trying to preserve the status quo or change it. The problem with this approach is that it glosses over the central prediction of our theory: that nuclear weapons are poor instruments of compellence, but potentially useful for deterrence.

Moreover, drawing the distinction between deterrence and compellence is useful for policy purposes. How should policymakers respond to potential proliferators, like Iran? Scholars may not believe that it is important to distinguish between attempts to wrest away territory and attempts to defend it, but policymakers certainly do. If a nuclear arsenal allows states to grab territory with greater ease, reverse unfavorable policies, or extract other concessions, the prospect of nuclear proliferation is quite threatening. In that case, aggressive nonproliferation policies – including, potentially, the use of military force – might be justified. However, if nuclear weapons serve primarily defensive functions, the consequences of proliferation would be less severe. To be sure, allowing another state to build the bomb might still be undesirable. Yet the argument for aggressive nonproliferation policies weakens considerably if nuclear weapons deter but don't compel.[16]

Gavin's final criticism is that our study provides only partial insight into the dynamics of nuclear coercion. The reason, he explains, is that "the question of resolve and the military balance has already come into play before a crisis is even initiated, so studying nuclear crises does not reveal the full story of whether and how military power plays affect world politics."[17] In other words, one also would need to study "the crises that never happened" in order to glean a complete picture of how nuclear weapons shape the dynamics of coercion. Studying only coercive threats, as we do in "Crisis Bargaining and Nuclear Blackmail," tells only part of the story.

Gavin highlights a thorny and well-known problem in international relations scholarship: the problem of so-called "selection effects."[18] The basic principle is that in observational studies (i.e., studies in which we cannot conduct experiments), the phenomena we observe

[16] The theory and evidence we presented in our article support the less alarmist view of nuclear proliferation. Sechser and Fuhrmann, "Crisis Bargaining and Nuclear Blackmail," 192-93; and Matthew Fuhrmann and Todd S. Sechser, "Would A Nuclear-Armed Iran Really Be So Dangerous?" *Christian Science Monitor* (12 January 2011).

[17] Gavin, "What We Talk About When We Talk About Nuclear Weapons," 14.

[18] See especially James D. Fearon, "Signaling Versus the Balance of Power and Interests An Empirical Test of a Crisis Bargaining Model," *Journal of Conflict Resolution* 38 (2): 1994, 236-69; and James D. Fearon, "Selection Effects and Deterrence," *International Interactions* 28 (1): 2002: 5-29.

are not random events: they are the result of strategic decisions and explicit choices. Those decisions, however, could skew our conclusions by concealing important pieces of the puzzle. In the sports drink example above, assessing the drink's effect on athletic performance could be complicated by the possibility that individuals who choose to consume the drink (i.e., "select themselves" into that category) probably are more likely to exercise, and thus be in better athletic condition. If we simply compared the performance of people who regularly purchase the drink against those who do not, we would likely find a sizable difference – but not necessarily because the drink itself has had any effect.

Gavin is undoubtedly correct that researchers must be conscious of potential selection effects to avoid drawing hasty inferences. He does not, however, offer any reason to believe that selection effects are responsible for the findings we report in our article. All observational studies – including the archival studies he touts – are subject to selection effects, but this does not by itself undermine their validity. One must also explain why that selection effect has produced an incorrect inference. Gavin has not done this.

In fact, our article explicitly considers the possibility that self-selection dynamics might be driving the results.[19] Specifically, we evaluate whether nuclear states might be achieving coercive "victories" without ever having to issue threats. Using several methods of varying technical complexity, we find little evidence to suggest that this is the case. Selection effects undoubtedly exist in the cases we examine, but they do not appear to explain the patterns that we have uncovered.

Our article addresses a central question in world politics: what political dividends might nuclear weapons bring to their owners? An emerging view holds that nuclear weapons have coercive utility. Nuclear weapons, according to this perspective, help countries compel changes to the status quo in ways that serve their political interests. Our paper challenges this emerging consensus. We argue that nuclear weapons are poor tools of blackmail, in part, because it is difficult to make coercive nuclear threats credible. Our analysis of 210 compellent threats provides support for our theory: we find that nuclear states are not more effective at coercive diplomacy, on average, than their nonnuclear counterparts.

Gavin's essay, in our view, does not undermine our central conclusions. Although we find his criticisms unpersuasive, it is important to note that there is much about which we agree. Most importantly, we share the view that questions relating to the causes and effects of nuclear proliferation are tremendously important, and worthy of dedicated study. We also share his view that historical analysis is an indispensable tool for understanding the political effects of nuclear weapons, especially when combined with statistical analysis. Our statistical tests were meant to be our first word on this subject, not our last. Since the publication of our article, we have been researching the role of nuclear weapons in dozens

[19] See the section titled "Selection Effects," in Sechser and Fuhrmann, "Crisis Bargaining and Nuclear Blackmail," 187-89.

of Cold War and post-Cold War crises for precisely the reasons that Gavin identifies.[20] A thorough analysis of these cases, we believe, only strengthens support for our argument that nuclear weapons are poor tools of coercion and intimidation, despite (and in some cases because of) their destructive power.

We thank Gavin again for his engagement with our work, and we hope that this exchange encourages further thinking and research about the role of nuclear weapons in world politics, and about the broader methodological issues raised in these essays.

[20] Our analysis of these cases will be included in our book with the working title *Nuclear Weapons and Coercive Diplomacy*.

## **"A Superior Theory of Superiority," Response** by **Matthew Kroenig**, Georgetown University

In his article, "What We Talk About When We Talk About Nuclear Weapons: A Review Essay," Frank Gavin critiques the recent quantitative turn in nuclear security studies and devotes much of his attention to my recent article "Nuclear Superiority and the Balance of Resolve: Explaining Nuclear Crisis Outcomes." Matthew Fuhrmann, Todd Sechser, and I address many of Gavin's broader methodological and disciplinary criticisms in our joint response. In this essay, I briefly address the specific concerns raised about my article.

Gavin makes three major criticisms of my piece. First, he claims that there are problems with my attempt to theorize and measure my key explanatory variable, nuclear superiority. Second, Gavin argues that I don't effectively account for selection into nuclear crises. Third, he makes other smaller criticisms of my analysis, which lead him to question my results. Gavin is one of the country's leading scholars on nuclear issues and I am grateful for his deep engagement with my work. In this particular instance, however, I believe that all three of his criticisms lack merit.

First, Gavin argues that there are problems with my conception of nuclear superiority. He claims that I measure which state has a nuclear advantage, but do not account for whether the superior state has a "splendid" first strike capability or whether the inferior state has a secure, second-strike capability. According to Gavin, therefore, I ignore the "most consequential aspect of the question" which is, "whether either side believed it possessed a robust enough capability to launch a first strike and escape the ensuing response with an acceptable level of damage."

This criticism is misguided for two reasons. First, as I clearly explain in the article, I do account for second-strike capabilities. Second, and most importantly, it is Gavin, not me, who ignores "the most consequential aspect of the question." By reverting back to the hoary notion that nuclear superiority only matters if one state has a splendid first strike capability, Gavin repeats an outdated and overly simplistic notion of nuclear deterrence theory that has caused far too much confusion in the past. He also appears not to have fully grasped my central argument. My article explains why nuclear superiority matters even when both states possess secure-second strike capabilities. Indeed, this is the article's central theoretical contribution.

For decades, international relations scholars were faced with a puzzle: if nuclear deterrence theorists were correct and nuclear superiority does not matter once both states possess secure, second-strike capabilities, then why did the United States and the Soviet Union expend so much effort during the Cold War striving for a nuclear advantage? Were these strategies illogical, as many scholars charged? Or did scholars not fully appreciate the true nature of nuclear deterrence? I believe my article helps to solve this puzzle by clearly articulating the strategic advantage provided by nuclear superiority even in a situation of Mutual Assured Destructin (MAD). Gavin focuses his attention on my article's evidence, but I believe its biggest contribution is theoretical.

Turning back to the empirics, Gavin criticizes my measurement of nuclear superiority for focusing on quantitative, to the exclusion of qualitative, indicators. I clearly explain my reasons for doing so in the published article and I stand behind them. It would be difficult, if not impossible, to code and measure the qualitative nuclear balance between every pair of states in the international system for every year from 1945 to the present. Moreover, there is good reason to believe that quantitative and qualitative superiority are tightly correlated and that my quantitative measure accurately gauges the concept I am trying to measure. Furthermore, for the sake of argument, even if there is measurement error present (and I have no reason to believe that there is), Gavin does not specify the nature of that error, how the error results in bias, or why the bias would cause me to overestimate, rather than underestimate, the strength of the true relationship between superiority and crisis outcomes.

Also on the issue of measuring superiority, Gavin asserts that the United States and the Soviet Union have pursued superiority and damage limitation more aggressively than other states, and that the United States has done this in part out of a desire to extend nuclear deterrence to its allies. He seems to imply that this somehow poses a problem for my analysis, but he does not clearly explain this point. I disagree with the premise of this criticism (other states, including Pakistan today, have pursued or are pursuing superiority and damage limitation quite aggressively) and whatever Gavin's intended criticisms might be (selection effects and robustness are the only two that come to mind), they are fully addressed in the published article and again below.

On to Gavin's second major criticism: selection into nuclear crises. Gavin writes that "to truly understand how important military balances or resolve are, you would not just analyze crisis outcomes; you would also need to include the crises that never happened, because a state calculated that either it was outgunned or didn't possess the requisite resolve to prevail…It is hard to imagine how you can effectively 'control' for such a thing." But, as I clearly explain in the article and the online appendix, I do control for such a thing. I perform three different tests to account for selection into crises and in each test I find that the nuclear balance of power remains correlated with nuclear crisis outcomes even after accounting for selection into crises.

Moreover, as I also explain in the article, if there is a selection effect it should bias against, not in favor, of my hypothesis. If leaders take the nuclear balance of power into account before initiating crises, then this would mean that nuclear superiority is an important factor in international politics, but that we should not expect to see its largest effects in the outcomes of the crises that actually occur. Selection is a bigger problem, therefore, for those who wish to argue that nuclear weapons do not affect crisis bargaining. The fact that I find a strong relationship between superiority and crisis outcomes in actual crises even in the face of possible selection effects, therefore, provides especially strong support for my hypothesis.

Third, and finally, Gavin makes a number of other minor criticisms of the piece. He inveighs against the widely-used International Crisis Behavior (ICB) dataset for coding the

Berlin Wall crisis as a victory for the Soviet Union, but if one wanted to argue that this was a victory for the nuclear superior United States (and I agree that such an argument could be made), then this would only strengthen my reported relationship between superiority and victory. Gavin claims that in some crises both states achieve their goals, "highlighting how the zero-sum 'win-lose' approach of large-N studies is ill-suited to this case and international politics more broadly," but, as I explain in my article, large-N studies do not require a zero-sum win-lose approach. Indeed, in my study, some crises are coded as producing multiple winners and others as producing multiple losers. Gavin questions whether some of the observations in my dataset belong in a study of nuclear crises, but, as I explain in my article, the findings are robust to the removal of each individual crisis and to each individual country, rendering such criticisms moot. He argues that a study of nuclear crisis outcomes must clearly define what counts as a nuclear crisis and that this definition should include some threshold for the level of danger and whether nuclear weapons were front and center in the crisis. In my article, I provide a clear definition of nuclear crises and explain why, for theoretical and methodological reasons, it would be a mistake to focus only on cases that escalate to a certain level of danger, or in which nuclear weapons were front and center. Next, in a footnote, Gavin raises the possibility that the findings of this article might be in tension with my policy recommendations for addressing the Iranian nuclear challenge, but, in actuality, they strongly support them. A nuclear-armed Iran would be able to hold U.S. assets at risk with the threat of nuclear war, (something that is not possible now), increasing Washington's expected costs in a conflict with Iran, and reducing its willingness to run risks in crises against Tehran. Moreover, in addition to reduced crisis bargaining power, there are many other more dangerous threats posed by a nuclear-armed Iran. As President Obama has clearly stated, a nuclear-armed Iran "is not a challenge that can be contained" and the United States must be willing to "do everything required to prevent it." Finally, Gavin claims that the distinction between deterrence and compellence is fuzzy, and that this distinction is "crucial to both sets of arguments." I agree that this distinction is often fuzzy and that it does pose potential problems for studies that attempt to focus on one or the other, but this distinction is not at all crucial to my analysis. Rather, I focus on the outcomes of all serious showdowns between nuclear-armed states regardless of whether the initiator is attempting to deter or compel the target.

