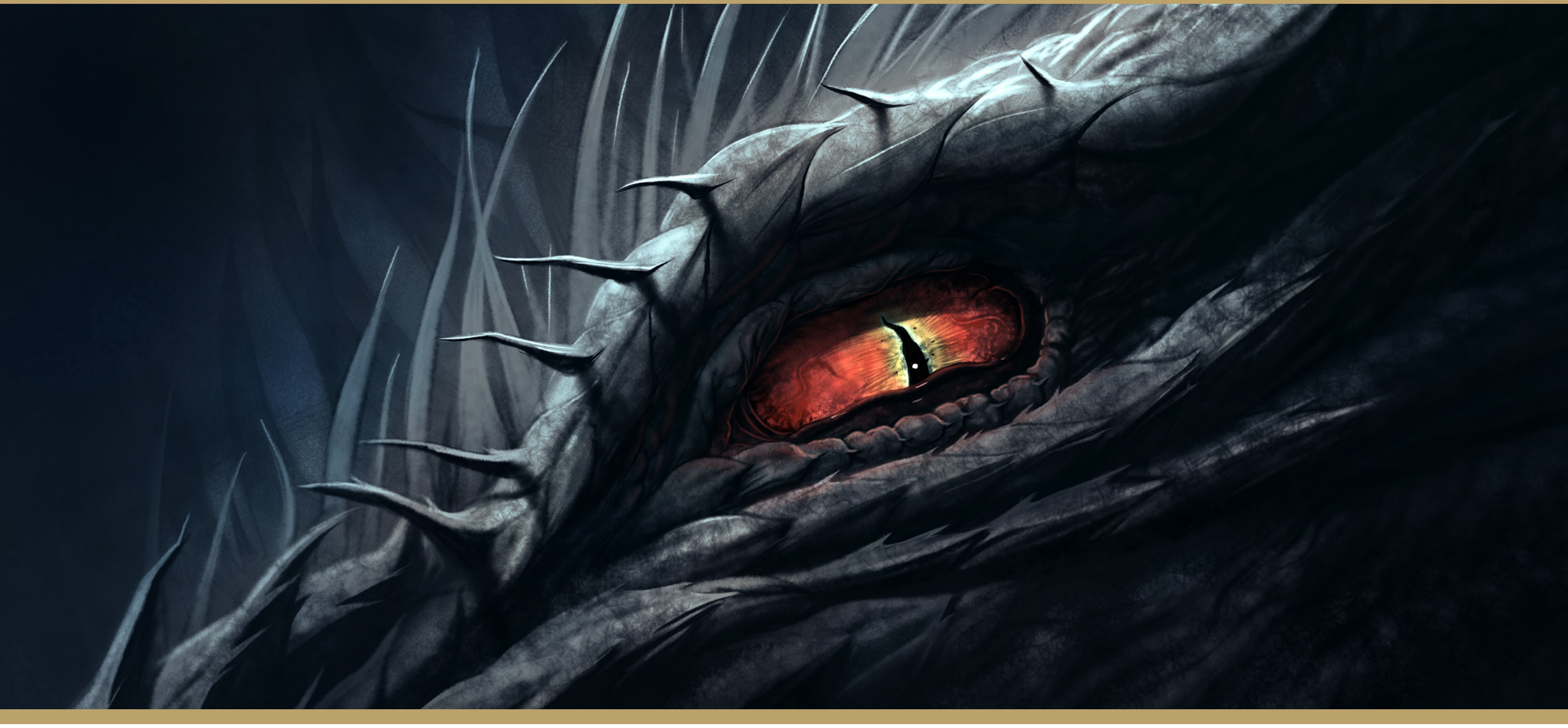
TICKLING THE SLEEPING DRAGON'S TAIL

Should We Resume Nuclear Testing?

National Security Report



Michael Frankel | James Scouras | George Ullrich

JOHNS HOPKINS
APPLIED PHYSICS LABORATORY

TICKLING THE SLEEPING DRAGON'S TAIL

Should We Resume Nuclear Testing?

Michael Frankel

James Scouras

George Ullrich

JOHNS HOPKINS
APPLIED PHYSICS LABORATORY

“Tickling the sleeping dragon’s tail” is a metaphor for risking severe consequences by taking an unnecessary provocative action. Its origin can be traced to the last year of the Manhattan Project at Los Alamos National Laboratory (LANL) in 1946. When investigating the critical mass of plutonium, LANL scientists usually brought two halves of a beryllium reflecting shell surrounding a fissile core closer together, observing the increase in reaction rate via a scintillation counter. They manually forced the two half-shells closer together by gripping them through a thumbhole at the top, while as a safety precaution, keeping the shells from completely closing by inserting shims. However, the habit of Louis Slotin was to remove the shims and keep the shells separated by manually inserting a screwdriver. Enrico Fermi is reported to have warned Slotin and others that they would be “dead within a year” if they continued this procedure. One day the screwdriver slipped, allowing the two half-shells to completely close, and the increased reflectivity drove the core toward criticality. Slotin immediately flipped the top half-shell loose with a flick of the screwdriver, but by then he had endured a lethal burst of fast neutrons. He was dead nine days later. Richard Feynman characterized the activities of the critical mass group as “tickling the tail of a sleeping dragon.”

NSAD-R-20-098

Contents

Figures

Image credits:

Figure 1. Worldwide Nuclear Testing. Data mostly from Arms Control Association, with modifications, “The Nuclear Testing Tally,” fact sheet, last reviewed July 2020, https://www.armscontrol.org/factsheets/nucleartesttally.

Figure 3. Subsidence Craters at the Nevada Test Site with the Sedan Ejecta Crater in the Foreground. From NNSA (National Nuclear Security Administration) Nevada Field Office, *Nevada Test Site Guide*, DOE/NV-715 Rev 1 (Las Vegas: NNSA, March 2005), 58, https://www.nnss.gov/docs/docs_LibraryPublications/DOENV_715_Rev1.pdf.

Figure 4. A Typical Target Chamber for a Horizontal Line-of-Sight Test. From Harold Drollinger, Robert C. Jones, and Thomas F. Bullard, *A Historical Evaluation of the U12t Tunnel, Nevada Test Site, Nye County, Nevada*, vol. 5, DOE/NV/26383–109 (Las Vegas: NNSA, February 2009), https://doi.org/10.2172/1010604.

Figure 5. Seismic Cavity Decoupling Strength as a Function of Frequency. Reproduced from John R. Murphy and Brian W. Barker, *A Comparative Analysis of the Seismic Characteristics of Cavity Decoupled Nuclear and Chemical Explosions*, PL-TR-95-2177/SSS-TR-95-14980 (Hanscom Air Force Base, MA: US Air Force Phillips Laboratory, 1995), https://apps.dtic.mil/sti/pdfs/ADA304812.pdf.

Figure 6. Sedan Crater at Nevada Test Site Excavated by a Plowshare Test. From NNSA (National Nuclear Security Administration) Nevada Field Office, *Nevada Test Site Guide*, DOE/NV-715 Rev 1 (Las Vegas: NNSA, March 2005), https://www.nnss.gov/docs/docs_LibraryPublications/DOENV_715_Rev1.pdf.

Figure 7. Castle Bravo Event. From Thomas Kunkle and Byron Ristvet, *Castle Bravo: Fifty Years of Legend and Lore, A Guide to Off-Site Radiation Exposures*, DTRIAC SR-12-001 (Washington, DC: Defense Threat Reduction Agency, January 2013), https://blog.nuclearsecrecy.com/wp-content/uploads/2013/06/SR-12-001-CASTLE-BRAVO.pdf.

Figure 8. Mk 21A Reentry Vehicle Damage after X-Ray Exposure in an Underground Nuclear Test. From Byron Ristvet, "A Brief History of US Nuclear Weapons Testing During the Cold War: 1947–1992," Defense Threat Reduction Agency briefing, September 11, 2013.

Figure 9. Calculated Electromagnetic Pulse Footprint from a Detonation over the United States. Originally from the Department of Defense. Appears in Michael J. Frankel, James Scouras, and George W. Ullrich, *The Uncertain Consequences of Nuclear Weapon Use*, National Security Report NSAD-R-15-020 (Laurel, MD: Johns Hopkins University Applied Physics Laboratory, 2015).

Figure 10. Exposure of US Troops during an Atmospheric Test in the 1950s. Federal Government of the United States, public domain, via Wikimedia Commons, https://commons.wikimedia.org/wiki/File:Exercise_Desert_Rock_I_(Buster-Jangle_Dog)_002.jpg.

Figure 11. Target Chamber of the National Ignition Facility at LLNL. From Lawrence Livermore National Laboratory.

Figure 12. The IMS for Verifying Compliance with the CTBT. Modified with permission from CTBTO, https://www.ctbto.org/fileadmin/content/verification/facilities/mapfacilities.html.

Figure 13. 1970 Baneberry Event Venting through Undetected Rock Fissure. NNSA (National Nuclear Security Administration) Nevada Field Office, *Nevada Test Site Guide*, DOE/NV-715 Rev 1 (Las Vegas: NNSA, March 2005), 57, https://www.nnss.gov/docs/docs_LibraryPublications/DOENV_715_Rev1.pdf.

Summary

This report addresses the questions of whether the United States should resume nuclear testing and, if not, whether it should better prepare to do so in the future. These questions involve high stakes with multiple, complex, and uncertain interacting considerations. Many publications that address these matters emphasize policy issues, presenting the geopolitical consequences that, as it is often argued, rule out contemplation of such a provocative step. Works that credibly argue the technical issues, both pro and con, are scarcer. Our goal is to provide a comprehensive and balanced consideration of all significant technical and policy arguments relevant to nuclear testing. After considering these arguments and relevant counter-arguments, we conclude that under present circumstances, the United States should not resume nuclear testing because of the lack of a compelling national security need combined with potentially significant negative geopolitical consequences for nuclear proliferation and reignition of a nuclear arms race. For now, let the sleeping dragon lie. Once it is awoken, there is no plan to put it back to sleep and there will always be the opportunity to awaken it in the future. At the same time, we identify a series of future technical and political developments whose occurrence would require revisiting our decision calculus. We end the report with recommendations to improve test readiness and, as a final thought, place the issue of whether or not to resume nuclear testing in the context of conflicting far- and near-term US national security goals.

The report starts with a short review of the international framework of treaties that have constrained and conditioned the evolution of US nuclear test history. The introduction goes on to define the scope of our analysis—it is limited to discussion of underground nuclear tests—and argues that addressing the question of whether to resume testing is not an esoteric intellectual exercise but rather relates to important issues of geopolitical stability and national security as the arsenal continues to age and new threats emerge. We then provide a synopsis of US nuclear test history that traces the evolution of the conduct of the nation's 1,054 nuclear tests. This section describes the trend over time from large—up to megaton-class—explosions in the atmosphere or underwater, generally in the Pacific, to much more modest-size devices exploded underground in the continental United States. It also describes the different test types conducted: weapon development, weapon effects, proliferation/treaty monitoring, peaceful nuclear explosions, safety, and stockpile maintenance. An important new retrospective highlights the role surprises have played in lessons learned and in justifying the need for nuclear testing and provides a taxonomy of the types of significant surprises encountered—surprises in weapon development and safety, vulnerabilities of military systems, and the nuclear weapon environments. We review the history of nuclear moratoria and the critical role played by the Stockpile Stewardship Program, which supports the present moratorium in lieu of testing. We also present the growing concerns of Stockpile Stewardship Program critics as nuclear warheads both age and are altered over time from their originally manufactured and tested configurations, and we express concern over the inherent conflict of interest at the heart of the present system of certification.

We then present the heart of this report, which are the arguments—both political and technical and both for and against—around the resumption of nuclear testing. Arguments for resumption of testing include the following:

- Underground testing is needed to underwrite deterrence by enabling the development of specialized lower-yield nuclear weapons.