In closing, I would like to thank Gavin again for bringing attention to the subject of nuclear superiority and nuclear crisis outcomes. I hope this brief response serves to clear up any confusion about my recent scholarship and to advance future research on this important subject.

## "Archives and the Study of Nuclear Politics"

**Hal Brands,** Duke University

In early 1979, Saddam Hussein, the de facto ruler of Iraq, laid out his vision for an Arab attack against Israel, one in which an Iraqi nuclear deterrent would play a pivotal role. As Saddam explained to some of his closest advisers, Iraq needed a nuclear weapon—perhaps to be acquired from the Soviet Union—if it were to reverse earlier Arab losses vis-à-vis Israel and subject that enemy to an unprecedented defeat:

> The most important requirement is that we be present in Iraq and Syria and will have planned ahead that the enemy, the air force, that the enemy will come and attack and destroy, etc. We should bear it and keep going—and go put pressure on our Soviet friends and make them understand our need for one weapon—we only want one weapon. We want, when the Israeli enemy attacks our civilian establishments, to have weapons to attack the Israeli civilian establishments. We are willing to sit and refrain from using it, except when the enemy attacks civilian establishments in Iraq or Syria, so that we can guarantee the long war that is destructive to our enemy, and take at our leisure each meter of land and drown the enemy with rivers of blood. We have no vision for a war that is any less than this.[1]

In Saddam's view, having a nuclear deterrent would foreclose the possibility of Israeli nuclear escalation, thereby allowing the Arabs to wage a "patient war" that would last "for twelve continuous months," result in tens of thousands of casualties, and gradually allow Iraq and its allies to liberate (at least some of) the territories conquered by Israel in previous wars.[2]

Saddam's comments can be found in the transcript of a Revolutionary Command Council meeting, one of several hundred Iraqi documents (nearly all of them dealing with issues of foreign policy and national security) first made available to scholars on an unclassified basis in 2010. As David Palkki and I have argued elsewhere, we think that this quotation (and many others like it) provides real insight into the way that Saddam Hussein viewed nuclear weapons—their military and strategic significance, their potential uses vis-à-vis Iraq's enemies, and their overall utility for Baathist statecraft. We also argue that Saddam's nuclear calculus has important implications for theoretical debates about nuclear

[1] Hal Brands and David Palkki, "Saddam, Israel, and the Bomb: Nuclear Alarmism Justified?" *International Security* 36, 1 (Summer 2011), 133-166, quoted at 148-149

[2] Ibid; see also Hal Brands, "Saddam and Israel: What Do the New Iraqi Records Reveal?" *Diplomacy & Statecraft* 22, 3 (Fall 2011), 500-520.

proliferation and its geopolitical consequences.[3] Accordingly, I think that this passage serves as an appropriate point of departure for a paper about the role of new sources and new approaches in the study of nuclear politics.

This paper does three things. First, it points out that it is truly an amazing time to be a scholar of nuclear issues, as archival revelations—from a number of different countries—are reshaping our understandings of old events and opening up new pathways for understanding international nuclear politics. Second, it looks at one set of records that I have worked with in some depth—the records of Saddam's regime in Iraq—as an example of the sort of insights that emerge from a close reading of newly available archival materials. Third and finally, it makes some very preliminary observations as to how political scientists and historians might best exploit the value of this emerging archival record. The overall goal of the paper is not to serve as anything approximating a formal or comprehensive literature review, but simply to provoke discussion about potential avenues for improving and building on the exciting work now being done in the field of nuclear politics.

**I**

It is a very good time to be a scholar of nuclear issues. Since the Cold War—and more recently in some cases—reams of documentation on nuclear statecraft have become available through declassification and other processes. These documents shed light on a wide range of subjects pertaining to the international politics of nuclear weapons, and they have the potential—and in some instances, they have already begun—to reshape the ways that scholars think about important aspects of the nuclear age.

Consider the case of U.S. nuclear history. Through the early 1990s, it was difficult to find extensive documentation on U.S. nuclear policy for any period following the 1962 Cuban Missile Crisis.[4] The situation is now very different. Through the process of declassification, the publication of compendiums like the *Foreign Relations of the United States* series, and the efforts of institutions like the National Security Archive, scholars can now access detailed, highly enlightening archival information on key nuclear issues from the 1960s through the end of the Cold War. The ways that U.S. officials dealt with the strategic challenges of nuclear parity; efforts to forge an international non-proliferation regime and gain the adherence of key states like West Germany, Japan, and South Korea; U.S. policy toward the Pakistani and Israeli nuclear programs; the rethinking of U.S. nuclear doctrine under the Nixon/Ford and Carter administrations; President Ronald Reagan's nuclear calculus and the end of the Cold War; the dynamics of Able Archer 83 and other crises; the

[3] *Ibid.* The relevant records are located (for now) at the Conflict Records Research Center in Washington, D.C., although funding issues may force relocation of the records in the future, or even force the CRRC to shut down altogether.

[4] Even here, the documentation was relatively sparse. By the early and mid-1980s, scholars had access to the ExCom minutes, but much other relevant documentation remained classified. Since then, new documentary releases have been incorporated into excellent works such as Michael Dobbs, *One Minute to Midnight: Kennedy, Khrushchev, and Castro on the Brink of Nuclear War* (New York: Vintage, 2008).

'Madman strategy' and other ways in which U.S. leaders approached nuclear statecraft: all of these subjects can now be treated in some depth using high-level archival materials. Indeed, some of the most provocative scholarship on U.S. policy in the second half of the Cold War has dealt with precisely these issues.[5]

What about the Soviet side of the Cold War? Here too there is lots of material to work with, even though declassification procedures (such as they are) are much more idiosyncratic. Without even traveling to Moscow (or learning to read Russian)[6], researchers can consult a wide range of materials that bear on Soviet nuclear history: the minutes of the Plenary Sessions of the Presidium between 1955 and 1964, which contain striking insights into Soviet Premier Nikita Khrushchev's uses and misuses of nuclear weapons; documents on the Soviet and Warsaw Pact approach to non-proliferation in the 1960s, which demonstrate some of the key nuclear dynamics of the middle Cold War and East-West collaboration on nuclear issues; records on Soviet arms control policies and perceptions of the strategic balance during the 1970s and 1980s; documents on the making of the Intermediate-Range Nuclear Forces (INF) Treaty and Reagan-Gorbachev nuclear diplomacy; records on the role of nuclear weapons in Soviet alliance relations with countries like Cuba and the People's Republic of China; and others.[7] These documents have

[5] Paul Lettow's book, *Ronald Reagan and His Quest to Abolish Nuclear Weapons* (New York: Random House, 2005), is one excellent example. For a representative sampling of some of these new documents, consult the relevant *FRUS* volumes for the Nixon-Ford years, as well as the relevant Electronic Briefing Books made available through the National Security Archive. For other sources that address some of these questions, see William Burr, "The Nixon Administration, the 'Horror Strategy,' and the Search for Limited Nuclear Options, 1969-1972," *Journal of Cold War Studies* 7, 3 (Summer 2005), 34-78; William Burr and Jeffrey Kimball, "Nixon's Secret Nuclear Alert: Vietnam War Diplomacy and the Joint Chiefs of Staff Readiness Test, October 1969," *Cold War History* 3, 2 (January 2003), 113-156; Hal Brands, *What Good is Grand Strategy? Power and Purpose in American Statecraft from Harry S. Truman to George W. Bush* (Ithaca: Cornell University Press, 2014), chap. 3; Dmitry Dima Adamsky, "The 1983 Nuclear Crisis—Lessons for Deterrence Theory and Practice," *Journal of Strategic Studies* 36, 1 (January 2013), 4-41; Hal Brands, "Progress Unseen: U.S. Arms Control Policy and the Origins of Détente, 1963-1968," *Diplomatic History* 30, 2 (April 2006), 253-285; Hal Brands, "Rethinking Nonproliferation: LBJ, The Gilpatric Committee, and U.S. National Security Policy," *Journal of Cold War Studies* 8, 2 (Spring 2006), 83-113; Brian J. Auten, *Carter's Conversion: The Hardening of American Defense Policy* (Columbia: University of Missouri Press, 2008); Kristina Spohr Readman, "Conflict and Cooperation in Intra-Alliance Nuclear Politics: Western Europe, the United States, and the Genesis of NATO's Dual-Track Decision, 1977-1979," *Journal of Cold War Studies* 13, 2 (Spring 2011), 39-89.

[6] Obviously, one can learn much more about Soviet policies if one does travel to Moscow and read Russian. The point here is simply that even non-specialists now have access to exponentially more archival material than would have been the case twenty-five years ago.

[7] Books and articles on these subjects are too numerous to cite comprehensively. As examples, see David Hoffman, *The Dead Hand: The Untold Story of the Cold War Arms Race and its Dangerous Legacy* (New York: Anchor, 2010); Douglas Selvage, "The Warsaw Pact and Nuclear Non-Proliferation," Cold War International History Project Working Paper No. 32 (Washington, D.C.: Cold War International History Project, 2001); Vladislav Zubok, *A Failed Empire: The Soviet Union in the Cold War from Stalin to Gorbachev* (Chapel Hill: University of North Carolina Press, 2007); Timothy Naftali and Aleksandr Fursenko, *One Hell of a Gamble: Khrushchev, Castro, and Kennedy, 1958-1964* (New York: Norton 1997); Hal Brands, "Non-Proliferation and the Dynamics of the Middle Cold War: The Superpowers, the MLF, and the NPT," *Cold War*

been made available through endeavors like the Cold War International History Project, the Parallel History Project on NATO and the Warsaw Pact, the National Security Archive's Electronic Briefing Book series, and the Kremlin Decision-Making Project, and they have made it possible for the nuclear dynamics of the Cold War to be studied more broadly and more thoroughly than ever before.

This progress has not been limited to the records of the Big Two—the last two decades have seen the emergence of documents regarding the nuclear policies of a wide range of important countries and subjects: British nuclear posture and doctrine in the 1960s and 1970s; French views on non-proliferation and the diplomatic uses of nuclear weapons; the Brazilian nuclear program in the 1970s and 1980s; the development and strategic role of an Israeli deterrent; apparent nuclear cooperation between Israel and South Africa; the ways that key non-nuclear states like West Germany, Italy, and Japan viewed the Non Proliferation Treaty (NPT); the reasons why countries like South Korea, Taiwan, and Australia chose not to develop nuclear weapons; even Chinese perceptions of nuclear deterrence and the potential uses of nuclear weapons during the early years following Beijing's first nuclear test in 1964. (In some cases, these countries have made their own documents available to researchers; in others, scholars have had to study the subjects in question through recently declassified U.S. documents or other, similar sources.) And as I'll discuss in greater detail below, an upshot of the U.S. invasion of Iraq in 2003 has been to make available to scholars a number of key Iraqi documents dealing with Saddam Hussein's approach to nuclear issues during his time in power.[8]

At the risk of belaboring the obvious, all of these records are very valuable to students of the nuclear age. In some cases, they shed light on little-known aspects of nuclear statecraft; in others, they illuminate to an unprecedented degree the ideas, motives, and perceptions that fed into nuclear politics. They can correct longstanding misunderstandings and push us toward the resolution of certain debates, while also opening (or reopening) others.

---

*History* 7, 3 (August 2007), 389-423; Viktor M. Gobarev, "Soviet Policy Toward China: Developing Nuclear Weapons 1949-1969," *Journal of Slavic Military Studies* 12, 4 (December 1999), 1-53.

[8] For a sampling of literature that draws on these sources, see Matias Spektor, *Kissinger e o Brasil* (Rio de Janeiro: Zahar, 2009); Sasha Polakow-Suransky, *The Unspoken Alliance: Israel's Secret Relationships with Apartheid South Africa* (New York: Vintage, 2011); Avner Cohen, *Israel and the Bomb* (New York: Columbia University Press, 1998); Cohen, *The Worst Kept Secret: Israel's Bargain with the Bomb* (New York: Columbia University Press, 2010); Alexander Lanoszka, "Protection States Trust? Explaining South Korea's Nuclear Behavior," October 2012, http://www.princeton.edu/politics/about/file-repository/public/A-Lanoszka-10-17-12-PST-ROK.pdf; Matthew Jones and John W. Young, "Polaris, East of Suez: British Plans for a Nuclear Force in the Indo-Pacific, 1964-1968," *Journal of Strategic Studies* 33, 6 (December 2010); William Glenn Gray, "Commercial Liberties and Nuclear Anxieties: The U.S.-German Feud Over Brazil, 1975-7," *International History Review* 34, 3 (Summer 2012), 449-474; Christine M. Leah, "US Extended Nuclear Deterrence and Nuclear Order: An Australian Perspective," *Asian Security* 8, 2 (Summer 2012), 93-114; and the relevant essays in Olav Njolstad, ed., *Nuclear Proliferation and International Order: Challenges to the Non-Proliferation Treaty* (New York: Routledge, 2011).

How dangerous was Able Archer?[9] How did the possession of nuclear weapons affect Soviet behavior during the Joseph Stalin and Khrushchev years? Was Ronald Reagan really a nuclear abolitionist? How did the Germans and Japanese really view the NPT? What role did the Soviet nuclear buildup of the 1960s and 1970s, or the Carter-Reagan buildup of the 1980s, play in the course of the Cold War? These are just some of the questions that the documents can help us begin to answer.

Yet what makes these records so valuable is not simply that they enhance historians' understanding of things that happened decades ago. Rather, these records have real significance for some key theoretical debates surrounding nuclear weapons and international relations. Why do states want nuclear weapons in the first place, and how do nuclear capabilities affect the behavior of states that get them? How is the NPT viewed by some of its key non-nuclear signatories? How many bombs does it take to constitute an effective deterrent? Does the relative nuclear balance (or, more accurately, perceptions of that balance) make a difference in the resolution of crises and disputes? These and related questions are at the center of scholars' efforts to understand the nuclear age. And many of them can be addressed more completely and accurately than before through an investigation of the relevant archival records.

**II**

As a brief illustration of the above points, think just about the Iraqi records. These are the documents with which I have worked most closely, and they provide a good example of the sort of insights—both empirical and theoretical—that can emerge from a close reading of archival records.

From my perspective, the most important information in the Iraqi records has to do with the question of why Saddam wanted the bomb in the first place—an issue that is of interest not just to scholars of Iraq, but to anyone interested in studying the broader theoretical issue underlying that question.[10] The Iraqi records show that Saddam, from the mid- and late 1970s onward, was interested in nuclear weapons for two reasons that are well established in the theoretical literature: deterrence vis-à-vis enemies like Israel and Iran, and considerations of national prestige.[11] But they also show that Saddam was attracted to the bomb for another reason, one that is less commonly referenced in the traditional literature on proliferation, and one not generally apparent to scholars of Iraqi policy prior

[9] This particular question has seen a number of recent, useful contributions based on archival documents. See, for instance, Vojtech Mastny, "How Able Was 'Able Archer'? Nuclear Trigger and Intelligence in Perspective," *Journal of Cold War Studies* 11, 1 (Winter 2009), 108-123; Adamsky, "The 1983 Nuclear Crisis," passim.

[10] Indeed, as noted below, because there are small-n problems involved in studying why states develop (or seek to develop) nuclear weapons, extensive qualitative work is all the more important—and rewarding—in addressing this question.

[11] See, for instance, Scott Sagan, "Why Do States Build Nuclear Weapons? Three Models in Search of a Bomb," *International Security* 21, 3 (Winter 1997), 54-86.

to the release of these documents: Saddam, it now seems clear, wanted nuclear weapons as a means of enabling *conventional* attacks on Israel. Time and again, Saddam came back to the idea that if Iraq had even a very small nuclear arsenal, it could establish nuclear deterrence vis-à-vis Israel, thereby allowing the Baathist regime and its Arab allies to wage a prolonged, conventional war to liberate at least some of the territories lost in 1967.[12] If Iraq lacked atomic weapons, on the other hand, then it would remain vulnerable to nuclear blackmail even as its armies advanced. In 1978, for instance, Saddam explained to military officials that he saw nuclear weapons as a means of neutralizing Israel's military advantages:

> When the Arabs start the deployment, Israel is going to say, "We will hit you with the atomic bomb." So should the Arabs stop or not? If they do not have the atom, they will stop. For that reason they should have the atom. If we were to have the atom, we would make the conventional armies fight without using the atom. If the international conditions were not prepared and they said, "We will hit you with the atom," we would say, "We will hit you with the atom too. The Arab atom will finish you off, but the Israeli atom will not end the Arabs."[13]

Saddam often came back to the same theme in the late 1970s and early 1980s, and even a decade later, in the late 1980s, as well. "Now, if the Arabs were to have a nuclear bomb," he said in early 1990, "wouldn't they take the territories that were occupied after 1967?" For Saddam, nuclear weapons were meant not simply for purposes of deterrence and prestige, but also to enable conventional war.[14]

If one accepts this conclusion, then there are some interesting theoretical implications that follow. The Iraqi records provide numerous illustrations of a statesman invoking the logic of the stability/instability paradox. They show that Saddam believed that nuclear weapons could have great strategic and military utility even if they were never used—and that he believed that even a very small and primitive arsenal would be enough to deter a much larger Israeli arsenal. Additionally, they bear on the long-running optimism/pessimism debate as well as newer questions about "nuclear alarmism."[15] On the one hand, the Iraqi documents are actually somewhat reassuring, in that there is no evidence that Saddam wanted nuclear weapons so that he could launch an unprovoked attack on Israel or any other country. On the other hand, the documents are troubling because they indicate that

[12] At times, Saddam spoke of the need to eliminate Israel altogether, but in general he focused on achieving more limited gains.

[13] Brands, and Palkki, "Saddam, Israel, and the Bomb," 149.

[14] Brands and Palkki, "Saddam, Israel, and the Bomb," passim. Amatzia Baram made a similar point in his "An Analysis of Iraqi WMD Strategy," *Nonproliferation Review* 8, 2 (2001), 25-39.

[15] On nuclear alarmism, see, for instance, John Mueller, *Atomic Obsession: Nuclear Alarmism from Hiroshima to al-Qaeda* (New York: Oxford University Press, 2010).

had Saddam obtained nuclear weapons, he might have used them for purposes that still would have been bloody and highly destabilizing. Finally, in an echo of other recent work on nuclear issues, the documents help flesh out just how amateurish and idiosyncratic Saddam's views on nuclear strategy (and international relations more broadly) could be.[16] As Palkki and I have noted in our article on this subject, Saddam's nuclear calculus seems to have been sincerely held, but it was flawed on numerous grounds—not least his belief that Iraq (even with a collection of Arab allies at its side) could defeat Israel in a conventional war once deterrence was established at the nuclear level.

Looking beyond the question of why Saddam wanted nuclear weapons and what he planned to do with them, there are other intriguing insights to emerge from the Iraqi records. Transcripts of Saddam's meetings with the Revolutionary Command Council and other senior groups help explain how he interpreted the ambiguous (potentially nuclear) threats made by the United States on the eve of the Persian Gulf War in 1991. In a recent chapter for an edited volume, Palkki and Shane Smith have examined the way that international sanctions and inspections impacted Saddam's perceptions of the utility of his nuclear program in the 1990s, and the reasons why the regime ultimately parted with much of that program while still seeking to retain basic facilities and know-how.[17]

To be sure, the Iraqi records now available do *not* give a complete portrayal of Iraqi nuclear politics under Saddam. The existing records provide a fragmentary picture of Iraqi statecraft, and they raise obvious questions about how literally one should take Saddam's statements about nuclear weapons and other important issues. As with any archival records, they need to be combined with a range of other sources in order to provide a more complete, better-rounded picture of the subject under investigation—and even then, there will be room for interpretation, speculation, and disagreement. But all in all, I would argue that these records do provide greater insight than any other sources into the strategic motivations underlying Saddam's nuclear decision-making, and that they have the potential to constructively influence theoretical debates as well as empirical ones.

**III**

[16] See, for instance, Paul Bracken, *The Second Nuclear Age: Strategy, Danger, and the New Power Politics* (New York: Macmillan, 2012); Jacques Hymans, *The Psychology of Nuclear Proliferation: Identity, Emotions, and Foreign Policy* (Cambridge: Cambridge University Press, 2006); or Francis Gavin, *Nuclear Statecraft: History and Strategy in America's Atomic Age* (Ithaca: Cornell University Press, 2012), 104-119. In Saddam's case, these idiosyncratic tendencies were not limited to his views of nuclear weapons. See, for instance, Hal Brands, "Why Did Saddam Invade Iran? New Evidence on Motives, Complexity, and the Israel factor," *Journal of Military History* 75, 3 (July 2011), 681-705; Hal Brands and David Palkki, "Conspiring Bastards: Saddam Hussein's Strategic View of the United States," *Diplomatic History* 36, 3 (June 2012), 625-659; Kevin Woods, James Lacey, and Williamson Murray, "Saddam's Delusions: The View from the Inside," *Foreign Affairs* 85, 3 (May/June 2006), 2-28.

[17] See David Palkki and Shane Scott, "Contrasting Causal Mechanisms: Iraq and Libya," in Etel Soligen, *Sanctions, Statecraft, and Nuclear Proliferation* (Cambridge: Cambridge University Press, 2012); Palkki, "Deterring Saddam Hussein's Iraq: Credibility Assessments in Theory and Practice," unpublished manuscript, 2013.

In many ways, it would seem superfluous to suggest ways that scholars should utilize these new materials, or to offer comments on potential directions for the field of nuclear studies more broadly. Both History and Political Science already have strong cohorts of scholars doing innovative and exciting work on nuclear issues, ranging from proliferation and its consequences, to arms control and reasons for nuclear abstention, to the ways that the nuclear balance affects the outcome of international crises. Indeed, a number of scholars—historians especially—have already been energetically utilizing new archival materials to offer some groundbreaking insights about a range of key issues.

In the spirit of generating discussion, however, here are a couple of brief observations about archives and nuclear studies. First, for political scientists particularly, I think that the new materials offer an opportunity to strengthen and perhaps reinvigorate the qualitative research agenda. As an outsider looking in, I am struck by the degree of which political scientists seem to be moving *away* from the sort of deep, archives-based, case-study analysis that would appear to be the most rewarding way of exploiting the emerging material. Political scientists have long asked the biggest, most ambitious, most interesting, most policy relevant questions about nuclear issues. But for a variety of reasons, they have long been much less likely than historians to do deep archival work on these issues.[18] And based on an admittedly impressionistic survey of the field, it seems as though they are now increasingly answering those questions through quantitative analysis. Much of the recent work on proliferation, for instance, is based on such techniques, as is the highest-profile recent scholarship on the relationship between nuclear weapons and coercion.[19] And just from looking at the nuclear-related articles run by leading journals like *International Organization, Journal of Conflict Resolution,* the *American Political Science Review,* and others, it would seem as though the subfield as a whole is trending in this direction. (*International Security* and *Security Studies* are still more of a mixed bag.)

Quantitative work certainly has its benefits, which I do not aim to dispute here.[20] But when dealing with nuclear issues, it should also be noted that quantitative work has limits.[21] There are, of course, the small-n problems that come with studying a number of nuclear

[18] There are prominent exceptions, of course. See, for instance, Scott Sagan, *The Limits of Safety: Organizations, Accidents, and Nuclear Weapons* (Princeton: Princeton University Press, 1995); Peter Feaver, *Guarding the Guardians: Civilian Control of Nuclear Weapons in the United States* (Ithaca: Cornell University Press, 1992); Hymans, *Psychology of Nuclear Proliferation*; Daryl Press, *Calculating Credibility: How Leaders Assess Military Threats* (Ithaca: Cornell University Press, 2007).