- Confidence in the stockpile is eroding as the state of weapons changes over time.
- Nuclear adversaries could exploit imperfect monitoring capabilities and Comprehensive Test Ban Treaty (CTBT) ambiguities.

Arguments against resumption of testing include the following:

- Other states would inevitably also resume testing, making the world more dangerous.
- US nonproliferation leadership would be undermined, condemnation by the rest of the world provoked, and US bipartisan support for nuclear policy threatened.
- There is no need for new weapons; for now, existing ones suffice.
- The Stockpile Stewardship Program provides sufficient confidence in our stockpile.
- Underground nuclear tests will inevitably create health risks to civilian populations.

After discussing these arguments, we provide our analysis and bottom-line determination. We assess the three arguments in favor of resuming nuclear testing as unconvincing. We are not persuaded that stockpile maintenance needs nuclear testing at this time, and while it well might at some point in the future, it is difficult to predict when that might be. Similarly, while we support the development of low-yield nuclear weapons, simpler designs (or those previously tested but not weaponized) likely can be developed without testing, and even specialized new designs would have no need to be tested for years. Finally, we believe that the ambiguity in definitions of zero yield do not seem exploitable to achieve military advantage. By contrast, we are in strong agreement with the first two of the arguments against resuming nuclear testing. While all forecasts are speculative, we think there is a reasonable likelihood that resumption of underground nuclear testing will spur nuclear proliferation and, possibly, reignite a nuclear arms race, both to the potentially grave detriment of geopolitical stability and US national security. Thus, we conclude that, at present, the United States should not resume nuclear testing.

Next we consider possible future developments—both scientific and political—that could change our decision calculus. These possibilities include a nuclear war or crisis, the prospect of Russia or China breaking from the moratorium, unambiguous discovery of Russian or Chinese clandestine testing, discovery of a common-mode arsenal failure issue, failure to certify the arsenal or loss of confidence in the certification process itself, new international incidents of testing due to failure of the Nuclear Non-Proliferation Treaty, and the possible development of a new design imperative in the face of emerging threats.

We then review the state of test readiness should the United States actually decide to resume testing. We discuss the history of the safeguards for potential resumption of nuclear testing and assess as doubtful the United States' readiness to comply with relevant legislation. In no small part, this unpreparedness is due to the dissipation of human capital, as the cohort of scientists and technicians uniquely knowledgeable in the combined science and art of nuclear testing have left the scene and a new cohort will have to be reconstituted without familiarity with such nuclear "art forms" as post-detonation containment, grounding and shielding, and other ill-documented nuclear test arcana.

The report culminates with our recommendations for moving forward. First, the United States should consider relaxing its interpretation of the CTBT limits to be consistent with the Russian definition, which allows tests of very low yields. Second, US leaders should more openly acknowledge the limitations inherent in the Stockpile Stewardship Program, and its potential for failure, and develop plans to mitigate these limitations, including operational measures meant to accommodate weapons with uncertain reliabilities. Third, the United States should revamp its current nuclear stockpile certification process to mitigate inherent conflict of interest and embed in the process a standing, independent review body. Fourth, national leaders should take more seriously the possibility that the United States may choose to test in the future and take specific steps that increase the nation's ability to execute the requirements of Presidential Decision Directive 15 (PDD-15).

Our final thought recasts the question of resuming nuclear testing in the context of conflicting far-term and near-term US goals. In the far term the United States seeks a world without nuclear weapons, but in the near term, with no viable alternative in a world with nuclear-armed adversaries, it seeks to maintain the efficacy of nuclear deterrence. Resumption of nuclear testing now would undermine the nation's far-term goal without substantially contributing to its near-term goal.

Introduction

With nuclear weapon tests widely accepted as an essential element of national strategy during the Cold War, the United States has conducted over one thousand tests of various types and purposes, starting with Trinity in 1945 and ending with Divider in 1992. US nuclear weapon testing is now constrained by four treaties in force:[1]

(1) The Limited Test Ban Treaty (1963), which prohibits nuclear testing in the atmosphere, in outer space, and under water;

(2) The Outer Space Treaty (1967), which extends the nuclear testing ban to the moon and other celestial bodies;

(3) The Threshold Test Ban Treaty (1974), which identifies allowed nuclear test sites and prohibits underground nuclear weapon tests with yields greater than 150 kilotons, as well as those that produce radiation that extends beyond the national boundaries of the testing state; and

(4) The Peaceful Nuclear Explosions Treaty (1976), which bans individual nuclear explosions with total yields above 150 kilotons and group explosions with total yields above 1,500 kilotons at locations not specified as test sites under the Protocol to the Threshold Test Ban Treaty.

In combination, these treaties allow only underground nuclear tests at designated test sites with yields below 150 kilotons, as well as "peaceful" nuclear explosions with the same yield threshold. This threshold of 150 kilotons is based on the United States' concern in the mid-1970s about its inability to use national technical means to verify compliance of lower-yield underground testing.

In addition, the United States has been operating under several policies that further constrain testing. In 1992 it adopted a self-imposed moratorium on all underground nuclear weapon testing. Subsequently, in 1996, the United States signed, but has not ratified, the Comprehensive Test Ban Treaty (CTBT), which would prohibit all nuclear explosions for any purpose, of any yield,[2] in any environment. Although the Senate rejected ratification of the CTBT in 1999,[3] as a signatory the United States may or may not be obligated under international law to avoid taking steps that would violate this treaty.[4] Finally, Article VI of the Nuclear Non-Proliferation Treaty (NPT), which states that "each of the Parties to the Treaty undertakes to pursue negotiations in good faith on effective measures relating to cessation of the nuclear arms race at an early date and to nuclear disarmament, and on a treaty on general and complete disarmament under strict and effective international control," may or may not impose a duty to stop nuclear testing under a comprehensive test ban.

The cumulative effect of these policies is that, since 1992, the United States has not conducted nuclear tests that involve a sustained nuclear chain reaction. It essentially adheres to its interpretation

---

[1] Woolf, Kerr, and Nikitin, *Arms Control and Non-Proliferation*.

[2] The CTBT, which never came into force, was a so-called "zero-yield" treaty. There is a long evolutionary development of the meaning of zero yield, which often specifies a threshold number of energy release or a threshold number for a permissible multiplication factor of the nuclear chain reaction, or a requirement that fission energy be less than the chemical energy released by the high explosive. The present US definition articulated by the US Department of State avoids specific numbers and instead "prohibits all nuclear explosions that produce a self-sustaining chain reaction of any kind." The omission of a specific definition of scope in the treaty was a deliberate decision the negotiating parties, including the United States, made to ensure that no loopholes were created by including a highly technical and specific list of what activities were and were not permitted under the treaty. Note that zero yield, even according to the United States, does not mean zero nuclear yield, as it permits less than self-sustaining nuclear chain reactions. AVC, "Scope of the Comprehensive Nuclear Test-Ban Treaty."

[3] Arms Control Association, "Senate Rejects Comprehensive Test Ban Treaty."

[4] Rogoff, "International Legal Obligations."

of the provisions of the CTBT. The questions we address in this report are straightforward, although their analysis is anything but. Simply put, should the United States resume underground nuclear testing and, if not, should the United States be better prepared to do so in the future?

The decision to resume or not to resume underground nuclear testing is one of high stakes with multiple, complex, and uncertain interacting considerations.

Before we begin considering these questions, we clarify a few issues of scope. First, we are only addressing the question of resumption of underground nuclear testing. This is the result of several considerations. First, atmospheric testing is not on anyone's agenda in the United States or other established nuclear states. Even North Korea is conducting only underground tests. Second, most of the benefits of renewed nuclear testing can be achieved with underground tests. The main limitations of underground tests relate to better understanding of certain nuclear weapon effects, especially combined effects on satellites and electromagnetic pulse effects on the infrastructures that sustain modern societies. Finally, compared with atmospheric testing, limiting underground testing to yields no greater than 150 kilotons would reduce adverse domestic and international reaction. It would adhere to all current treaties, minimize damage to the environment, and allow for full-scale weapon design testing for weapons below that threshold.

Addressing this question of nuclear testing is not an esoteric intellectual exercise. With the aging of the US nuclear arsenal, the reemphasis on the importance of nuclear weapons—including new designs—in maintaining national security, the possibility of both Russian and Chinese testing inconsistent with the CTBT "zero-yield"[5] standard as interpreted and adhered to by the United States, and the recent dismantlement of important pillars of the arms control edifice constructed largely during the Cold War, one of the next dominoes to fall may well be a resumption of underground nuclear weapon testing.[6]

In the United States these factors have motivated a debate that has risen to the highest levels of government. In May 2020, the Trump administration reportedly held a high-level meeting of defense officials to discuss the possibility of conducting a nuclear test. According to a *Washington Post* article, "The meeting did not conclude with any agreement to conduct a test, but a senior administration official said the proposal is 'very much an ongoing conversation.' Another person familiar with the meeting, however, said a decision was ultimately made to take other measures in response to threats posed by Russia and China and avoid a resumption of testing."[7] We do not know the level of nuclear expertise of participants in this and potentially future discussions. Unfortunately, however, many participants in the broader public debate do not appear to be as informed as they could be. Moreover, few published discussions are both comprehensive and balanced. Many vociferously argue only one side of the issue. As a result, the nuclear weapon testing debate has not been particularly productive.

Resumption of nuclear testing could have momentous impacts on the prospects for a renewed qualitative and quantitative arms race, deterrence and the likelihood of nuclear war, nuclear weapon safety, international relations, nuclear proliferation, the future of arms control, and US international standing and moral leadership. Thus, the decision to resume or not to resume underground nuclear testing is one of high stakes with multiple,

---

[5] Arms Control Association, "Senate Rejects Comprehensive Test Ban Treaty."

[6] AVC, *Adherence and Compliance.*

[7] Hudson and Sonne, "Trump Administration Discussed."

complex, and uncertain interacting considerations. We hope this report will make a positive contribution to making the wisest decision.

Nuclear Testing Overview

Before the United States conducted the world's first nuclear test, code-named Trinity—a fifteen-kiloton plutonium implosion device detonated near Alamogordo, New Mexico, at a height of one hundred feet in July 1945—a concern surfaced that the extreme heat of the explosion, never previously experienced, might be sufficient to ignite the atmospheric nitrogen, which could cause a nearly instantaneous destruction of all life on the Earth's surface.[8] With seemingly—in retrospect—hubristic self-confidence, the Los Alamos National Laboratory (LANL) scientists decided that the risk of precipitating instant global destruction was not high enough to derail their test plans. That catastrophic possibility avoided, confidence was reinforced with the employment of the Little Boy atomic bomb against Japan in history's first wartime use of a nuclear weapon. Little Boy, employing a different fissile core material of uranium and a gun assembly design radically different from the Trinity experiment, had never been tested. But understanding of the underlying physics was validated by the devastating bombing of Hiroshima.

Since that first nuclear explosion in 1945, the United States has conducted, according to officially released accounts, 1,054 nuclear tests.[9] Eight other nuclear states have also conducted tests, as shown in Figure 1, for a worldwide grand total of slightly over two thousand tests. US tests were designed for a variety of purposes and in a variety of environments: in the atmosphere, in outer space, on the surface, underground, underwater, and deep in salt caverns.[10] Primary objectives[11] have included weapon design, stockpile sustainment, weapon safety and reliability, nuclear weapon effects studies, treaty and/or proliferation monitoring, civil engineering exploration, and some other special-purpose efforts. The great majority of tests had a single purpose, but the United States also carried out some multipurpose tests addressing, for example, both weapon design and safety, weapon effects and weapon design, or other combinations.