[19] See, for instance, Robert Rauchhaus, Matthew Kroenig, and Erik Gartzke, *Causes and Consequences of Nuclear Proliferation* (New York: Routledge, 2011); Matthew Kroenig, "Nuclear Superiority and the Balance of Resolve: Explaining Nuclear Crisis Outcomes," *International Organizations* (Volume 67, Issue 1, January 2013), 141-171; Todd S. Sechser and Matthew Furhmann "Crisis Bargaining and Nuclear Blackmail" by *International Organizations* (Volume 67, Issue 1, January 2013), 173-95.

[20] On the benefits, see Erik Gartzke's contribution to this forum.

[21] The best discussion of those limits is Scott Sagan and Alexander Montgomery, "The Perils of Predicting Proliferation," *Journal of Conflict Resolution* 53, 2 (April 2009), 302-328.

issues. More broadly, relying on statistical analysis makes it difficult to capture the complexity that is often at work in studying nuclear politics. As Francis Gavin has recently pointed out, for instance, the very attempt to code the outcomes of crises involving nuclear powers—to answer, in a straightforward and accurate fashion, the question of who won and who lost—is fraught with nasty complications.[22] To put it another way, if reasonable people can code the Cuban Missile Crisis as a U.S. victory (because the missiles came out), a draw (because the United States secretly traded the Jupiter missiles in Turkey and openly pledged not to invade Cuba), or even a Soviet victory (because the crisis essentially ensured the survival of the Castro regime, an outcome that had looked quite dubious in mid-1962), then it is hard to know what to make of any coding scheme of winners and losers in nuclear crises.[23]

In fact, I would argue that deep qualitative research—often in archival records—is central to making sense of many of the questions that political scientists are currently seeking to answer. If you want to know how the strategic balance affected a crisis, for instance, you need to understand, as accurately as possible, how that balance was *perceived* by the officials in question. If you want to understand how acquiring nuclear weapons affects the behavior of a given state, you have to look at the complicated and often conflicting ways in which those weapons altered (or not) the strategic calculus of leaders in power. If you want to know how many bombs it takes to constitute an effective deterrent, you need to find out when leaders of the country in question—as well as the leaders of the opposing country—*believed* that this threshold had been passed.[24] In other words, so much of nuclear history revolves around *perception* as much as reality. And the best way to understand those perceptions, and to capture the immense complexity and nuance of many nuclear issues more broadly, is through serious archival research.

This is not to dismiss the need for ambitious, large-n quantitative studies, or to disparage the scholarship of those political scientists who are already doing exceptional qualitative work on nuclear issues. Nor, for that matter, do I think that *every* important topic in nuclear studies can or should be tackled through extensive archival research. I simply want to note that the emerging archival materials underscore the need for a vigorous, archives-based program of qualitative political science research on key nuclear issues. And at the very least, the possibilities for doing such work at tolerable costs in time and effort are greater now than ever before.

[22] See Gavin's contribution to this forum.

[23] On these aspects of the crisis, see, for instance, John Lewis Gaddis, *We Now Know* (New York: Oxford University Press, 1997), chapter 8.

[24] For an interesting effort to grapple with this issue, see Lyle Goldstein, "Do Nascent WMD Arsenals Deter? The Sino-Soviet Crisis of 1969," *Political Science Quarterly* 118, 1 (Spring 2003), 53-80.

As for historians, there are already a number of scholars who are doing terrific work with these materials on a wide range of subjects.[25] Rather than suggesting new topics for them to consider, I would simply suggest that as historians continue to do their work, they should think about new ways of wringing maximum value out of the records they use and the insights they generate. Broadly speaking—and this is probably somewhat ironic in light of the foregoing—historians need to be more 'political sciency' when they address nuclear issues. They need to be more attuned to the theoretical implications of their work and more willing to explicitly engage the relevant political science and strategic studies literature; they need to be willing to try their hand at comparative history; and they need to work—as political and other social scientists often do—on more collaborative projects.

With respect to the first of these tasks, historians need to be far more theory-conscious in dealing with nuclear issues. There is a vast body of theoretical literature, produced by the political science and strategic studies communities, on these subjects, but historians too often write their books and articles as though this theoretical work did not exist.[26] This silence may be because historians simply aren't conversant in the relevant theoretical literature, or because there isn't much professional payoff to engaging that literature if you want to make a living working in a traditional history department.

But the lack of theory-literate historians imposes real limitations on the work we do. Some of the most interesting historical on nuclear issues has been produced when scholars like Marc Trachtenberg have used archival materials to test common theoretical propositions.[27] And when historians fail to do this sort of work, they are missing a chance to bring cutting edge archival work to bear on important theoretical and policy debates, they are failing to put two closely related bodies of literature in conversation with one another, and they are ignoring what may be some very interesting ways of framing their subjects. On top of all this, they are giving political scientists little incentive to engage historical work, further impoverishing the dialogue between the disciplines. So if historians really want to make good use of the exciting archival revelations they are constantly discovering, they need to treat those revelations as opportunities to test, refine, and refute the prominent theoretical concepts in nuclear studies.

Historians also need to be willing to do more comparative history. Doing serious comparative, archival history is hard for lots of reasons—it is immensely time-consuming, it is costly, and it often requires multiple languages. But historians have overcome these obstacles in other aspects of diplomatic/international history, and there are good reasons

[25] See, for instance, many of the sources cited previously.

[26] There are exceptions, of course, most notably Mark Trachtenberg, Francis Gavin, and John Gaddis. But at least two of these are exceptions that prove the rule—Trachtenberg and Gavin are historians who work (or will soon work) in political science departments. I should also note that I include some of my own earlier work on non-proliferation in this indictment.

[27] See Trachtenberg, *History and Strategy* (Princeton: Princeton University Press, 1991).

to try it in nuclear history as well.[28] How has the nuclear balance affected crisis behavior? How has the acquisition of nuclear weapons affected the behavior of different states? How did different governments approach the negotiation of the NPT in the 1960s, or the nonproliferation regime that followed? How have different governments or administrations approached similar challenges of the nuclear age? Comparative history can shed light on these and other questions. It can give scholars more analytical leverage than they would gain by simply looking at a single case, allowing them to explore similarities and differences across cases, to see which dynamics are common and which seem to be more idiosyncratic, and to see how much issues like politics, ideology, and culture influence the way that states approach nuclear weapons. Comparative history can thereby be a powerful tool for illuminating the nuances, patterns, and anomalies of the nuclear era.

Finally—and related to the first two points—historians need to do more collaborative work. Historians have generally been reluctant to do such work, for some reasons that make sense and some that do not. But here as elsewhere, there is much to be gained from trying new approaches. Collaborative work represents a natural way of bringing together the various expertise and language skills that are often necessary to do good comparative history. It is also an easy way of bringing together people who know the archival material with those who have the theoretical knowledge and methodological skills necessary to engage with other disciplinary literatures. Indeed, some very interesting work on nuclear issues has been done by teams that have brought together just these sorts of skill sets.[29] So there is lots of interesting work to be done with the new archival material, but like political scientists, historians will need to be flexible and creative in thinking about how to do it.

**Dr. Hal Brands** is an Assistant Professor of Public Policy and History at Duke University. He previously worked at the Institute for Defense Analyses outside of Washington, DC. He is the author of *What Good is Grand Strategy? Power and Purpose in American Statecraft from Harry S. Truman to George W. Bush* (2014); *Latin America's Cold War* (2010); and *From Berlin to Baghdad: America's Search for Purpose in the Post-Cold War World* (2008), as well as numerous articles on grand strategy, arms control and nuclear issues, and other subjects. Dr. Brands holds a Ph.D. in history from Yale University.

[28] See, for instance, Erez Manela, *The Wilsonian Moment: Self-Determination and the International Origins of Anticolonial Nationalism* (New York: Oxford University Press, 2007); Jeremi Suri, *Power and Protest: Global Revolution and the Rise of Détente* (Cambridge: Harvard University Press, 2003); Hal Brands, *Latin America's Cold War* (Cambridge: Harvard University Press, 2010); Tanya Harmer, *Allende's Chile and the Inter-American Cold War* (Chapel Hill: University of North Carolina Press, 2011).

[29] I am thinking particularly of two articles: Scott Sagan and Jeremi Suri, "The Madman Nuclear Alert: Secrecy, Signaling, and Safety in October 1969," *International Security* 27, 4 (Spring 2003), 150-183; and Gordon Chang and He Di, "The Absence of War in the U.S.-China Confrontation over Quemoy and Matsu in 1954-1955: Contingency, Luck, Deterrence," *American Historical Review* 98, 5 (December 1993), 1500-1524. The first article represents a collaboration between a political scientist and a diplomatic historian; the second is a cooperative effort involving one scholar from the United States and one from China.

## **"An Apology for Numbers in the Study of National Security . . . if an apology is really necessary"** by **Erik Gartzke**, University of California, San Diego

**I**

One way to gain understanding is to count things and then use this information to assess relationships. This is an approach that has gradually swept through the sciences, engineering and is now increasingly practiced in the social sciences and humanities. There are other ways of gaining knowledge, including deduction (rational or otherwise), expert judgment, and qualitative assessment. Quantification is not a substitute for qualitative evaluation or theoretical reasoning. It is a complement. That said, it poses important potential (comparative) advantages that have proven fruitful in an enormous and enormously diverse set of intellectual disciplines. What is the case for not using quantification to study the causes and consequences of nuclear weapons?

For more than sixty years, it was widely assumed that the study of nuclear security was not amenable to quantitative analysis. There were few countries with nuclear capabilities and nuclear weapons had existed for only a short time. For a large portion of that period, this was indeed the case. It is not any longer. The last decade has witnessed a flowering of quantitative empirical studies of correlates of the nuclear age. These studies have explored the reasons that nations proliferate[1] and why they do not.[2] Research has assessed the impact of nuclear capabilities on alliance ties[3] and extended deterrence. Studies have begun to assess the relationship between nuclear status and crisis escalation[4]

[1] Sonali Singh and Christopher Way, "The Correlates of Nuclear Proliferation." *Journal of Conflict Resolution* 48(6) 2004:859–885; Jo Dong-Joon Jo and Erik Gartzke, "Determinants of Nuclear Proliferation: A Quantitative Model." *Journal of Conflict Resolution* 51(1) 2007:167–194; Matthew Fuhrmann, "Taking a Walk on the Supply Side: The Determinants of Civilian Nuclear Cooperation." *Journal of Conflict Resolution* 53(2) 2009:181–208; Matthew Kroenig, "Importing the Bomb: Sensitive Nuclear Assistance and Nuclear Proliferation." *Journal of Conflict Resolution* 53(2) 2009 :161–180.

[2] Philipp C. Bleek and Erik Lorber. 2012. "Can Security Assurances Prevent Allied Nuclear Proliferation?: A Mixed-Method Assessment." 2012 Typescript. Monterey Institute of International Studies and Duke University.

[3] Matthew Furhmann and Todd S. Sechser, "Signaling Alliance Commitments: Hand Tying and Sunk Costs in Extended Nuclear Deterrence," 2012 Typescript. Texas A&M University and University of Virginia.

[4] Kyle Beardsley and Victor Asal, "Nuclear Weapons as Shields." *Conflict Management and Peace Science* 26(3) 2009: 235–255.

brinkmanship,[5] (and nuclear arsenal size and conflict.[6]

In this very preliminary effort, I seek not to criticize the critics, but rather to make the case that quantitative analysis of nuclear security is a useful addition to the tools available with which to make sense of extremely important questions, indeed ones that may decide the fate of humanity. Pointing to the flaws in any approach is useful if it leads to refinements that improve the products of inquiry. It will be less constructive if the effort is designed to squash what is seen as competition in its infancy or to dismiss a new perspective out of an instinctive desire to preserve established approaches or hierarchies. Fortunately, much of the criticism is of the former type, which is both informative for quantitative research and in a few instances helps to identify and hopefully clarify confusions that persist about the process of reducing important subjects to 'mere numbers.'

**II**

A strong critique of the quantitative approach is simply that it is not possible or is perhaps pointless to measure nuclear capabilities. Some theoretical research asserts that nuclear weapons have symbolic, rather than tangible security consequences .[7] To the degree that such claims are true, one should find at best a tenuous relationship between nuclear status or capabilities and consequences, such as the exercise of influence or the advent of conventional war and peace. This may be, in fact, what has been found. Longstanding controversies about the effects of nuclear arsenals serve to emphasize the ambiguity surrounding such weapons. Two of the leading scholars of international relations have spent at least fifteen years in a public and active controversy over whether nuclear weapons deter or inflame.[8] This embodiment of the adversarial system has led to a number of insights, but no firm resolution of the basic controversy.

The adversarial system doesn't always work as well as one would wish. Answers may prove elusive because the kind of information needed to arrive at answers does not exist. Opposing counsel in a criminal trial can speak to facts that are at least potentially available.

[5] Todd S. Sechser and Matthew Fuhrmann, "Crisis Bargaining and Nuclear Blackmail." *International Organization* 67(1) 2013 :173–195; Beardsley and Asal, "Winning with the Bomb: The Advantages of Being a Nuclear Power." *Journal of Conflict Resolution* 53(2): 2009, 278-301.

[6] Matthew Kroenig,"Nuclear Superiority and the Balance of Resolve: Explaining Nuclear Crisis Outcomes." *International Organization* 67 (1) 2013: 141-171.

[7] Nina Tannenwald, *The Nuclear Taboo: The United States and the Non-use of Nuclear Weapons since 1945*. Cambridge: Cambridge University Press, 2007. Press, Sagan and Valentino conducted a survey experiment in which U.S. respondents seem to exhibit no particular inhibition against advocating the use of nuclear weapons, apparently contradicting the nuclear taboo. See Daryl G Press, Scott D. Sagan & Benjamin A. Valentino, "Atomic Aversion: Experimental Evidence of Taboos, Traditions, and the Non-use of Nuclear Weapons." *American Political Science Review* 107 (1): 2013, pp. 188-206.

[8] Scott D. Sagan and Kenneth N. Waltz. *The Spread of Nuclear Weapons: An Enduring Debate.* 3rd ed. New York: W.W. Norton, 2012.

A murder happened or it did not; the defendant owned a pump-action shotgun, or she did not. However, information of this type may not be forthcoming in international relations and particularly not in studying nuclear security, with (almost) no cases of usage, few countries that have nuclear weapons, and relationships that are tendencies (at best) with contrasting tendencies, or just plain noise.

The dialectic is not particularly effective at providing scale; perhaps both tendencies exist, though neither is large, at least in small samples. The argument that nuclear weapons are mostly ineffectual is not far different observationally from the view that nuclear weapons are so enormously potent that all are afraid to use them or that indeed, they are doing their intended job when they are not being used (Brodie 1959).[9] It would not be particularly surprising if early attempts to unravel ambiguity and surrounding controversy of this type through quantitative methods were subject to some controversy of their own, especially since three generations of qualitative research have been unable to resolve these issues.

Qualitative work got first pick of the pool of questions and the easy stuff will tend to have been figured out by now (perhaps). Qualitative researchers have also had a considerable amount of time to work through deficiencies and difficulties.[10] The issues that remain at this point are those over which qualitative debates have proven less than definitive, such as the Kenneth Waltz-Scott Sagan debate about nuclear stability. There may be many flaws that need to be ironed out in quantitative research, just as there have been with qualitative accounts of complex, multi-causal relationships. Indeed, some of these difficulties remain and it may be in these areas that quantitative research proves valuable.

Ambiguity can also result from a problem that might be described as 'Swiss cheese.' Depending on where one looks, Swiss cheese is either cheese, or it is not. There are big holes in the *fromage* that, depending on where one is looking, can make the cheese, not cheese. Controversies about relationships in which behavior or effects are heterogeneous can prove difficult to evaluate qualitatively because evidence will exist for contrasting claims. This is not an issue of noise or of subtle effects. There is most definitely cheese in some places and decidedly no cheese elsewhere. Imagine the debate between two observers, one who received a sample of the cheesy part and the other whose sample is the center of a hole. 'Yes it is,' 'no its not,' and so on. The problem of dueling examples is familiar to any historian. How is one to interpret a world where every tendency has a countervailing tendency? The Waltz-Sagan debate again can be used to illustrate the problem, precisely because it takes on this character, but also because of the deficiencies in the adversarial system already mentioned. Whether effects reflect important contrasting

[9] John Mueller, "The Essential Irrelevance of Nuclear Weapons: Stability in the Postwar War." *International Security* 13(2): 1988, 55–79; Nina Tannenwald, "The Nuclear Taboo: The United States and the Normative Basis of Nuclear Non-Use." *International Organization* 53(3): 1999, 433–468; and Bernard Brodie, *Strategy in the Missile Age.* Princeton: Princeton University Press, 1959.

[10] I will not claim that Thucydides had nuclear issues pegged 2500 years ago, as some claim about international security generally. Schelling argues that nothing interesting has been produced in deterrence since roughly 1968.

tendencies of lots of not-so-much, the temptation is to exaggerate the pervasiveness of one's preferred evidence. Whatever differences exist are hard to scale, even while debate pushes participants toward rhetorical hyperbole.

One version of the Swiss cheese problem arose during work on the 2009 special issue in *the Journal of Conflict Resolution*. Alexander Montgomery and Scott Sagan pointed out that different scholars were using different definitions of nuclear status and that this belied some confusion about what, precisely, was being measured.[11] The critique was a valid one on its merits; countries like Israel and India have not been entirely frank about their nuclear status at various times. Researchers for their part hold discrepant opinions about when North Korea, for example, crossed the nuclear threshold. There has been intense debate among experts about which date is correct. The old adage that an intellectual is someone who cares more and more about less and less seems applicable in this instance. Choosing dates arbitrarily primised an intellectual train wreck, since the intensity and diversity of expert opinion seemed to invite some disapproval of whatever dates were chosen. As it turned out, a surprising and useful thing happened that was only possible because of the application of statistical methods. What could have degenerated into a lengthy and ultimately unresolvable debate about when, really, highly secret nuclear programs achieved certain milestones quickly established a very simple fact. Whatever plausible dates one wished to champion, they did not matter in terms of the results of any of the analyses conducted by the studies involved in the project.[12]

Academics can disagree about anything and there is no shortage of things about which to disagree. There is certainly an interesting debate to be had about numerous details of the development of nuclear weapons in each of the nuclear powers. However, it would be a disservice to a critical set of questions to let the search for better answers become mired in details, especially when it is possible to determine whether the answers to these questions hinge on particular values in any given controversy. If in addition it turns out that individual judgment or opinion is not critical in a given area, then this is an extremely liberating discovery for all involved. We can then leave such debates to be conducted at leisure by the appropriate scholarly community. Rather than submitting to the untenable position that every detail must be hammered out and agreed upon before moving forward, or imposing arbitrary winners and losers in such debates, we can use statistics to tell us where values are or are not critical to the types of conclusions that may or may not be supportable.

This leads me to a basic difficulty in all forms of analysis, not just those posed by the use of

---

[11] Erik Gartzke and Matthew Kroenig, "A Strategic Approach to Nuclear Proliferation." Journal of Conflict Resolution 53(2): 2009, 151–160; and Alexander Montgomery and Scott Sagan, "The Perils of Predicting Proliferation." Journal of Conflict Resolution 53(2): 2009, 302–328.

[12] A similar phenomenon occurred in the effort to develop quantitative measures of democracy and democratization. Intense debate over the details of index construction that had sustained diverse efforts in the 1980s and early 1990s subsided as few actual empirical differences could be identified in the resulting applied research.

numbers. Researchers make an enormous number of assumptions to move forward with a particular line of inquiry. Francis Gavin's laudable and carefully reasoned critique of two recent quantitative studies of nuclear brinkmanship is, like all works of non-fiction, packed with assertions, overt and otherwise. How do we know that what Gavin read and read into history is really there? One defense is that Gavin does not stray far from the documents. However, as he himself points out, the documents merely chronicle what participants thought, or worse yet said, not what actually happened. It is necessary to interpret history, to emphasize some things while discounting others. All of this is done through a screen of beliefs about what history is really about. This is why, in part, different historians can read the same case differently. The more we rely on the precision of a case, the more we rely on the objectivity of the scholar, which in the end is not clearly demonstrable but is an article of faith. This may be easy to do with a researcher of Gavin's caliber, but it is something that is not scalable, replicable, or readily assailable. How can one separate the quality of conclusions from the quality of the researcher, a quantity that is based on reputation as much as reality?[13]

One way to assist in relying less on the genius of the researcher is to use techniques that facilitate replication. Quantitative analysis clearly helps in this regard. Replication standards in Political Science and other allied fields require researchers to share their data with readers of a journal article and to make accommodation so that it is possible to verify reported findings. On occasion, it is discovered that replication is not possible, or that changes to the analysis alter published results. This can cause a minor scandal and is not infrequently used as evidence of the deficiencies of quantitative analysis. The actual consequence is just the opposite; rather than a flaw in the method, it is a virtue of quantitative work that it assists critics and advocates in conducting an evidence-based debate. I am happy to see such controversies because of what often follows and because they focus attention on the right sorts of issues and questions. We can have fairly high confidence that, where there is interest, conclusions will be drawn and a consensus will form from it.

Gavin, in his detailed and thoughtful dissection of the controversy between Matthew Kroenig and Todd Sechser and Matthew Fuhrmann, makes an astute connection to an earlier controversy between Paul K. Huth and Bruce Russett and Ned Lebow and Janice Stein.[14] One thing that Gavin does not mention in his review is the disposition of this earlier debate. In a response to their critics, Huth and Russett simply dropped all of the cases that Lebow and Stein objected to as being miscoded.[15] The results, surprisingly, were

[13] As we have found in numerous instances, researchers of stellar reputation can produce flawed research. I do not need to go into details here, but the problem is perhaps reliance on authority rather than verification.

[14] Paul K. Huth and Bruce Russett, "What Makes Deterrence Work?: Cases from 1900 to 1980." *World Politics* 36(4): 1984, 496–526; and Richard Ned Lebow and Janice Gross Stein, "Rational Deterrence Theory: I Think, Therefore I Deter." *World Politics* 41(2): 1989, 208–224. The subtitle of the Lebow and Stein article is one of the most witty and evocative in international relations.

[15] Paul K. Huth and Bruce Russett, "Testing Deterrence Theory: Rigor Makes a Difference." *World*

much the same. Whatever one thought about the coding of individual cases — a qualitative issue that ultimately may not be resolvable — the authors' observations about tendencies in this debate did not depend on their preferred coding of the cases in dispute.

Even while quantitative analysis is better equipped to address certain kinds of controversy, it is a more permissive environment within which to ignore other issues involved in making inferences. Qualitative and quantitative research differs, importantly, in how each reacts to error. There is a justifiably strong tradition in scholarship that focuses on veracity, precision and truth. It goes against these instincts when a researcher sees what he or she deems to be an error in describing a case, for example. This is indeed extremely important in qualitative inference. Not only is a sloppy or otherwise mistaken case wrong and very likely to lead to incorrect conclusions, but errors are unlikely to allow the researcher to infer even the direction of the mistake. A less-than-perfect bit of qualitative inference confounds conclusions in an indeterminant manner. This is an excellent reason for lavishing enormous care on details of a case, and why debates evolve over when, precisely, India became a nuclear nation, for example. It does not follow, however, that the effort to avoid error is invariably, or even often, successful, or that the result is satisfactory in any given case. How perfect must data collection, weighting, and the narrative be before we can say with confidence as a community that a case is "right" and that inference can commence? Might there be examples on the margins where it is difficult to determine whether this threshold has been achieved? Indeed, could it even be the case that there is no particular threshold that makes logical sense in the majority of cases, since the effect of an error will differ depending on its impact on inference, not just on its absolute empirical veracity? In short, there exists in qualitative research no metric or even rule-of-thumb by which to evaluate what constitutes adequate precision and no method is available by which to conduct quality control, other than the adversarial system mentioned above.

Quantitative analysis is vastly more forgiving of errors. It is not that we should wish for errors, but that they are an inevitable part of all human activity, even with careful scholarship. One of the things that seems most difficult for researchers to absorb, even for many quantitativists, is the enormous robustness of statistical analysis to miscodings. A statistical study can even be based on data that is mostly in error and still provide the 'true' relationship, as long as the errors are constructed in such a way that they are not intentionally, or systematically, biasing of results.[16] Statistical estimation treats 'random' errors as noise, removing their impact on inference. Since there is so much noise in social scientific data anyway, and since in most circumstances we would rather not commit the error of believing the noise, whether by coding or generated by the environment that we are studying, separating out the signal and the noise is extremely useful.