US Nuclear Testing

After World War II, the Atomic Energy Act of 1946 placed nuclear warhead design as well as all atomic energy developments under civilian control with the establishment of the Atomic Energy Commission.[12] Concurrently, the Armed Forces Special Weapons Project was created[13] to maintain operational control of nuclear weapons and related aspects, such as stockpile maintenance, storage, and security, and to determine the effectiveness of nuclear weapons in an operational context.

---

[8] Konopinski, Marvin, and Teller, *Ignition of the Atmosphere*. Although this report was published after the end of World War II, it reflects analysis conducted prior to Trinity. See also Toton and Scouras, *Trinity and Ivy Mike*.

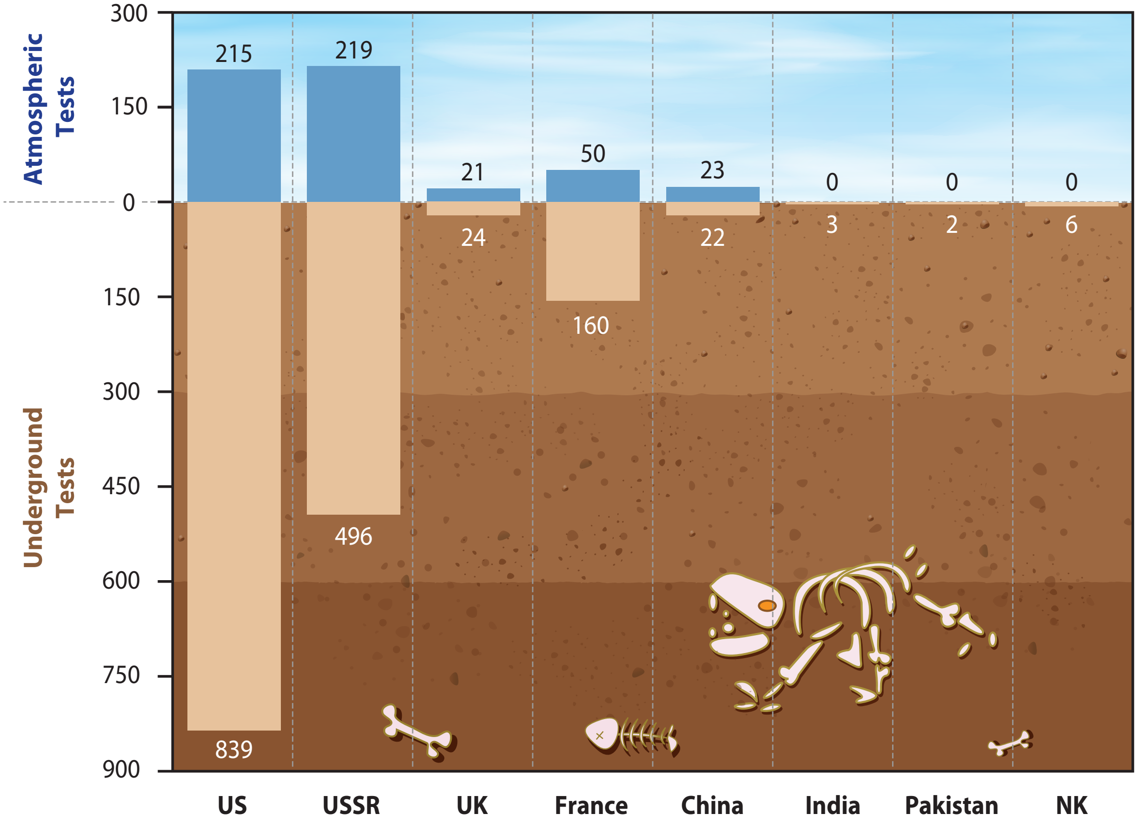
[9] NNSA Nevada Field Office, *United States Nuclear Tests*. The actual number of nuclear devices exploded exceeds 1,054, as a number of tests, each counted by the Department of Energy as a single unit, involved multiple simultaneous, or near-simultaneous, separate explosions. US total detonations total 1,149, including 28 joint detonations. One test employed 6 separate explosions. Some listings cite 1,030 as the number of US tests, subtracting from the total the 24 tests conducted jointly with the United Kingdom.

[10] Although most US underground nuclear tests have been conducted at the Nevada Test Site (since renamed the Nevada National Security Site), some have been conducted in Colorado, New Mexico, Mississippi, and Alaska.

[11] The NNSA Nevada Field Office report *United States Nuclear Tests* taxonomizes all US tests in one of six categories: weapons related, weapons effects, safety experiment, Plowshare, Vela Uniform, and joint US–UK. In the following section, we break these categories into related but not identical categories. In particular, the NNSA report does not maintain, or discuss, a separate category for stockpile maintenance activity.

[12] Atomic Energy Act of 1946.

[13] See the memo from Mary F. Shelley, with the attached document "Establishing and Early Development of the AFSWP."



	US	USSR	UK	France	China	India	Pakistan	NK
Total Tests	1,054	715	45	210	45	3	2	6
Test Years	1945–1992	1949–1990	1952–1991	1960–1996	1964–1996	1974, 1998	1998	2006–2017

Notes: (1) The US total does not include the atomic bomb attacks on Hiroshima and Nagasaki. (2) The US and UK totals each include twenty-four joint underground tests. (3) The India and Pakistan tests involved multiple explosions, so the total number of explosions for each country is six. (4) Israel and/or South Africa may have tested, but official confirmation is lacking. There are discrepancies with numbers listed here and with some other published lists. Some of this confusion is due to different methods of counting test events with multiple near-simultaneous explosions.

Figure 1. Worldwide Nuclear Testing

The US nuclear test program started in earnest in July 1946, almost exactly one year after the Trinity event, with a pair of tests, dubbed Operation Crossroads, conducted at Bikini Atoll in the Marshall Islands. Two stockpile Fat Man devices were used, one detonated above and the other below the water's surface, to investigate blast and shock effects on ships.[14] These tests were soon followed by numerous other atmospheric detonations at Bikini and Enewetak Atolls, including the highest-yield pure-fission test (Ivy King, 500 kilotons) and the first test with an experimental thermonuclear device (Ivy Mike, 10.4 megatons), both in 1952.

---

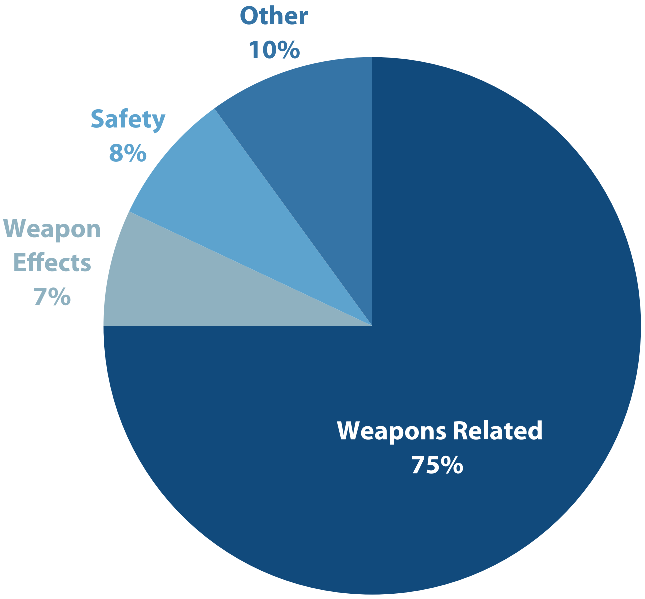
[14] The underwater Baker detonation created a so-called "base surge," a highly radioactive mist resulting from the collapse of the over-dense stem of the mushroom cloud that contaminated a flotilla of some fifty-seven ships well beyond what was anticipated, in what has been referred to as the world's first nuclear disaster.

In 1954, the first operational thermonuclear weapon was demonstrated at Bikini Atoll in an event code-named Castle Bravo. It yielded fifteen megatons, considerably more than what was expected, and contaminated an area of thousands of square miles with radioactive debris, exposing Marshall Islanders situated in neighboring atolls. Operation Hardtack I, a series of thirty-five nuclear detonations conducted in 1958, marked the end of nuclear surface detonations in the Pacific and the beginning of a nuclear moratorium that lasted until 1961 with the resumption of testing by the Soviet Union.

The United States responded in 1962 with Operation Dominic, a series of twenty-six nuclear tests using bombs dropped from aircraft and detonated in the atmosphere over Christmas Island and another five high-altitude tests launched by Thor missiles[15] from Johnston Island. The latter five, constituting Operation Fishbowl, were the first to demonstrate the hitherto unknown effect of electromagnetic pulse.

Testing at the Nevada Test Site started in 1951. A hundred atmospheric tests, including several cratering tests, were conducted during 1951 to 1962, with yields ranging from sub-kiloton to about seventy-five kilotons. With the imposition of the Limited Test Ban Treaty in 1963, all subsequent testing was conducted underground at the Nevada Test Site. While considerable care was taken to ensure the containment of radioactive debris, a few tests did have vents, most notably the Baneberry event of 1970 that released eighty thousand curies of iodine-131 into the atmosphere,[16] more than the combined releases from all subsequent underground tests. With public sentiment decidedly against the continuation of underground testing and political pressure building internationally, the Soviet Union announced a moratorium on nuclear testing in 1991 and the United States followed suit in 1993.

Nuclear tests are conducted for a variety of purposes, the most significant of which are indicated in Figure 2 and discussed below. Of course, any test can accommodate more than a single objective.[17]



The category "other" encompasses the Plowshare Program, joint US–UK testing, nuclear test monitoring support, and stockpile confidence testing.

Figure 2. US Nuclear Tests 1945–1992

Weapon development. After the Atomic Energy Commission was disestablished in 1974, responsibility for nuclear weapon development passed to another civilian agency, the newly created Energy Research and Development Administration. In 1977, President Jimmy Carter signed legislation

---

[15] Rademacher, *Plutonium Exposures*. One of the Thor missiles blew up on the launchpad, destroying the nuclear payload and significantly contaminating the island with plutonium.

[16] CTBTO, "Baneberry Incident."

[17] In the Department of Energy taxonomy in the NNSA Nevada Field Office report *United States Nuclear Tests*, all the listed US nuclear tests fall into one of six categories: weapons related, weapons effects, safety, Plowshare, Vela Uniform, and joint US–UK. We have classified the test types here in a closely related but somewhat different organization. In particular, we have broken out stockpile maintenance from weapons related as a separate category, and we do not retain Vela Uniform as a separate category.

creating the Department of Energy, which assumed responsibility for nuclear warhead development, testing, and sustainment. Since 2000, that mission has been carried out by the National Nuclear Security Agency (NNSA), a semiautonomous agency within the Department of Energy.

The early nuclear test program featured atmospheric tests, many of which combined weapon tests with effects tests. The focus of warhead development testing was on fission-fusion physics[18] as realized in various device configurations, with a push to make more efficient use of the fissile materials (i.e., to extract more bang per pound). A main driver of these design efforts was the push to miniaturize weapon packages, so they might fit on a MIRVed reentry vehicle. Typically, a large number of tests (of order ten) were required, which might have included some fizzles, before a new design was operationalized, while a number of other designs may have been tested but never entered the stockpile. Other development test efforts focused on tailored output devices, such as enhanced neutron and other special warhead types.

Presently the treaty monitoring community claims a "well-honed ability to monitor militarily significant nuclear test explosions anywhere in the world."

After the cessation of atmospheric tests in 1962, the focus remained the same but the method changed. Weapon development testing was generally carried out underground in a vertical line-of-sight configuration, where the nuclear devices were situated at the bottom of deep vertical shafts emplaced in canisters with test instrumentation measuring device performance. Within minutes, the underground cavity formed by the detonation collapses as hot explosive gas pressure leaks away, leaving a signature subsidence crater to mark the test site location, as shown in Figure 3.