This leads to the paradoxical point that attempts to 'fix' or correct errant data, by

*Politics* 42(2): 1990, 466–501.

[16] The differences between error and bias are a somewhat technical subject that I hasten to avoid in this setting.

interested parties, can actually lead to errors in inference. Even Gavin's criticisms of coding, many of which appear to me to be correct, are not without unintended consequence. Few among us will be equally attentive to cases that support both (or all) sides in an argument. The tendency is to identify errors in cases that, once repaired, are more likely to support one's own preferred conclusions. Thus, even if the error is genuine and the proposed correction correct, one could well end up degrading the quality of the inference by converting random errors in the coding of data into nonrandom (i.e., directional) errors, something that is much more difficult for any method of inference to address.[17]

Even as there is a tradition of precision in scholarship, there is also a tradition of searching for essence, of the core truth of a problem, context or situation. If this objective is lost in the pursuit of what might ungenerously be described as the 'little truths' of particular cases, then surely this is a high price to pay. As I will emphasize throughout this essay, my claim is not that certain approaches should be replaced, but that a big tent is a better one when it comes to the study of the important questions of national and international security. A pragmatic standard may be the best one to apply in applied social scientific research; let us use whatever works best in a given situation, and perhaps there are some places where quantitative analysis possesses considerable potential.

## III

By far the most common, and arguably the strongest critique of existing quantitative studies is that particular authors have mis-specified relationships (omitted or included the wrong set of variables, chosen the wrong samples, etc.), or that scholars have failed to code cases correctly. This represents the majority of the concerns raised by Gavin, and also others such as Sagan .[18] It may be useful to note that this is really not a criticism of method, but rather one of application. Noting that someone is 'doing it wrong' is a refutation of how one is doing something, not of what one is attempting to do. A bicycle poorly mounted is a problem of the cyclist rather than of bicycles.

Interestingly, rather that critiquing method, these criticisms serve to highlight (more) virtues of the quantitative enterprise. The purpose of empirical research is not to demonstrate truth — though some would have readers believe this — it is to eliminate mistruth. Hypothesis testing literally does not affirm but fails to reject, an outcome that reflects the tenuous nature of inference. All that one can say is that a relationship exists or does not exist under a given set of assumptions. A critic who rejects any of an author's assumptions has no reason to sustain the author's conclusions.

There is value in achieving even this limited form of clarity, however. The purpose of

[17] It can be argued that attempts to code one's own data — usually something lauded by the research community — are problematic in that researchers are inclined to code their cases in particular, supportive ways.

[18] Scott D. Sagan, "The Causes of Nuclear Weapons Proliferation." *Annual Review of Political Science* 14: 2011, 225–244.

statistical techniques is to solidify the process of hypothesis testing, not to ensure that particular applications are free of fault or even objection. The naive notion of science as truth is precisely wrong; science is a method of removing mistruth. It is also a considerable fallacy to confuse the method by which errors are detected more effectively with the origin of error. Thus, when a scholar like Gavin notes flaws in a quantitative study or studies, three things seem to attach to the criticism:

First, quantitative techniques are intended and often appear in practice to facilitate the process of developing objections to research. The ease with which a critic can identify sources of concern is a strength of the method, not a weakness. The specificity and detail with which quantitative researchers must detail their assumptions does not mean that they have more assumptions. Rather, the assumptions are laid bare to inspection. This is not an excuse for errors, even as the presence of errors is not an excuse for rejecting a method that makes it easier to identify errors. Clarity will often lay bare deficiencies, but we don't reject clarity.

Second, quantitative methods make it less difficult to address concerns about coding or specification. If you don't like what someone did, change it and see if the changes make a difference for the findings. Just as it is a mistake for a quantitative researcher to assert truth from a given set of results, so too it is a mistake to assert the faultiness of a conclusion from cases where errors can be shown to exist. One cannot know whether the results of particular cases obtain elsewhere. All theories involving human behavior are wrong in given instances (we do not expect these types of theory to be determinant). Tendencies require samples, not examples.

Third, quantification provides a common basis on which to adjudicate debates. This does not mean that scholars will always agree. Such a standard, like several applied informally in methodological discussions like this one, allude to a nonexistent reference point in which an approach, because it is not without fault, is deemed faulty. The appropriate standard is to inquire as to what alternative methods of adjudication are available and whether, in fact, one or another approach poses advantages. The deterrence debate alluded to by Gavin was resolved to the satisfaction of many scholars in international relations, though certainly not all. This was, and remains, a distinctive feature in a literature that seldom manages to find much more than its own intellectual tail. Many, though not all, objections can be raised, quantified and evaluated to determine which among existing alternative accounts is most nearly correct.

Uncertainty itself is an important argument for the utility of quantitative methods, not because quantification provides sufficiency, but because it may be necessary. Recent thinking emphasizes the problematic nature of inferences about the causes of war and peace when the causes of war involve uncertainty about whether war will take place.[19] For war to occur, participants themselves must not to know whether war will occur, or at least

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

one side must be mistaken in their beliefs about the onset, length or intensity of the looming contest.[20] To the degree that uncertainty is a cause of war, there exist important limitations on a researcher's ability to rely on, dissect, and interpret the historical record. The fact that key actors in history must have incorrectly evaluated key variables or processes for war to occur means that their descriptions must be treated with some caution.

A series of difficulties are perhaps best illustrated by referencing Gavin's own method in critiquing the studies previously mentioned. I highlight these concerns not as a challenge to Gavin's work, which I have already said is polished, insightful, and probably correct, but to make slightly clearer the potential pitfalls in even the best conceived and executed historical research:

> As a historian interested in these questions, my way of assessing Sechser, Furhmann, and Kroenig's arguments is straight-forward. I would identify the most important example where these issues are engaged, look at the primary documents, see how the authors coded crucial variables and determine how good a job their analysis does in helping us understand both the specific crisis itself and the larger issues driving nuclear dynamics. Political scientists might describe this as running both a 'strong' and a 'critical' test; in other words, if the authors' theories don't fully explain the outcomes and causal mechanisms in the most important and most representative case, how useful are the findings in explaining the broader issues?

What makes an example important? Clearly, at minimum, to be important, an event must be an event, a happening that actually occurs. Given what we think we know about war, however, it is very likely that many of the most important cases never occur at all and thus are unavailable for evaluation, precisely because the issues at stake were recognized to be potent enough that discretion was chosen by participants over a challenge. The problem of non-events is a crucial issue in studying deterrence and is presumably no less vexing for historians studying the nuclear age. The fact that an event, crisis or whatever happened suggests that at least one side in a conflict did not think the issues at stake were important, or at least that they underestimated how an opponent would react.

How does one validate a case, particularly in a stochastic world where causality is multiple, and contingent and where uncertainty is thought to be a critical causal element? Gavin's is a 'best case' version of a case study; I am inclined to believe his interpretation of events. However, it is not impossible that his take on things is shaded in a particular direction. My faith in the validity of his inferences rests heavily on my confidence in his judgment. This itself is tricky. Historians will note that the facts are there in the historical record, but the question is which facts are chosen and how are they weighted. Every analysis is an assessment of what matters and what does not, and what matters more than other things

[20] Erik Gartzke, "War is in the Error Term." *International Organization* 53(3): 1999, 567–587.

that also matter. The emphasis given particular elements of a historical case is art, not science, not because it is not true, but because it depends on judgment that cannot be replicated by others using the same facts, unless they just happen to share the same prejudices and biases.

It may be useful to consider the situation in terms of Jervis's discussion of misperception.[21] Knowing that others misperceive is not in itself a problem, provided that one makes allowance for the errors and provided perhaps that others recognize that they have this problem. The danger lies in those who do not know, who believe incorrectly that their own (or others') judgments are neutral and objective, or conversely in those who believe that they or others are biased, when they are not. Politics is never just about facts, about the record, but about how facts are interpreted. While statistical estimation can be quite crude in the treatment of these issues, it has the important advantage of being overt and replicable. It forces the researcher to make his or her weighting explicit, even as it removes some of the burden of guessing by estimating these effects mechanically.

Evaluating the impact of different factors is made considerably more difficult when the subject matter requires that the participants misperceive, or at least err, such as in war causation. The challenge is not just to infer intent and action, but to correctly identify misperception, error, excess, or insufficiency. If scholars are infallible, then there is no basis for concern.[22] However, the presence of errors as a necessary causal condition says that the historical record itself must be suspect. Participants are not going to give a complete, impartial and accurate account of what happened. Indeed, theory strongly suggests that those in power must have made errors for war to occur, errors that participants may wish to discount or obscure in contemporary or future description.

Finally, how often is a case at once both most important and most representative? I do not mean to be harping on semantics; there is more than just a rhetorical critique in such a challenge. The president of the United States is an important person, but for this very reason he is likely to be different in critical ways from the average citizen. Note for example scholarly fascination with the First World War. The Great War was important, but precisely because of this it does not represent other wars or their causes. Indeed, it was atypical in many ways, not least because it was extremely large.[23] The intensity of a contest can be affected by an enormous number of factors that have nothing to do with the reasons for war onset. The first days of World War I appeared to contemporaries much like preceding

---

[21] Robert Jervis. 1988b. "War and Misperception." *The Journal of Interdisciplinary History* 18(4) 1998b :675–700.

[22] I once pressed Jack Snyder that surely his view of politics meant that scholars, too, must make biased decisions. His response was that academics were smarter than politicians. I will not add any additional comment.

[23] The distribution of wars by intensity follows a power law, a common feature of natural processes. See Lars-Erik Cederman, "Modeling the Size of Wars: From Billiard Balls to Sandpiles." *American Political Science Review* 97(1): 2003, 135–150.

contests, so much so that many discounted what was to become the largest conflagration in history (up to that time). 'We will be home by Christmas' became an unfortunate cliché, a sentiment in some tension with the belief that primary documents serve to shed light on the true causes of the contest. Then too, the temptation is to give more weight to sources that turned out to be correct, something that contemporaries clearly did not do.

Human beings have a natural affinity for the remarkable. The evening news does not tell one much about how the world works precisely because it focuses on events that are at odds with the normal functioning of institutions, populations and politics in general. We want to know why World War I happened, but may miss the causes of war generally — and may even miss indicators of the next big war — if we rely on evidence that is itself idiosyncratic or which relies on an association between events that coincide with, but did not cause, warfare or its expansion, duration, etc.

Finally, I do not believe that political science would characterize Gavin's core criticism, though important and deserving of attention, as constituting a strong or critical test. The nature of such a test is the belief that, if a claim works anywhere, it must work in this instance. There is again a common tendency to conflate the notable with the indicative or typical. Outliers are a bad test of theory, unless they are placed in the context of the larger sample. At the same time, the stochastic nature of causal claims in the social sciences in general, and for theories of war in particular, mean that, with certain exceptions, no case is likely to serve this function. Events can always occur that derail cause and effect. Stated in a stronger way, the need for uncertainty as a condition for war must mean both that war can occur and that it can fail to occur for precisely the same set of observable conditions. It is inconsistent with the theory of war to claim that war must result in some instances and not in others, unless war is in fact the product of these discrete conditions and not due to uncertainty about the status of these conditions, as is increasingly believed.

The reaction by quantitative and theoretically-oriented researchers to recent studies by Jack Snyder and Erica Borghard and Trachtenberg helps to highlight this concern.[24] Both Snyder and Borghard and Trachtenberg raise doubts about the empirical validity of audience cost theory in international relations. Audience cost theory, originally proposed in a somewhat off-hand manner by James D. Fearon, and later clarified by Alastair Smith, Schultz and others, argues that leaders constrained by domestic public opinion are more likely to be able to credibly signal resolve, less likely to bluff, and therefore better able to avoid the bargaining failures that lead to war.[25] Both Trachtenberg and Snyder and

[24] Jack Snyder and Erica D. Borghard, "The Cost of Empty Threats: A Penny, Not a Pound." *American Political Science Review* 105(3): 2011, 437–456; and Marc Trachtenberg, "Audience Costs: An Historical Analysis." *Security Studies* 21(1): 2012, 3–42.