Such craters are formed by the collapse of the cavity created by nuclear tests in deep vertical shafts.

Figure 3. Subsidence Craters at the Nevada Test Site with the Sedan Ejecta Crater in the Foreground

Weapon effects. Gaining an understanding of nuclear weapon effects through a program of nuclear explosions and extensive development of aboveground simulator facilities has been a decades-long national enterprise requiring an enormous investment of resources. Despite this, significant uncertainties remain and are unlikely to be resolved anytime in the foreseeable future.[19]

The early days of atmospheric testing largely focused on understanding air blast and thermal effects on structures, vehicles, ships, and various military equipment, as well as the simulated humans in these structures. A primary emphasis of US underground effects testing has been understanding how x-rays affect the survivability

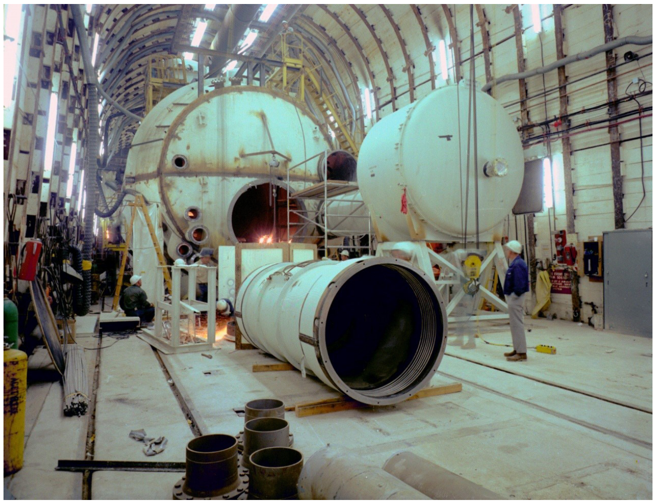
---

[18] A number of the development tests to produce a very low fission-fusion fraction device were stimulated by Project Plowshare, which sought to minimize potential radioactive contamination of its proposed civil engineering projects.

[19] Frankel, Scouras, and Ullrich, *Uncertain Consequences*. This report focuses on uncertainties in the present state of knowledge on the physical effects of nuclear weapons. However, the full consequences encompass economic, medical, social, and other considerations as well, so any assessment of actual weapon effects based just on testing is a lower bound.

of missile systems, reentry vehicles, and satellites.[20] Some tests gave soldiers operational experience in simulated nuclear battlefield conditions.

Unlike the majority of Department of Energy weapon tests conducted in a vertical configuration drilled from the surface, Department of Defense underground tests were more usually conducted in a horizontal line-of-sight tunnel configuration or, in at least four instances, in an underground cavity configuration. Department of Defense tests, involving a much more complex tunneling operation into the side of a mountain, tended to be more complicated and heavily instrumented than the Department of Energy vertical line-of-sight weapon tests. Figure 4 shows a typical target chamber under preparation.



Test articles are situated in an evacuated target chamber at the end of a long conical pipe conveying x-rays from the nuclear detonation. The pipe is rapidly pinched off at the source to prevent any hydrodynamic flow of device debris.

Figure 4. A Typical Target Chamber for a Horizontal Line-of-Sight Test

After the move underground, aboveground simulators (e.g., explosive test beds for blast waves and large radiation simulators) were developed in parallel to simulate (some with reasonable fidelity, others with much less) aspects of the nuclear environment. But nothing could simulate all aspects of a nuclear detonation along with associated target response complexities.[21]

Proliferation/treaty monitoring. Many have raised the technical possibility of cheating (i.e., conducting clandestine tests with detectable signatures reduced below those revealed by long-range detection capabilities). Presently the treaty monitoring community claims a "well-honed ability to monitor militarily significant[22] nuclear test explosions anywhere in the world, above ground or below, and to distinguish them from mine collapses, earthquakes, and other natural or nonnuclear phenomena."[23] The National Academy of Sciences concludes that "the threshold levels for IMS [International Monitoring System] seismic detection are now well below 1 kt worldwide for fully coupled explosions."[24]

At the Nevada Test Site, several tests were carried out to validate the CORRTEX system,[25] developed to determine yields of explosions as part of the Threshold Treaty cooperative on-site monitoring provision. The satellite-based component of the US Atomic Energy Detection System (USAEDS),[26] one of the national technical means used to monitor

---

[20] There have also been important non-space system efforts, such as the HYBLA GOLD experiment, which investigated the survivability of underground missile basing schemes.

[21] It is important to realize that an underground test is also a simulation of a real aboveground nuclear explosion, and important phenomenology (e.g., combined effects) will not be captured.

[22] We address the issue of military significance later in this report.

[23] Richards and Kim, "Advances in Monitoring."

[24] NRC, *Technical Issues*, 1.

[25] CORRTEX was the US yield measurement system fielded in 1988 as part of the Joint Verification Experiment whereby the United States and the Soviet Union cooperated to jointly field detonations at the Nevada Test Site and Semipalatinsk. It worked by measuring the time of reflection of electrical impulses along a cable length whose end point was being successively ionized by the propagating shock front.

[26] USAEDS is managed by the Air Force Technical Applications Center, an Air Force surveillance organization assigned to the 16th Air Force.

potential atmospheric test activity, was calibrated with nuclear test data as well.[27] Radioactive releases from underground test explosions were routinely monitored, and today, ground stations' detection of gases and particulates remains one of four measurement techniques[28] used by the Provisional Comprehensive Test Ban Treaty Organization's IMS, while USAEDS also monitors such leaks from high-flying aircraft.

The United States has also conducted a number of tests to assess the feasibility of cheating by masking the yield. Two such tests—Salmon and Sterling, part of the Vela Uniform series—were carried out in a deep underground salt dome in Mississippi to investigate the degree of decoupling of the seismic signal achievable by detonations in an underground cavity. As illustrated in Figure 5, the Sterling data indicate a potential decoupling factor of about seventy at sampling frequencies up to approximately one hertz, the frequency domain relevant for long-distance detection.[29] This level of decoupling makes it feasible for an experienced nuclear state such as Russia or China to contemplate conducting, albeit with some practical difficulty,[30] clandestine testing up to two kilotons.[31]

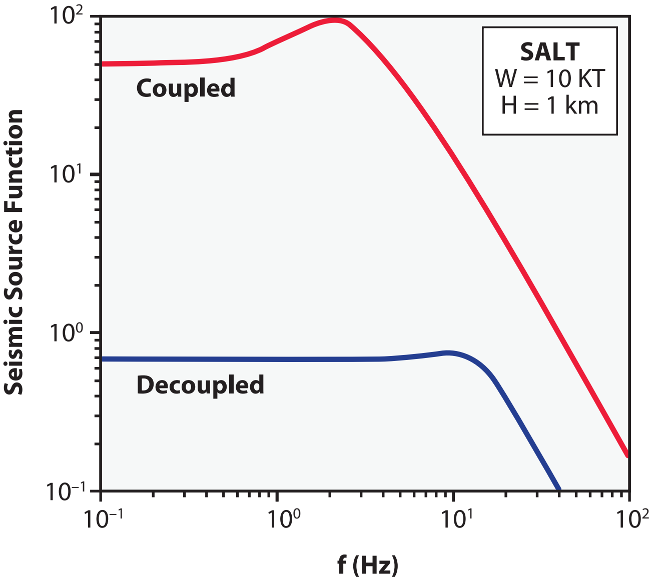


Figure 5. Seismic Cavity Decoupling Strength as a Function of Frequency

Peaceful nuclear explosions. In 1957 the Atomic Energy Commission conducted its first contained underground test, a 1.7-kiloton device detonated in an underground tunnel 274 meters below Rainier Mesa at the Nevada Test Site. Demonstrating the feasibility of nuclear excavation, it served as the harbinger for the US Plowshare Program, formally begun the next year, that would eventually number twenty-six distinct tests spanning 1961 to 1973 and would explore various civil engineering applications, such as widening of the Panama Canal,[32] stimulation of oil and gas deposits to facilitate extraction, and steam production for generation of electricity.[33] Figure 6 is a photograph of the Sedan

---

[27] In 1979, a Vela satellite detected a double-hump optical signature characteristic of a nuclear blast in the atmosphere over an unpopulated region of the South Atlantic/Indian Oceans, presumably executed by an unnamed nation-state. This interpretation of the Vela signal has not been officially confirmed.

[28] Seismic, hydroacoustic, infrasound, and radioactive leak monitoring.

[29] Figure 5 indicates a decoupling factor of about seventy at lower frequencies with drop-off at higher frequency measurements. While the decoupling factor drops off at higher frequencies, the seismic signal amplitude also drops off and a denser array of more close-in detectors would be necessary for high-confidence detection. Murphy and Barker, *Seismic Characteristics*.

[30] The complexity includes the need for a large and deep underground excavation and to hide the excavation activity.

[31] NRC, *Technical Issues*, 95.

[32] Back then, it was not viewed as important to ask what the Panamanians might think of such a project.

[33] A wide variety of earth-moving projects and oil and gas production stimulation efforts were proposed. As an example of the type of thinking entertained, in 1963 the Atomic Energy Commission proposed Project Carryall, whereby the California Department of Transportation would have employed twenty-two nuclear blasts to cut a highway pass for the construction of I-40 through a mountain pass in the Mojave Desert. Fry, Stane, and Crutchfield, "Preliminary Design Studies." Even more ambitious but less well-developed suggestions were bruited about but were not developed to the point of testing a nuclear

crater, the result of one of the Plowshare excavation tests. Funding for Plowshare formally ended in 1977, by which time it was quite clear that the downside of such nuclear exploitation—spreading radioactive fallout and contaminating watersheds—had completely dwarfed any realistic benefit and thoroughly aroused public protest. Natural gas stimulation was the only Plowshare Program initiative that ever garnered sufficient interest for industry to participate, but this too was short-lived.

Figure 6. Sedan Crater at Nevada Test Site Excavated by a Plowshare Test

The parallel Soviet effort, Nuclear Explosions for the National Economy, surpassed the US Plowshare Program both in terms of the number of tests and the diversity of applications.[34] The Soviet peaceful nuclear explosions program started in earnest in the mid-1960s and would ultimately number 122 events involving 129 explosions with yields ranging from 0.01 to 140 kilotons.[35] The Soviet civil engineering projects generally fell short of expectations and faced mounting international criticism because they released radioactivity into the atmosphere.[36] The last Soviet nuclear excavation test event was conducted in 1974, yet Soviet arms control negotiators adamantly insisted on exempting large-cumulative-yield explosions for future excavation projects throughout the negotiations of the Peaceful Nuclear Explosions Treaty, which was signed in 1976. One of the more interesting applications of Soviet peaceful nuclear explosions was the use of underground nuclear explosions to extinguish runaway gas well fires. During 1966 to 1981 the Soviet peaceful nuclear explosions program executed a number of attempts to stem five runaway gas wells using fully contained nuclear explosions—most of them successful. The Soviet program ended in 1984.

> **Nuclear weapons and their component materials are now aging long beyond their initial planned service lifetimes.**

As the "explosive" component of the otherwise civilian nuclear technology pursuits under the Eisenhower administration's "Atoms for Peace" movement, the peaceful nuclear explosions program always raised suspicions that it was a cover for weapon development. Indeed, both the United States and the Soviet Union conducted numerous device development tests under the peaceful nuclear explosions rubric, emphasizing

---

device, including the use of nuclear devices to dig a sea-level canal across Nicaragua, defeat killer asteroids, or power space ships as part of Project Orion.

[34] Nordyke, *Soviet Program*.

[35] The Soviet focus on developing inherently low-contamination devices during the peaceful nuclear explosions era gives Russia today a distinct advantage in the possible scientific pursuit of a pure fusion device. A more limited number of the US test events similarly involved validating "clean" device designs suitable for use in peaceful Plowshare endeavors. OSTI, *Executive Summary: Plowshare Program*.

[36] The Soviets interpreted the Limited Test Ban Treaty provision for *no radioactive release beyond state borders* to apply only to particulate matter and not gaseous releases. Accordingly, they argued that a subsidence crater produced by the collapse of an underground cavity formed by a nuclear explosion was not a treaty violation if only radioactive gases escaped into the atmosphere.

high fusion-fraction yield with minimal fission-related residual contamination. While none of these devices may have been weaponized, their fungibility was not lost on critics of the program.

Safety. In 1968, the United States established a requirement that all warheads be one-point safe. That means should accident or misadventure (e.g., an accidental drop of an unarmed bomb from an airplane)[37] initiate an explosion at a single point on the warhead's high-explosive driver, it would not lead to a sequence of events resulting in nuclear energy release greater than four pounds of TNT equivalent.[38] Additional requirements focus on isolating the warhead's electrical initiation system from accidentally firing a detonation signal and preventing dispersal of radioactive pit material in the event of accidental fires even without nuclear yield. A history of mishaps points to the need to upgrade the safety systems of nuclear weapons. The change to more insensitive high-explosive driver was a major design change for a number of systems[39] and could not be integrated into the arsenal absent a series of nuclear tests. Other positive design changes to enhance safety over the years include mechanical safing and enhanced electrical isolation, which reduces the calculated risk of inadvertent electrical firing of detonators in an accident scenario to less than one in a million.

Along with concerns about accidents or even terrorist scenarios, in the post-test era there is a new concern that nuclear weapons and their component materials are now aging long beyond their initial planned service lifetimes. Aging high explosives, the nuclear pit, and other material components subject to multiyear radiation exposure from weapon fissile components present a clear threat to safety. Department of Energy nuclear laboratories are addressing this threat through Life Extension Programs, which include, *inter alia,* remanufacture of some elements, enhanced surveillance, and nonnuclear testing of components.

A persistent theme throughout the history of nuclear weapon testing has been the element of surprise.

Stockpile confidence. Stockpile confidence tests, sometimes referred to as production verification tests, were intended to build confidence in the long-term stability and readiness of the strategic stockpile. In a 1987 draft response[40] to a congressional inquiry, a Lawrence Livermore National Laboratory (LLNL) report stated:

> A current nuclear explosive test necessary to consider a weapon adequately tested is the detonation of a war reserve production unit or preferably a unit withdrawn from stockpile . . . To the extent that it is feasible, it is desirable that it have been subjected to a simulated stockpile-to-target sequence of the enabling actions and most severe operating conditions it will encounter before detonation in actual use.

The report goes on to explain that it became routine practice to conduct such nuclear tests on the first production unit of a weapon in actual stockpile configuration, usually within the first year of deployment. The number of stockpile confidence tests was redacted from the unclassified version of the 1987 report.

The report disclosed that problems were encountered with fourteen (later expanded to fifteen) nuclear weapon designs that had been part of the

---

[37] All of which have happened, in various forms and configurations, multiple times. See, e.g., Perrow, *Normal Accidents*, on the Goldsboro incident, or the NPR report "Nuclear Bomb Lost" on the Tybee, Georgia, incident. There have been a variety of other incidents, including at least one where a crashing airplane impacted stored nuclear weapons on the ground.

[38] Kidder, *Report to Congress*.

[39] Employed on all warheads entering stockpile before 1978.

[40] Kidder, *Maintaining the U.S. Stockpile*.

inventory since 1958; post-deployment nuclear tests were required to correct the problems. The report distinguished between what it termed the "Sixties Nine" and the "Eighties Five." The former were the product of a crash effort to complete tests before the looming onset of the 1958 moratorium[41] and then a hasty rush to build and stockpile weapons during the three years of the moratorium.[42] In the report's estimate, the latter group of five's post-deployment testing suffered from a lack of sufficiently rigorous stockpile confidence tests. After further post-deployment testing, problems were assessed as corrected and weapon performance as satisfactory in thirteen cases, with only one weapon type requiring significant corrective action. Since that assessment, the United States has reduced the number of basic weapon designs in the present force structure to seven,[43] and more than thirty years have passed. So direct projections from that experience to the current state of the stockpile do not seem relevant.

A History of Nuclear Test Surprises

A persistent theme throughout the history of nuclear weapon testing has been the element of surprise. In retrospect, surprises surfaced in the gamut of nuclear test types with some regularity. It is convenient to categorize such unexpected nuclear test results in a threefold taxonomy: surprises in (1) weapon development and safety, (2) nuclear weapon environments, and (3) vulnerabilities of military systems to nuclear environments.[44]

Surprises in weapon development and safety. Weapon design testing, which comprises more than 80 percent of all US tests, has been a source of surprise ranging from fizzles (detonations with no or little nuclear yield) to other variations from expected yields. In some instances, these surprises revealed important safety issues.

During the test moratorium of 1958–1961, hydronuclear criticality experiments uncovered and resolved significant one-point safety concerns for some weapons with boosted fission primaries already in the stockpile and others on the verge of entering. Nonetheless, production of weapons was halted, and only after the resumption of testing in late 1961 did a new series of nuclear tests allow for a confident retrofitting of the US stockpile.[45] A 1987 report to Congress[46] identified fifteen US nuclear weapon systems with yield and safety deficiencies discovered as unwelcome surprises after deployment. As late as 1997, during the present voluntary testing moratorium and well after the era when further nuclear testing might be used to validate any corrective actions, congressional testimony identified the nation's most sophisticated strategic weapon, the W88 nuclear warhead, as eliciting safety concerns.[47] During the testing era, other issues related to aging of the high explosives and low-temperature performance were also unwelcome surprises.

---

[41] During the month preceding the moratorium, the United States conducted thirty nuclear tests, ten of which were conducted in the last five days. Kidder, *Maintaining the U.S. Stockpile*, 16.

[42] For example, in the case of the since retired W52, the high-explosive driver—incredibly, in retrospect—was swapped out for a new explosive and the weapon was then stockpiled without testing. The chemical high-explosive characteristics can affect yield, range, and other big-ticket considerations. Kidder, *Maintaining the U.S. Stockpile*, 17.

[43] B61, W76, W78, W80, B83, W87, and W88 (and variants). OASD(NM), *Nuclear Matters Handbook 2020*.

[44] The last two categories are subsumed in the category of weapons effects in Figure 2.

[45] Thorn and Westervelt, *Hydronuclear Experiments*.

[46] Miller, Brown, and Alonso, *Report to Congress*, 24; and *Safety and Reliability: Hearing 105–267*, 134.

[47] Schlesinger, "Clinton Defers." This *Wall Street Journal* article was appended to and published as part of the Congressional Record of *Safety and Reliability: Hearing 105–267*.

The utility of hydronuclear[48] testing to assess problems that might arise in modern thermonuclear designs is quite limited,[49] and a 1994 JASON study[50] concluded that hydronuclear testing (or supercritical testing in general) was not required to maintain the existing US stockpile, as long as a robust science-based stockpile stewardship program was established and maintained.

In 1957, Castle Bravo (shown in Figure 7), a test of the first US air-deliverable thermonuclear bomb design, was detonated at Bikini Atoll in the Marshall Islands. It delivered the largest measured yield in US testing history at fifteen megatons, significantly exceeding the expected yield. It also produced unexpectedly widespread fallout contamination over thousands of square miles of ocean; islands to the east of the blast required evacuation of native populations and personnel on ships at sea were exposed to radiation doses. It caused numerous incidents of radiation sickness, including one fishing boat fatality. The excess yield was triggered by the unanticipated production of additional energy from the bomb's U-238 component in a fission-fusion-fission reaction.[51] The fears of atmospheric fallout engendered by Castle Bravo and subsequent explosions gave the major impetus for the eventual signing of the Limited Test Ban Treaty in 1963.

Surprises in vulnerabilities of military systems. One key goal of nuclear weapon testing was to understand how military systems responded when exposed to actual and simulated nuclear test environments. Open discussion of all these instances is constrained by security and classification restrictions, but a number of specifics can be described.

This image was taken 3.5 seconds after detonation at a distance of 75 nautical miles east of ground zero from an altitude of 12,500 feet.

Figure 7. Castle Bravo Event

Our strategic nuclear systems' survivability in a nuclear environment, as well as their nonnuclear command and control, can be tested, to a limited degree, in aboveground nuclear environment simulators.[52] But shortfalls of fluence, spectrum, timing, exposure, and volumes, as well as limited ability to exercise complex systems' synergistic response modes to a combined nuclear environment, limit the confidence we can place in simulator test data. The technically more satisfying alternative that remedied many such shortfalls, although vastly more expensive, was to test such critical systems

---

[48] Traditionally, hydronuclear tests involve full-scale nuclear weapon assemblies complete with high-explosive driver but with insufficient fissile material to sustain a chain reaction. So named because the high-explosive implosion creates pressure levels causing materials to behave as a fluid, such tests were primarily used to investigate weapon safety in accident scenarios resulting in the detonation of the high-explosive driver. Some also lump current subcritical testing activities at Tunnel U1a of the Nevada Test Site under the rubric of hydronuclear. These experiments investigate material properties of plutonium under extreme pressure and temperature conditions.

[49] Conditions for thermonuclear boosting are established only after a considerable fission yield has already been achieved.

[50] Drell et al., *Science Based Stockpile Stewardship*.

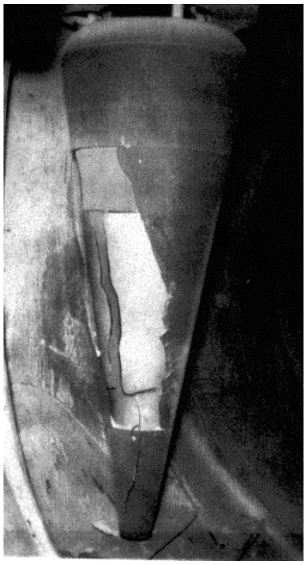
[51] Kennedy, *Fallout Forecasting*.

[52] These included x-ray simulators such as Double Eagle, Blackjack, Pithon, Casino, and Decade; gamma-ray simulators such as Aurora, HERMES, and PulseRad 1150; prompt neutron simulators such SPRIII and FBR; electromagnetic pulse simulators such as ARES, Trestle, EMPRESS, and Pax River HPD; and air blast simulators such as LBTS and LIHE. A number of these simulators have been either mothballed or permanently retired.

underground. However, not all shortfalls stemming from lack of atmospheric testing, including high-altitude electromagnetic pulse and other high-altitude effects, can be well simulated in an underground test.[53] But for a period of twenty-five years, until the cessation of the nuclear test era in 1992, the Department of Defense continued to test all major strategic systems, primarily to x-ray effects, in a series of underground nuclear tests.

The results of such tests highlight the unique value of the nuclear test program. A senior official in the nuclear effects test community once remarked[54] that every major military system introduced for the first time to the underground test environment had produced surprises, up to and including system failure. As a stark example, Figure 8 is a photograph of the post-test Mk 12 reentry vehicle, where catastrophic damage mechanisms had been overlooked in pretest analysis. Other less dramatic examples of surprises occurred with testing of the Mk 2 Mod2, Mk 3, Mk 4, Mk 12A, and Mk 21 reentry systems; the Spartan anti-ballistic missile; and the Trident I and II guidance systems.[55]

Another major vulnerability, which came to light more than thirty-five years into the testing program, was the discovery of potential cold (low-energy) x-ray vulnerabilities in the optical and power conversion systems for critical space-based surveillance systems.[56] These vulnerabilities were confirmed and mitigation strategies explored in some of the last underground tests at the end of the nuclear testing era.



This image shows unexpected heat shield failure after x-ray exposure in an underground nuclear test.