[25] James D. Fearon, "Domestic Political Audiences and the Escalation of International Disputes." *American Political Science Review* 88(3): 1994, 577–592; Alastair Smith, "International Crises and Domestic Politics." *American Political Science Review* 92(3): 1998, 623–638; Kenneth A. Schultz, "Domestic Opposition and Signaling in International Crises." *American Political Science Review* 94(4): 1998, 829–844; Schultz, "Do Domestic Institutions Constrain or Inform?: Contrasting Two Institutional Perspectives on Democracy and War." *International Organization* 53(2): 1999, 233–266; and Schultz, *Democracy and Coercive Diplomacy*.

Borghard take audience cost theory to task for the lack of evidence, which is indeed a valid problem for the theory.[26] However, the method of criticism — involving attempts to look at instances where audience costs 'should' have been present — has received considerable pushback, even among those skeptical about audience cost theory.[27] In short, the absence of evidence is not evidence of absence.

Once these factors are considered, it is by no means clear that a straight-forward approach to assessing arguments is destined to yield straight-forward conclusions. Indeed, much of the justification for convoluted theory and assessment in the social sciences is based on the fact that a straight-forward approach, though appealing on many grounds, does not work. To repeat, this is not a criticism of the substance of Gavin's objections, or indeed of those of others. The scholars whose work has been criticized should respond, justifying their coding decisions and making clear whether changes in these decisions alter their results. Still, under many, possibly most, circumstances in international security, counter examples, even ones involving extremely important cases, will be both ubiquitous and far from definitive disconfirmation of a theory. The stochastic, multi-causal and indeed indeterminant nature of conflict processes suggests perhaps that causation will be clearly supported in no individual case, even when causal mechanisms are actually operating as predicted.

**IV**

There can be no question but that considerable room remains to conduct research on all fronts in studying something that may well, eventually, cause the end of humanity as we know it. Asserting a role for quantitative nuclear security research says very little about the value of other approaches. It simply suggests that there is room for the 'big tent' that many of us believe in fact can prove most fruitful. Part of this is reflected in the inherent insufficiencies of all available methods. Criticism often takes the form of identifying where some alternative falls short, while ignoring similar or comparable deficiencies in alternative methods. As before, I intend no slight or confrontation in referring directly to Gavin's comments:

> [I]t may not even be possible to code resolve ex ante: the crisis itself reveals that one or both sides misread the other's resolve, which is the very reason for the crisis in the first place. In other words, it is the crisis behavior that reveals resolve, not the other way around.

There is a basic tension in criticizing attempts to code attributes of crisis bargaining such as which state was the initiator, had more to lose/gain, or was more resolved because the

---

Cambridge: Cambridge University Press, 2001.

[26] Kenneth A. Schultz, "Looking for Audience Costs." *Journal of Conflict Resolution* 45(1): 2001, 32–60.

[27] Erik Gartzke and Yonatan Lupu, "Still Looking for Audience Costs." *Security Studies* 21(3):2013, 391–397.

participants themselves did not know these things, while at the same time asserting that such an effort at coding can legitimately form the basis for criticism of an approach. Acknowledging that motives are complex or indistinct does not clearly lead to the conclusion that one should abandon a method that is permissive of ambiguity in favor of one that is much less flexible in its need for precision.

Put another way, if coding cases is hard, are we better off as a research community coding more cases or coding cases more carefully? This is not actually a question to which there is a ready answer. However, two possibilities present themselves. First, the tradeoffs between these alternatives are perhaps minimized if we attempt both approaches, determining the effectiveness of either approach by what we are able to find in doing so. Second, to the degree that one alternative must be chosen, there is reason to believe that quantitative imperfection may trump qualitative precision. Precisely because coding is likely to be fraught with error, a method that handles error relatively well is to be preferred. As in so many things, the enemy of the good may be the search for the best. The attempt at refutation of inference from a sample by identifying flaws in the coding of an example, while intuitive, makes no more sense than rejecting quantitative analysis because of one or a few studies that are said to fall short. Indeed, it highlights the demanding nature of qualitative inference. The external validity of claims is only as good as the generalizability of one's cases, which itself depends on an intimate understanding of the properties of the larger sample. For centuries, research on international security was hampered by a simple but important omission; researchers simply ignored all the many non-wars in inferring the properties of war causation. It took 400 years of the analysis of war before someone said, clearly, that the cause of war was not about power but about what actors believed about power, something that seems so obvious today.

In sharing the intellectual domain, quantitativists have a great deal to learn from their qualitative colleagues who, after all, have been toiling away for decades if not centuries. Rather than asking what quantitative analysis has done for the subject lately, let us ask where we stood a decade or more ago. Qualitative analysis was unaccompanied (or unopposed) for much of the nuclear era. It is not as if scholars using traditional methods were prevented from figuring out how nuclear security worked. Important advances were made throughout this period. Still, much remained to be resolved, either because the evidentiary record was unclear or because it could be interpreted in more than one way. Quantitative analysis is very good at finding relationships in areas where trends are not obvious, where cross-cutting effects are present or where multiple factors intercede.

There is also a false consciousness of qualitative opposition to quantitative methods. Quantitativists benefit from the careful insights and coding recommendations of qualitative researchers. Qualitative researchers may benefit in turn from feedback about what assumptions seem most valid systematically and which are perhaps more anecdotal. There will be little progress in any field of inquiry without this interaction, just as there will not be progress without the melding of deductive and inductive approaches. Increasingly, the best research is becoming multi-method.

It has been roughly two and one half millennia since Thucydides revolutionized the study of international relations and security policy with qualitative analysis fortified by rational causal reasoning. In the interim, students of world affairs have used qualitative analysis to resolve numerous riddles. At the same time, many questions remain to be answered. The passage of 2500 years and the fruits of many scholarly lifetimes suggest two things. First, that Thucydidian qualitative analysis, by itself, is at best not a blindingly rapid solution to endemic global problems. Second, that it may be a bit soon to judge the potential of alternatives or complements, particularly those introduced in the last generation or so. What is the hurry to condemn a method that is still in its infancy? Indeed, to the extent that qualitative approaches have been successful, they are bound to have consumed some of the territory in which their advantages are most distinct. What remains is not a virgin forest but one in which the easiest pickings have already been exploited. What remains disproportionately are questions that have proven difficult for qualitative analysis alone to resolve.

There are differences in the respective enterprises. Qualitative research is inspired at least in part by the conviction that the precise representation of one or a few cases can elicit relationships of interest. Quantitative analysis relies on numbers and approximation to accomplish the same objective. Rather than viewing one another with suspicion, it may be best to move forward in the mutual acknowledgment that each approach is imperfect and that the best opportunity for real insight depends on a more virtuous and reaffirming interaction than has taken place to date.

**Erik Gartzke** is a Professor of Government at the University of Essex and an Associate Professor of Political Science at UC San Diego. Gartzke's primary area of study involves the impact of information and institutions on war and peace. He applies bargaining theory, rational choice institutional theory, concepts of power and social identity, and statistical analysis to four substantive areas of interest: 1) The liberal peace, 2) international institutions, 3) diplomacy, and 4) the system overall. Examples of Gartzke's recent publications include "The Determinants of Nuclear Force Structure,'' with Jeffrey Kaplow and Rupal Mehta, *Journal of Conflict Resolution*, 58(2); "Under Construction: Development, Democracy, and Difference as Determinants of the Systemic Liberal Peace,'' with Alex Weisiger, *International Studies Quarterly*, 58(2); "The Myth of Cyberwar: Bringing War on the Internet Back Down to Earth,'' *International Security*, 38(2):41—73; "Permanent Friends?: Dynamic Difference and the Democratic Peace,'' with Alex Weisiger, *International Studies Quarterly*, 57(1):171-185; "Trading on Preconceptions: Why World War I Was Not a Failure of Economic Interdependence,'' with Yonatan Lupu, *International Security*, 36(4):115-150. He holds a Ph.D. in Political Science from the University of Iowa.

## "The Use and Abuse of Large-n Methods in Nuclear Studies"[1]

An Essay by **Vipin Narang**, MIT

The reemergence of the study of nuclear weapons in political science coincides with the widespread diffusion and use of increasingly sophisticated quantitative methods, including large-*n* regression analysis and the so-called revolution in causal inference. Improvements in these two methodological areas carry with them the promise of more systematically isolating correlates or, in the latter case, the promise of the holy grail of identifying causal variables of either nuclear proliferation or deterrence. What are the promises and limits of these methods as applied to nuclear studies? In general, as with any method, it is important for scholars to be transparent about the promise, and especially the limitations, of their choices. In order to get published in the pages of *Journal of Conflict Resolution* (*JCR*), *American Political Science Review* (*APSR*), or *International Organization* (IO), scholars have to sell their results to reviewers and editors. But often, that comes at the expense of transparency in terms of limitations, and there has been a disturbing trend in nuclear studies of quantitative work to sometimes 'oversell' the robustness and magnitude of results to an unnecessary degree. This perceived overselling of results has had the perverse effect of potentially undermining and discrediting the larger enterprise because the results are simply unbelievable to most specialists, including not only other political scientists and methodologists but also historians and international relations (IR) theorists. This is unfortunate because this trend has overshadowed the positive effects that the large-*n* enterprise *has* had on the study of nuclear proliferation and deterrence. This essay briefly outlines the promise and limits of the large-*n* enterprise in nuclear studies as I see it, as well as examining how the growing 'causal inference revolution' affects nuclear studies in political science.

### *The Large-n Enterprise in Nuclear Studies*

The incorporation of large-*n* methods in the study of nuclear proliferation and deterrence has been motivated by some very good reasons that are worth enumerating. Several of the most important methodological problems that a large-*n* research design is intended to solve are as follows:

**The data structure helps avoid selection on the dependent variable**. One of the biggest limitations of case-study work in the study of nuclear proliferation and deterrence was the tendency to select on the dependent variable, or those cases where we only observed, as with the case of successful proliferation. This has the potential to bias our understanding of the causes of proliferation because no-variation designs can erroneously identify factors that are believed to be important but which, in fact, are not because they are also present in the negative cases. By sampling the full universe of actual proliferation as well as nonproliferation, the large-*n* enterprise is in principle able to more systematically isolate

[1] I thank Mark Bell, Christopher Clary, Erik Gartzke, Frank Gavin, Robert Jervis, Nick Miller, and Scott Sagan for their comments and feedback on earlier drafts.

correlates of proliferation (or deterrence/compellence) than single or limited case study methods.[2]

**Large-*n* analysis enables multivariate analysis.** The large-*n* enterprise allows scholars to analyze the relative importance of many variables at once. This is particularly important in outcomes like nuclear proliferation or deterrence/compellence that are often not monocausal or monocorrelated. As these methods and approaches have improved, conditional and interaction hypotheses can also be tested in ways that would be difficult in single or small case study method analysis.

**Large-*n* analysis allows for more precise and standardized metrics for correlates that 'matter'**. By being forced to quantify independent and dependent variables and subject the analysis to econometric techniques and standards, the large-*n* enterprise carries with it the promise of providing more precise estimates of which independent variables are significantly correlated with an outcome of interest and how much they matter. This allows the debate to progress from 'security environment matters' to 'countries that face a conventional military disadvantage of some X amount against a bordering state are Y times more likely to acquire nuclear weapons.' This effort toward precision and measurement allows scholars to directly and more precisely engage theory, empirics, and each other on the same terms.

**Large-*n* data and analysis enables replicability.** One of the biggest advantages of the large-*n* enterprise is the ability for peers to know exactly how to replicate the reported results if provided the original dataset and procedures for the analysis. This allows the codings and assumptions in the analysis and models to be laid bare for other scholars to see, and is very much in the normal science tradition. This transparency has led to debates—whether productive or not—about whether reported results are particularly robust or fragile and, in the process, represents some degree of progress in our understanding of the correlates associated with a particular nuclear outcome of interest. Though subjectivity is still injected in the original coding of covariates in many cases, there is less subjectivity in the reporting of results, and other scholars have the ability in theory to reproduce and thereby probe them from the same playing field.