Figure 8. Mk 12 Reentry Vehicle Damage

Surprises in the nuclear weapon environments. Acquiring understanding of nuclear weapon effects has been a rich, and usually disconcerting, source of surprise throughout the testing era. Among the surprises are effects that simply had not previously occurred to Department of Defense scientists, including some that first became evident through observations of naturally occurring phenomena.

Early in the testing period—during what some have termed the "cook and look" era—scientists observing nuclear tests were regularly surprised

---

[53] Underground tests were never a perfect simulation of the radiation threat environment. Mining costs constrained the nuclear devices to have the smallest practical yield, often orders of magnitude less than the threat yield. This resulted in x-ray spectra that were not ideal and pulses of shorter duration than desired. Also, the neutron pulse was not properly sequenced, arriving too soon behind the x-ray pulse. Scatter stations and filters were often used to better tune the temporal width and the relative magnitudes of the neutron, gamma, and x-ray pulses. Although there were practical limits to how much the x-ray spectrum and pulse width could be modified, the x-ray fluence was essentially unlimited and well beyond what is achievable today with aboveground simulators.

[54] Discussion during the 1980s among two of this report's authors and senior leaders of the Defense Nuclear Agency, the US government organization charged with conduct of the nation's nuclear weapon effects program.

[55] Miller, Brown, and Alonso, *Report to Congress*, 24.

[56] Conrad et al., *Collateral Damage*. The vulnerability was first identified in a calculation by Gerry Gurtman in the early 1980s and subsequently substantiated in simulator facilities emitting line x-rays. Experiments with real, continuous nuclear spectra were conducted in the Mineral Quarry and Hunters Trophy underground tests.

by unanticipated phenomena. A notable example from that era includes the discovery of what came to be termed "nonideal" air blast, which could result in significantly enhanced damage to targets sensitive to dynamic pressure. Other examples include unanticipated incidents of thermal "flash-over" phenomena in target structures, aspects of radioactive spread and deposition, and long-range atmospheric ionization and blackout.

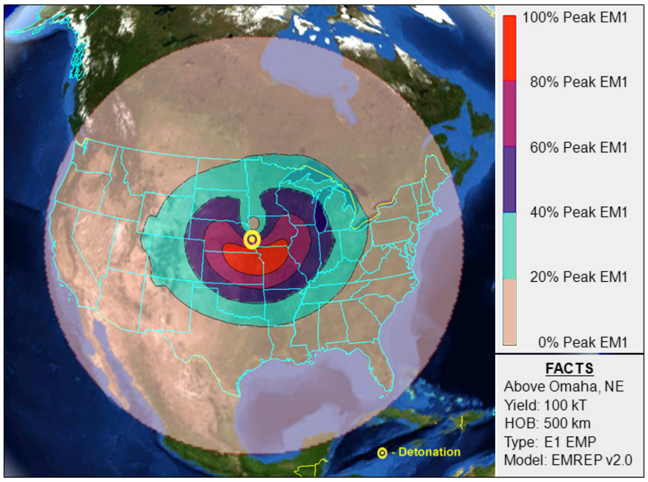


Figure 9. Calculated Electromagnetic Pulse Footprint from a Detonation over the United States

One of the most glaring surprises, the entirely unanticipated phenomenon of nuclear electromagnetic pulse, was unearthed by an atmospheric test in 1962, an exo-atmospheric detonation of the 1.4-megaton Starfish Prime four hundred kilometers above the Pacific Ocean. The inhabitants of Hawaii were among the first to discover the existence of electromagnetic pulse as streetlights suddenly went off in Honolulu, eight hundred kilometers away from the ocean location under the blast, and an Air Force radar station on the island of Kauai failed. As indicated in Figure 9, a high-altitude nuclear burst is capable of exposing the entire contiguous forty-eight united states to electromagnetic effects.[57] The intensity of the exposure is a function of weapon yield, design, and other factors.

Starfish Prime was also the source of another disconcerting surprise that was not discovered until months later: over time it disabled all known Earth satellites in orbit[58] through the unanticipated mechanism of Van Allen belt pumping. Other phenomena that had been completely missed by the nuclear community, such as the potential for ozone depletion and nuclear winter, raise analytic issues arguably worthy of further study. Thus, despite an extensive investment in acquiring knowledge of nuclear environments over five decades of nuclear testing, "significant uncertainties in physical consequences remain because important phenomena were uncovered late in the nuclear test program, have been inadequately studied, are inherently difficult to model, or are the result of new weapon developments."[59]

The long-term medical consequences of exposure to the nuclear environment were essentially unknown during the first decades of the nuclear test era. Figure 10, showing the deployment of US Army troops near a nuclear blast to familiarize them with the nuclear battlefield environment, underscores that point. Understanding of the full consequences is still being pursued today, in both scientific laboratories and court venues[60]

---

[57] Electromagnetic pulse exposure is a line-of-sight phenomenon extending to the horizon, so actual coverage depends on detonation location and altitude. Figure 9 shows electromagnetic pulse exposure for a nuclear detonation at an altitude of five hundred kilometers above Omaha, Nebraska.

[58] This was 1962 so there still weren't many satellites in orbit nor a highly space-dependent civilian and military telecommunications infrastructure. Still, the test managed to kill Telstar, the AT&T satellite that first demonstrated feasibility of transmitting television signals by space relay. There were also classified satellites in orbit at the time whose fate remains classified.

[59] Frankel, Scouras, and Ullrich, *Uncertain Consequences*. See note 19.

[60] The Department of Defense, through its designated lead the Defense Threat Reduction Agency, maintains the Nuclear Test Personnel Review program, which supports the Department of Veterans Affairs and the Department of Justice during review of veterans' medical claims by maintaining an archive of veterans

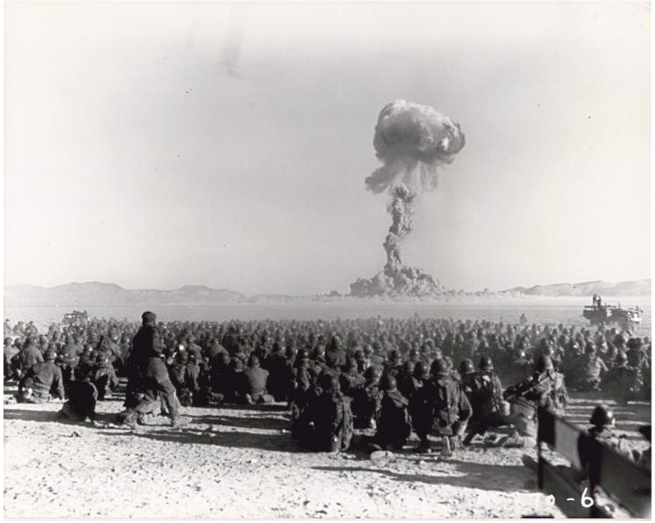


Figure 10. Exposure of US Troops during an Atmospheric Test in the 1950s

as studies have attributed more than ten thousand excess cancer deaths in the United States between 1951 and 1962 alone to radiation exposure from pre-1963 atmospheric tests.[61]

Science-Based Stockpile Stewardship and the Certification Process

Shortly after taking office in 1993, President Bill Clinton issued a Presidential Decision Directive[62] extending President George H. W. Bush's 1992 moratorium on nuclear testing until at least September 1994. It included a mandate to ensure by other means the safety, reliability, and performance of what was presumed to become a reduced but stagnant nuclear stockpile. Implicit in that mandate was the need to retain the essential core competencies, which in the past were continuously honed by the "design, build, test, and renew" approach to nuclear stockpile confidence.[63] The challenge was met by a team chaired by Victor Reis, then assistant secretary for defense programs in the Department of Energy, and including key members from each of the three national nuclear laboratories.[64] Together they crafted a strategy that became known as Science-Based Stockpile Stewardship (the Department of Energy name for the stewardship program), which, *inter alia*, featured enhanced surveillance of the stockpile, increased reliance on cutting-edge scientific computing, and new simulators to address key physics issues that, in the past, might have been addressed with underground tests. Figure 11 is an image of one such simulator that uses high-intensity lasers to compress nuclear materials in studies of nuclear ignition.

The Stockpile Stewardship Program was no doubt a masterful stroke.

Subsequently, the Stockpile Stewardship Program was established by the National Defense Authorization Act of 1994 with complementary implementing language[65] specified by another Presidential Decision Directive.[66] In 1995 President Clinton established a requirement that a panel of technical experts assess the safety, reliability, and performance of the active stockpile annually and that the secretaries of defense and energy, with the concurrence of the Nuclear Weapons Council, the directors of the nuclear laboratories, and the commander of US Strategic Command, certify the results to the

---

present at atmospheric nuclear tests and providing estimated or actual radiation dose information. DTRA, "Nuclear Test Personnel Review (NTPR)."

[61] Garbe et al., *Health Consequences*.

[62] White House, *Moratorium on Nuclear Testing*.

[63] The Department of Energy nuclear laboratories afforded their warhead designers multiple testing opportunities to empirically tweak key design features in what some mockingly referred to as "cook and look."

[64] For a detailed assessment of the Stockpile Stewardship Program at twenty years, see Reis, Hanrahan, and Levedahl, "Big Science of Stockpile Stewardship."

[65] National Defense Authorization Act for Fiscal Year 1994, Pub. L. No. 103-160, § 3137 (1994).

[66] White House, *U.S. Policy on Stockpile Stewardship*.

president.[67] Additional safeguards addressed the need to maintain the vitality of the nuclear weapons complex and to retain the ability to resume underground testing should the need arise and a presidential directive be issued to do so.

Figure 11. Target Chamber of the National Ignition Facility at LLNL with 192 Laser Beams Converging on a Target in the Containment Sphere

While the Stockpile Stewardship Program was no doubt a masterful stroke, at the time it seemed striking how quickly the nuclear laboratories acceded to it, given that they had previously vociferously argued for the need to continue testing. Resistance likely was softened by the program's generous payout.[68] Each of the three nuclear laboratories—LANL, LLNL, and Sandia National Laboratories (SNL)—were rewarded with a state-of-the-art supercomputer plus a cutting-edge laboratory facility. LANL was funded to build the Dual-Axis Radiographic Hydrodynamic Test (DARHT) facility for analyzing primary implosion physics; LLNL, the National Ignition Facility (NIF) for investigating radiation-hydrodynamic phenomena associated with nuclear detonations; and SNL, the Microsystems and Engineering Sciences Applications (MESA) complex for fabricating trusted microelectronics and micromechanical systems. Additionally, LANL and LLNL would maintain testing expertise by conducting subcritical plutonium aging experiments in deep underground chambers at the Nevada Test Site with similar processes and comparable rigor as exercised in past underground tests. It was argued that these facilities and activities would not only meet the needs of the stockpile but also maintain critical skill sets and attract new talent.

Originally submitted to Congress with a ten-year funding profile, the Stockpile Stewardship Program was billed as providing high confidence in the safety, reliability, and performance of the stockpile "indefinitely."[69] It has now been funded for over a quarter century with no officially stated need for a return to underground testing. Yet, there is a growing chorus of voices, including some former weapon designers, who question the adequacy of the Stockpile Stewardship Program as it continues with a patchwork of warhead component modifications and upgrades, validated solely by computational means and nonnuclear experimentation. A recent article[70] coauthored by John Hopkins, former head of nuclear testing at LANL, argues that the life-extended nuclear weapons in our arsenal today differ sufficiently from the original tested designs that we can no longer be fully confident in their reliability and performance. Aging effects, remanufactured components, and other departures from tested designs all contribute uncertainties that collectively could result in catastrophic failures that would not be unveiled by the Stockpile Stewardship Program. Proponents of the program argue to the contrary—that virtually all the lifetime extension modifications can be validated with aboveground tests and the unique design components that can

---

[67] Medalia, *"Safeguards" and Net Assessments*.

[68] The Stockpile Stewardship Program was estimated to cost about four to five billion dollars per year for ten years, about the same as the nuclear testing program it replaced. See, for example, *Safety and Reliability: Hearing 105–267* (testimony of Dr. Vic Reis).

[69] *Safety and Reliability: Hearing 105–267* (testimony of Dr. Vic Reis).

[70] Hopkins and Sharp, "Scientific Foundation Eroding."

only be tested in an underground test have not undergone significant change, if any.

Testing Moratoria and Treaties

The Castle Bravo detonation of March 1, 1954, which unexpectedly came in significantly higher than the expected yield and created major downwind contamination, galvanized debate on nuclear testing and spurred an international call for a comprehensive test ban. Public concern had already been high in light of research confirming an alarming buildup of strontium-90 in the teeth and bones of children who consumed the milk of cows that had fed on contaminated grass. Such fears were further inflamed in no small part by the publication of *On the Beach*,[71] a wildly popular best seller in 1957 describing the submariner survivors of nuclear Armageddon tragically anticipating their fate as radioactive poison continues its inexorable spread around the globe.

The Soviet Union proffered several test moratorium deals in 1957, but the Eisenhower administration eventually rejected them all because of limitations of US national technical means to adequately monitor potential Soviet cheating, coupled with Soviet unwillingness to accede to on-site inspections. On March 31, 1958, Soviet Premier Nikita Khrushchev chose to preemptively stake out the moral high ground and declare a unilateral halt to nuclear testing, provided the other nuclear powers would follow suit. The United States and the United Kingdom had little political choice but to join this test moratorium.

In 1961, to the apparent surprise and consternation of the United States, the Soviet Union broke out of the moratorium with an underground test. This was followed by the Tsar Bomba—at greater than fifty megatons, the largest ever nuclear test—whereupon President John F. Kennedy authorized resumption of US atmospheric tests as well. In contrast to the Soviet Union,[72] in the United States personnel experienced in nuclear testing had already been released to pursue other work, and it took heroic efforts to reassemble the essential talent. During this period, until 1963, when the Limited Test Ban Treaty entered into force, the United States executed a series of thirty-one atmospheric tests, code-named Operation Dominic, that included five detonations in space, revealing the previously unknown phenomenon of electromagnetic pulse in the very last such test in 1962. However, the hasty reconstitution of the US atmospheric test capability severely impacted the data return, which suffered from insufficient planning. In contrast to the US weapons community, which had been caught flat-footed, the Soviet Union continued planning throughout the moratorium, conducting its first atmospheric test the very next day after Khrushchev's announcement of the moratorium's end.

After the Cuban missile crisis in October 1962, both leaders sought to reduce tensions by reopening a dialogue on a nuclear test ban. This led to the successful negotiation of the Limited Test Ban Treaty, which entered into force in October 1963. It banned all nuclear explosions in the atmosphere, underwater, and in space, but allowed underground explosions, provided that any release of radioactive debris would not propagate beyond the borders of the country conducting the test.

In 1974, the United States and the Soviet Union reengaged in discussions on further nuclear testing restrictions. This time agreement on the Threshold Test Ban Treaty was reached within months. The major provisions of the treaty were to limit the yield of any test to 150 kilotons; to conduct tests

---

[71] Shute, *On the Beach*, published initially as a four-part series "The Last Days of the Earth" in the *Sunday Graphic* (London, 1957). The *New York Times* reviewer called it "the most haunting evocation we have of a world dying of radiation after an atomic war."

[72] Khrushchev announced the end of the moratorium on August 30, 1961, and testing resumed on September 1, 1961.

only at agreed-on locations to facilitate verification; and to exchange seismic and yield data at these sites for verification purposes. The treaty verification protocols were strengthened in 1990 with additional provisions for each party to have the option to conduct on-site hydrodynamic yield measurements on any test conducted by the other party in excess of 50 kilotons.

Subsequently, the Soviet Union, under the leadership of Mikhail Gorbachev, announced a moratorium on all Soviet testing in 1985. That moratorium lasted until 1987 when the Soviets resumed testing after complaining, accurately, that the United States ignored the Soviet pause and continued testing throughout that period. In 1992 the United States conducted its last underground nuclear test, and both the United States and Russia—which had conducted its last nuclear test in 1990 just before the collapse of the Soviet Union—began observing another voluntary test moratorium in anticipation of the signing of a comprehensive test ban treaty.

The CTBT was completed in September 1996 and has since been signed by 184 nations and ratified by 168. But it cannot enter into force until ratified by 8 additional countries, including the United States and China. It prohibits "any nuclear weapon test explosion or any other nuclear explosion." Article 1 of the CTBT specifies a "zero-yield" criterion; the United States interprets this as allowing only subcritical testing, whereas Russia appears to have a more liberal interpretation, allowing supercritical testing at yields that are undetectable by the CTBT IMS.[73] After the US government signed the treaty, the US Senate rejected the Clinton administration's efforts to ratify it, although the United States currently abides by its provisions.

Both China and France continued to test during the present moratorium, with the last tests by both countries in 1996.[74] India and Pakistan both conducted tests in 1998—all of them underground—after the start of the US–Russia moratorium. At present, as far as is known, the only nation that continues to either test or threaten to test is North Korea, whose last known test was in 2017.

The Arguments: To Test or Not to Test?

Arguments for and against the resumption of nuclear testing have both policy and technical dimensions. The technical arguments focus on the Stockpile Stewardship Program's adequacy—or inadequacy—to maintain the requisite scientific understanding and generate an acceptable level of confidence in the functioning of the US arsenal, as tested by its designers, and in its ability to detect "black swan" issues that might arise. They also address the adequacy of the United States' simulator facilities to replicate nuclear environments as might be needed, as well as its technical ability to detect cheating to preclude adversaries from garnering tactical or strategic advantage through clandestine testing. The policy arguments point to the claimed erosion in the effectiveness of the country's deterrent and the potentially grave implications for extended deterrence and geopolitical stability. This group of arguments also includes concerns regarding the consequences of resuming testing for nuclear proliferation and a renewed arms race.

On the next several pages, we summarize the cases for and against resumption of nuclear testing.

---

[73] AVC, *Adherence and Compliance*, executive summary.

[74] CTBTO, "France's Last Nuclear Test" and "First Chinese Nuclear Test."

Major Arguments in Favor of Resumption of Testing

Underground testing is needed to underwrite deterrence by enabling the development of specialized lower-yield nuclear weapons. The United States' present nuclear arsenal evolved with the Soviet Union as the country's principal foe. It emphasizes high-yield weapons effective against hardened intercontinental ballistic missile silos, other strategic and conventional military forces, command and control, and political and military leadership. High-yield weapons are also effective against soft economic facilities that present "area" targets or are close enough together that more than one can be adequately damaged with a single weapon. Collateral damage, whether to humans, structures, or the environment, originally was seen as a bonus, rather than something to be avoided as it is now more commonly viewed. But, because high-yield weapons create undesirable collateral damage, there is a concern that the United States might be self-deterred from retaliating in the face of attacks that do not directly target it and involve low-yield weapons from either established or nascent nuclear states. And, even if the United States is not self-deterred, its adversaries might judge that it will be, thereby undermining deterrence.

Because of this, some argue that the US arsenal needs to be bolstered by a more diverse set of low-yield, tailored-output weapons[75] whose availability enhances the credibility of the threat of retaliation by mitigating the prospect of self-deterrence. While the United States now has several lower-yield weapons in its arsenal, they are insufficient in quantity and diversity of delivery systems. In particular, sub-kiloton-class specialized weapons deliverable by survivable platforms have been proposed as a counter to Russia's nonstrategic nuclear arsenal governed by its "escalate-to-deescalate" doctrine. Such weapons are unlikely to be available absent testing.

On the other hand, it has also been argued that lower-yield weapons will inevitably lower the threshold for nuclear use.[76] Unfortunately, this argument usually fails to clearly distinguish between US nuclear use and adversary nuclear use and between nuclear first use and nuclear retaliation. Lower-yield weapons may indeed lower the threshold for nuclear first use for both the United States and adversaries in certain scenarios. It is reasonable that a state would more likely contemplate crossing the nuclear threshold in an armed conflict by using lower-yield nuclear weapons with the hope that nuclear war might be contained below the level of Armageddon. But our adversaries already possess lower-yield weapons. According to a declassified Central Intelligence Agency (CIA) memorandum,[77] "public statements by Russian scientists and officials since 1993 indicate that the last nuclear warhead designed during the Soviet era was a device tailored for enhanced output of high energy X-rays with a total yield of only 300 tons."

Thus, US matching of these adversary capabilities would not affect the threshold of adversary nuclear first use. Moreover, the likelihood that the United States would be tempted to use nuclear weapons first—whatever their yields—is not a primary concern. The scenarios under which this action might be plausibly contemplated are rare. To the extent it is a concern, the solution is to just not conduct a

---

[75] Tailored weapon design examples include enhanced electromagnetic pulse and enhanced neutron output.

[76] Kastetter, "Destabilizing Implications."

[77] CIA, "Evidence of Russian Development," 3.

nuclear first strike. Rather, the principal concern is that deterrence is undermined by failure to match adversary lower-yield capabilities with the ability to *respond* in kind. Lowering the threshold through enhanced nuclear strike flexibility is an argument explicitly rejected by the Pentagon, which argued the exact opposite in the Nuclear Posture Review.[78]

Beyond shoring up deterrence by closing gaps in the retaliatory *capability* of our arsenal, resumption of nuclear testing would, it can be argued, also buttress the *credibility* of deterrent threats to employ it in both first-strike and retaliatory scenarios. In contrast to the immediate aftermath of World War II when the United States held a monopoly in nuclear weapons and the earlier days of the Cold War when it enjoyed dominance in nuclear capabilities, we are now in an era where nuclear threats are increasingly less believable. Unconstrained nuclear war with either our principal adversaries, especially Russia but increasingly also China, would destroy the United States as a functioning entity. And any US nuclear use, even in retaliation, risks escalation to that end state. Even a single North Korean detonation on a US city could throw the United States into a paroxysm of rage and retribution with fatal consequences for constitutional rule. Thus, it is in the United States' interest to avoid crossing the nuclear threshold except under the most severe provocation. Yet deterrence, and especially extended deterrence, depends on a credible threat to violate this self-interest. Resumption of nuclear testing could reinforce the perception, of friend and foe alike, that the United States regards its nuclear strategy with seriousness and is willing to employ nuclear weapons when required.

Confidence in the stockpile is eroding as the state of weapons changes over time. With time, fissile materials in the weapon age and the other materials that compose the weapon may deteriorate as they continue to be bombarded by radiation from the weapon's fissile materials. Radioactive decay of plutonium produces energetic uranium atoms and alpha particles, which in turn create crystal lattice defects in the plutonium pit that may not self-heal. Also, alpha particles capture electrons, creating helium atoms that can aggregate to produce voids. Both effects can potentially lead to changes in key material properties that affect the performance of the pit. The nonfissile materials in the weapon may similarly undergo material property changes over time. The accumulation of small degradations in multiple components of a complex weapon leads to uncertainty in overall system performance. While surveillance of aging effects can be used to examine potential degradation of fissile materials and replacement of other material components may mitigate this risk, they cannot eliminate it.