**Quantitative analysis allows researchers to explicitly model probabilistic processes.** Most of our theories in international relations, but particularly nuclear security, are probabilistic. But modeling probabilistic processes in qualitative work is very difficult since the case study method necessarily truncates the sample size. As a result, the large-*n* enterprise carries with it the ability to explicitly and systematically model probabilistic processes and estimate these probabilities (for example, probability of proliferation or deterrence success) more rigorously than the case study method.

[2] See, for example, Erik Gartzke and Dong-Joon Jo, "Determinants of Nuclear Weapons Proliferation," *Journal of Conflict Resolution*, vol. 51, no. 1 (February 2007), pp. 167-194; and Sonali Singh and Chris Way, "The Correlates of Nuclear Proliferation: A Quantitative Test," *Journal of Conflict Resolution*, vol. 48, no. 6 (December 2004), pp. 859-885.

These are non-trivial methodological advantages that medium and large-*n* methods can offer to scholars of nuclear security. But they must be balanced against a variety of pitfalls that accompany large-*n* analyses, particularly in the area of nuclear proliferation and deterrence/compellence. We must be transparent about the following limitations and what they mean for how we report and interpret results.

Several critical limitations to be aware of are as follows:

**Large-*n* data in international relations often employs very poor measures.** There is an old adage from information sciences: garbage in, garbage out. That is, the output is only as good as the inputs. Too often, the measures deployed in quantitative IR studies, but especially nuclear studies, are total garbage. Some of the variables used to measure critical concepts such as 'power' (for example, the Composite Index of National Capability (CINC) score, of which I am also guilty) may be the best out there given the herculean task of coding these variables for every country in every year, but no one really believes it measures what we want it to. We can make arguments for why they are good proxies, but they are a tough sell. We treat some measures as if they can be consistently operationalized over two centuries and are temporally stable, but no one really believes that that is the case. In nuclear studies, we have often collapsed nuclear weapons states into a binary condition even though clearly some nuclear states are different than others in important and measurable ways. Recent efforts to analyze concepts such as 'nuclear superiority' which move beyond the simple binary condition of nuclear/nonnuclear are important, but *post hoc* measures of aggregate numbers of warheads fail to measure what theorists and practitioners really meant by superiority, a concept which includes variables such as relative geography and number of targets to hit and a psychological component which is simply very difficult to quantify. So if our measures and inputs are trash….

**One has to be exceptionally careful with analysis of rare dependent variables**. We are often trying to analyze the causes and correlates of very rare events: nuclear proliferation, deterrence and compellence success, nuclear assistance, stability/instability paradox. Gathering millions of observations when the dependent variables occur only, say 10 times, presents very real mathematical challenges to the standard models we deploy (especially logit/probit). Although, for example, rare events logit (ReLogit) techniques exist to correct for the low incidence of some of these dependent variables, these approaches are usually not implemented or deployed correctly. Furthermore, if a dependent variable is relatively rare, it is almost certainly the case that reported results are quite sensitive to coding decisions and/or noise or miscodings. It is a mathematical truth that all results break at some point. When the dependent variable is as rare as it tends to be in nuclear studies, the results tend to break early and often. We need to be better about reporting the point at which they break and allow readers to judge whether that point constitutes a robust breaking point.

**Large-*n* data structures and analysis must better account for time dependencies**, **and dependencies in general.** There are virtually no available techniques that model time, or uncorrelate temporal dependencies, accurately enough to really convince readers that 'country-year' datasets as they are often deployed in quantitative nuclear studies really represent independent observations year-to-year. Splines or $t, t^2, t^3$ are important corrections but it is clear that they are not generating independent observations for every country-year (the independent and identically distributed (IID) assumption for the standard errors is almost certainly violated). This means that we are not modeling history and repeated interactions accurately in the large-*n* enterprise. Conflicts and proliferation behavior have a history and operate through feedback loops that are not being captured by these modeling approaches, meaning that the estimated relationships and findings between the independent variables and dependent variable of interest are almost certainly erroneous. In addition to temporal dependencies, the problem of other dependencies (such as spatial) creates a 'degrees of freedom' problem if our millions of supposedly independent observations are actually only tens of independent observations. This has significant implications for the validity and unbiasedness of reported results since treating all these dependencies as independent observations massively and artificially deflates the standard errors, meaning our 'true' level of uncertainty is definitely much higher than is ever reported in the pages of *IO*, *JCR*, or *APSR*.[3]

Furthermore, we have to be very careful about how we code time-series dependent variables. If the theory and outcome of interest is the *onset* of a particular nuclear phenomenon, then the dependent variable needs to be coded 1 for the first instance of the outcome and then the observations should be dropped from the dataset; coding them as 1 for every country-year after onset results in an estimate for *duration*, not onset, since these are all temporally correlated. Finally, the temporal dependence of most of our time-series cross-sectional (TSCS) data makes the problem of reverse causality much more complicated than lagging the IVs by a year.

**Regression analyses have strong assumptions about functional form/linearity.** Almost all econometric models thus far deployed to analyze large-*n* datasets in nuclear security studies have very strong assumptions about the distribution of independent and dependent variables, the distribution of error terms, and assumptions of linearity. Almost all of these assumptions are egregiously violated by the nature of our data. Frankly, many of the processes of interest are in fact decidedly not linear but either cumulative or follow completely unpredictable or unknown distributional forms since they are inherently strategic. So, many analysts simply accept this but continue on, arguing that linearity is as good an assumption as any. But this is almost certainly wrong, even if nothing else is right. The implication is that we may not even know if the findings in the large-*n* work are real findings, or how to project these findings out as predictions (that is, increasing a border conventional threat by 5x increases the probability of proliferation by 7x. Well, maybe, maybe not. The assumptions built into that prediction are manifold and assumes some functional relationship that may in fact be completely wrong given the above concerns, but also may follow threshold or non-linear processes).

[3] See Gary King and Margaret E. Roberts, "How Robust Standard Errors Expose Methodological Problems they do not Fix," Unpublished Manuscript (November 27, 2013). Available at http://gking.harvard.edu/files/gking/files/robust.pdf.

**There is still no good way to deal with selection effects or strategic behavior.** While we often claim to have the 'full universe' of observations in large-*n* analyses, they are usually actually not the full universe. Instead, they are often a nonrandom selected-into set of observations (e.g. crises, nuclear suppliers, etc). One could attempt to deploy a family of selection models that involves several stages, but which require essentially finding an instrument or exclusion criteria that is ultimately uncorrelated with the final outcome of interest (say nuclear deterrence). The first stage exclusion criteria is extremely difficult to find and almost everything deployed in the extant literature is clearly correlated with the final outcome variable. Finding instruments is hard, so this is not a surprise, but we should not brush the difficulty of the selection effects problem in even large-*n* analysis under the rug. The large-*n* enterprise can solve the truncated observation problem, but it cannot solve the selection effects problem if *all the observations* are themselves subject to a selection criteria.

**Scholars in the quantitative enterprise must be better about performing and reporting sensitivity analyses.** As noted above, all analyses break at some point, but the large-*n* enterprise could be much more transparent about when results break and allow the reader to judge whether the results are sufficiently robust at that breaking point. For example, it is very different for results to be sensitive to whether, for instance, the 1912-1913 Balkan Wars are coded as a single or multiple wars versus whether France would have to not exist. Or, in nuclear studies, whether coding the 1999 Kargil War as a war or not breaks the results.

***Nuclear Studies and the 'Causal Inference Revolution'***

A growing proportion of methodologists in political science have increasingly shifted from the analysis of big observational data to focusing on methods of causal inference. Because the analysis of observational data, as noted above, often only reveals correlations, not causation, and because of the numerous threats to causal inference in observational data (e.g. endogeneity, reverse causality, significant distributional assumptions, model dependence, etc), there has been a shift in the past several years to focusing on identifying causation, rather than identifying simply correlates. The gold standard of causal inference is randomized experiments with a single treatment randomly assigned to a treatment group and with a placebo assigned to a control group. There is also a class of quasi-experimental studies that rely on observational data to try to 'match' treatment and control observations to extrapolate causal relationships, but one of the very hefty assumptions is that treatment and control observations are randomly (and not systematically) different.

How do these new methods fit with the study of nuclear weapons, or more generally, macro-phenomena that we cannot manipulate as researchers? The answer is, unfortunately, not very well. We cannot randomly assign nuclear weapons to some states and not others in order to identify their causal effect on deterrence success. We cannot randomly assign security threats or domestic political incentives in order to discern the causal variables for proliferation. And there are few—if any—believable natural experiments that do so. Endogeneity and complicated multicausality—which present a problem for identification strategies that often require a single identifiable and manipulable 'treatment' since the research designs required for multiple treatments are complicated—are facts of life both in the study of nuclear weapons and most *strategic* decisions and interactions in international relations. It is very difficult to identify exogenous variables or shocks that are uncorrelated with the causes or consequences of proliferation that might enable so-called natural experiments either.

There are of course exceptions. These methods, particularly survey experiments, can be quite powerful in identifying individual level preferences, such as mass opinion related to nuclear weapons.[4] They can be incredibly useful in identifying microfoundations at the individual level, since sample sizes can be large, and the 'treatment' (especially when it comes to things like public opinion) can be manipulated. But this is a limited class of questions and problems for which these methods might be useful. On the bigger questions of the causes and consequences of proliferation, there are few plausible identification strategies because of the strategic and macro nature of the phenomena and the multiple, complex pathways through which they work. Thinking about the developing trends in political science and how nuclear security studies fits into them is important if we aim to remain an integral part of political science and be able to engage with other subfields. This is one trend in which nuclear security studies future place is highly uncertain.

Ironically, the causal inference revolution in quantitative methods may lead to a resurrection in the discipline's valuation of qualitative methods in nuclear security, since qualitative methods in this particular area are much better suited to identifying and teasing out causal mechanisms and processes than the big-data enterprise. To paraphrase Jasjeet Sekhon, one of my own former methods professors: we know more about the foundations of democratic peace theory from Kant than from the thousands of pages of dubious quantitative work on the subject. The view of a purist is: deploy quantitative methods correctly or do not deploy them at all (much like SSBNs?). In nuclear studies, qualitative methods probably do a better job of illuminating the 'selection into treatment' than quantitative methods. They may not be able to estimate the average treatment effect of our treatments of interest (nuclear weapons, security pressures, etc), but neither can quantitative methods in nuclear security issues. However, with careful case selection and research design, qualitative methods can at least allow us to convince ourselves—and others—that in nuclear security studies, there are non-zero treatment effects.[5] Indeed the conclusion for nuclear security studies with the causal inference revolution should be that research design trumps methods. More than the divide between qualitative and quantitative methods, this trend should be a wakeup call that as scholars and students of nuclear proliferation, we should pay much better attention to general research design than we presently do.

### *Conclusion*

---

[4] See Daryl G. Press, Scott D. Sagan, and Benjamin A. Valentino, "Atomic Aversion: Experimental Evidence on Taboos, Traditions, and the Non-Use of Nuclear Weapons," *American Political Science Review*, vol. 107, no. 1 (February 2013), pp. 188-206.

[5] I thank Mark Bell for highlighting this point.

Does all of this mean we should abandon the large-*n* or quantitative enterprise in nuclear studies? Absolutely not. What it does mean is that we simply need to be more transparent and honest about the limitations of the enterprise, which will more credibly allow us to argue what insights it *does* provide. The quantitative enterprise today finds itself in a somewhat untenable position: under fire from both methodologists and the theorists and historians of nuclear security. It is difficult for an enterprise to survive if it continues to take fire from multiple fronts simultaneously. The quantitative enterprise offers some very real advantages over single or small case study methods, and these advantages are important. But because of the inherent limitations of the multitude of quantitative approaches to nuclear studies, they cannot be devoid of theory, lack an internal logic, or be ahistorical. The advantages of quantitative approaches can and should be supplemented by theory and qualitative empirics, suggesting that for nuclear studies the best pathway forward may be—instead of arguing about methods—accepting that a mixed-methods approach that deploys and exploits the advantages of both quantitative and qualitative methods would be more fruitful. For the quantitative enterprise side, good hard theoretical and qualitative work that teases out causal mechanisms and corroborating convincing evidence is critical to avoiding the charge that our results are nothing more than bad data mining.