When originally manufactured, weapons were not expected to last indefinitely. Their expected service lifetimes are being extended by a decade or more through Lifetime Extension Programs. These programs address aging and performance issues, enhance safety features and improve security, and determine whether to reuse, refurbish, or replace a weapon's components to extend its estimated service life. However, it has been argued that Lifetime Extension Programs need validation as systems continue to evolve from their original tested and trusted configurations. Test experts emphasize that there is no guarantee that even small deviations from the original construction in the United States' highly yield-to-mass-optimized designs might not lead to failures "owing to strong nonlinearities in

---

[78] Secretary of Defense James Mattis in the preface to the *2018 Nuclear Posture Review* (US Department of State): "In no way does this approach lower the nuclear threshold. Rather, by convincing adversaries that even limited use of nuclear weapons will be more costly than they can tolerate, it in fact raises that threshold."

the system dynamics and cumulative and cooperative effects arising from various sources."[79] In the recent words of those same LANL weapon testers:[80]

> It is important to note that legacy warhead designs often had to be highly optimized so as to achieve the required yield while satisfying tight limits on the weight and size of the delivery system and rigorous safety and security requirements. In highly optimized designs, small defects can seriously impair performance. The result is that some small details (but which ones?) must be accounted for in making predictions. For nuclear weapons this can be a very difficult undertaking, and its eventual success in the absence of relevant data cannot be assured.

In congressional testimony, former secretary of defense and secretary of energy James Schlesinger argued that a decline in stockpile reliability matters greatly. In Schlesinger's words, the United States "has both acquired and had thrust upon it international responsibilities. It is still pledged to hold a nuclear umbrella over its NATO allies and Japan. It has a semi-commitment also to hold an umbrella over other states, possibly including those non-nuclear states that have signed the NPT. Its forces are stationed in many countries." Under such circumstances, he concluded that "if confidence in the reliability of the U.S. nuclear deterrent were to decline, other nations that have been content to rely on American protection might feel impelled to seek their own protection."[81]

Nuclear adversaries could exploit imperfect monitoring capabilities and CTBT ambiguities. With cavity decoupling and other techniques in an undetermined geology, explosions may be masked. While the IMS and USAEDS have demonstrated a capability to detect low-yield tests,[82] the 2012 National Research Council (NRC) report concedes, somewhat begrudgingly,[83] that Russia or China might well be able to conduct clandestine tests of up to two kilotons of yield. Russian clandestine testing, in particular, can lead to a disadvantage and surprise that undermines deterrence. Because China's arsenal is, in general, smaller and based on simpler designs, and because China has conducted only forty-five tests between 1960 and 1990,[84] China might benefit the most by exploiting this opportunity to modernize its designs.

Ambiguity in the CTBT—a treaty that the United States has not ratified but to which it is conforming—leaves the door wide open to the conduct of tests that may advance adversary military nuclear technology. The treaty is often claimed to be a "zero-yield" instrument, but in fact it does not specify any specific yield number that might describe its violation. According to the US interpretation, "under the CTBT, supercritical hydronuclear tests (which produce a self-sustaining fission chain reaction) are

---

[79] Hopkins and Sharp, "Scientific Foundation Eroding," 24.

[80] Hopkins and Sharp, "Scientific Foundation Eroding," 24.

[81] *Safety and Reliability: Hearing 105–267*, 7–8 (testimony of James Schlesinger, former defense secretary, former chair of the Atomic Energy Commission, and former energy secretary).

[82] CTBTO/IMS has demonstrated capability in the detection of North Korean nuclear tests of a few kilotons. Richards and Kim, "Advances in Monitoring."

[83] NRC, *Technical Issues*, 10.

[84] Medalia, *"Safeguards" and Net Assessments.*

banned by the Treaty, but subcritical hydrodynamic experiments, which do not produce a self-sustaining fission chain reaction, are permitted."[85] Thus, the so-called "zero-yield" treaty actually permits some nuclear release in a subcritical test. Historically, the US interpretation of the hydronuclear test regime views a violation to have occurred only at an energy release greater than two kilograms of high explosive equivalent. However, senior Russian officials have stated that the nuclear energy release of permitted experiments might range up to a metric ton.[86]

According to the US State Department Bureau of Arms Control, Verification, and Compliance (AVC),[87]

> China maintained a high level of activity at its Lop Nur nuclear weapons test site throughout 2019. China's possible preparation to operate its Lop Nur test site year-round, its use of explosive containment chambers, extensive excavation activities at Lop Nur, and lack of transparency on its nuclear testing activities—which has included frequently blocking the flow of data from its International Monitoring System (IMS) stations to the International Data Center operated by the Preparatory Commission for the Comprehensive Nuclear Test-Ban Treaty Organization—raise concerns regarding its adherence to the "zero yield" standard adhered to by the United States, the United Kingdom, and France in their respective nuclear weapons testing moratoria.

Johnny Foster, an eminent nuclear weapon designer, stated that hydronuclear tests "of less than one ton" yield could provide high confidence in the "performance [of nuclear weapons] at low yield,"[88] although the value of hydronuclear tests would seem limited to the initial fission explosives, as the utility of hydronuclear testing diminishes once boosting is part of the mix.[89]

Major Arguments against Resumption of Testing

Other states would inevitably also resume testing, making the world more dangerous. If the United States were to resume underground nuclear testing, there is no reason to believe that its principal adversaries—Russia and China—would not follow suit, perhaps after some period of time during which they castigate the United States for taking such a reckless, warmongering action and forcing their hands, however reluctantly, to do likewise. Clearly, this action–reaction dynamic could precipitate another nuclear race among the great powers, which would be extraordinarily expensive, and which the United States might even lose. Further, absolutely no thought has been given to how such a nuclear arms race might end. As a point of reference, Cold War levels of the US and Soviet Union arsenals peaked at some thirty thousand and forty thousand weapons in the mid-1960s and mid-1980s, respectively, before arms control agreements, the end of the Cold War, and the collapse of the Soviet Union led to much lower levels.

---

[85] AVC, "Scope of the Comprehensive Nuclear Test-Ban Treaty."

[86] NRC, *Technical Issues*, 103.

[87] AVC, *Adherence and Compliance*, 49.

[88] Quoted in Blank, *Russian Military*, chap. 9.

[89] Ashley, "Russian and Chinese Trends."

Of course, smaller nuclear states—North Korea, India, and Pakistan—might also resume nuclear testing. Again, there is no reason to think they would not. Moreover, while the United States might stick to underground nuclear testing, other states could resume atmospheric testing in response. Currently, if nuclear testing were resumed, nonnuclear states might feel less constrained to develop their own nuclear weapons as well. And they could be further motivated by the geopolitical instability associated with a renewed nuclear arms race.

Because the United States has conducted the most nuclear tests, with Russia not far behind, some argue that the United States has the most to lose in a world with unconstrained nuclear testing. This argument seems plausible on the surface, but accurately evaluating it requires access to classified information. In addition, it appears that Russia is already deploying low-yield tactical weapons and high fusion-fraction weapons, which suggests that the United States might benefit more from a resumption of testing than its principal adversary would.

US nonproliferation leadership would be undermined, condemnation by the rest of the world provoked, and US bipartisan support for nuclear policy threatened. Without a compelling US justification for resuming nuclear testing, all these predictions seem eminently plausible.

Article VI of the NPT commits all parties to the treaty to "pursue negotiations in good faith on effective measures relating to cessation of the nuclear arms race at an early date and to nuclear disarmament, and on a treaty on general and complete disarmament under strict and effective international control."[90] Resumption of nuclear testing could easily be portrayed as a violation of this central provision, eviscerating the United States' nonproliferation leadership regardless of whatever justification it might provide for renewed testing. Thus, after resuming nuclear testing, it is hard to imagine the United States arguing convincingly with nonnuclear states that they should maintain that status. It is far easier to envision nuclear wannabes withdrawing from the NPT, pursuing clandestine nuclear weapon development programs, or both, ultimately leading to further proliferation of nuclear testing and additional nuclear-armed states. Even nonnuclear allies, currently under the US nuclear umbrella, might abandon the NPT as the international nuclear order becomes more chaotic and they determine they must pursue their own nuclear arsenals.

It is also easy to envision our adversaries scoring propaganda points from US resumption of nuclear testing, condemning this ostensibly dangerous abandonment of moral leadership, even if some of them see a net benefit to themselves of being freer to resume testing as well. Even allies and friends might join this chorus of declared indignation. Of course, eventually the clamor of condemnation will abate and life will go on. The long-term effects on US influence around the world are hard to assess but are not likely to be helpful.

To date, support for the nuclear deterrent, modernization of the nuclear complex, and arms control has been bipartisan, with no indications of pending political peril to its continuation. Resumption of testing through presidential edict as the result of a partisan decision process is likely to upset the bipartisan congressional consensus that nuclear policy has enjoyed over many decades—however

---

[90] Treaty on the Non-Proliferation of Nuclear Weapons.

tenuous at times. We glimpsed a hint of this in the negative reactions to the *Washington Post* revelation that Trump administration officials discussed resumption of nuclear testing.[91] The intensely polarized political environment of the past few years provides little support to projections of such continued bipartisan comity, which is not worth risking by a unilateral administration decision to resume testing.

There is no need for new weapons; for now, existing ones suffice. There are a limited number of objectives for which it might be useful to have a low-yield weapon with a very small fission-fusion fraction that would not pollute the environment, a low-yield weapon penetrator, or other "new" weapons with specialized outputs. But the present stockpile inventory should be adequate for the task.

On the other hand, the adequacy of the existing stockpile will be sufficient only insofar as no common-mode failures are lurking. For example, a particular concern with old weapons relates to the plutonium pit. Spontaneous radiation from the plutonium core is an issue both for the pit itself as well as for other components, but with the closure of the Rocky Flats Plant because of environmental concerns, the United States lost ability to manufacture new pits in quantities needed in the near future. Two former LANL directors, Siegfried Hecker and Terry Wallace, have expressed concern over unresolved aging plutonium issues and the performance of aged pits in a high-radiation environment, and Donald Cook, a senior NNSA official, is concerned with the buildup of helium in the plutonium metal matrix.[92] Thus, it might be argued that plutonium pit degradation holds the possibility of a common-mode failure that would require testing to revalidate.

Nevertheless, a 2007 JASON report estimated that plutonium pits are expected to last eighty-five years without significant performance degradation.[93] In addition, a Department of Energy initiative is ramping up to meet the need for manufacture of new pits.[94] However, other technical experts have raised issues regarding plutonium aging, noting that it may be more directly—and cheaply as compared with nuclear testing—addressed by devoting the resources necessary to update our understanding. The latter step was called for in the JASON report, but apparently, to date, the NNSA has not prioritized it.[95]

We finally raise one additional argument, rarely, if ever, made in the literature, for one advantage enjoyed by the legacy US arsenal developed during the Cold War. Skewed toward high yields, and with a dearth of proportionate-response tactical weapons in its deployed stockpile,[96] its retaliatory employment during a conflict must not occur, must be limited to available low-yield weapons, or must include high-yield weapons. Adversaries contemplating a nuclear first use with lower-yield nuclear weapons must consider the likelihoods of each of these responses. As counting on no retaliation or

---

[91] Hudson and Sonne, "Trump Administration Discussed."

[92] See Kramer, "Concerns about Aging Plutonium," 24.

[93] Hemley et al., *Pit Lifetime*.

[94] NNSA, "Plutonium Pit Production."

[95] Caldwell, letter to the NNSA. The letter reviews progress on pit aging since the 2007 JASON study and responds to questions on what is needed to estimate plutonium lifetime on a sound scientific basis.

[96] Frankel, Scouras, and Ullrich, *Nonstrategic Nuclear Weapons*.

a severely limited retaliation seems excessively risky, this might enhance deterrence by making an escalatory retaliation more plausible.

The Stockpile Stewardship Program provides sufficient confidence in the US stockpile. There is universal acknowledgment that the Stockpile Stewardship Program has led to increased scientific understanding of the weapon detonation process. The new experimental facilities such as the DARHT facility for analyzing primary implosion physics and the NIF for investigating radiation-hydrodynamic phenomena associated with nuclear detonations, along with a new generation of supercomputers, have given our nuclear scientists unparalleled insight into nuclear weapon physics. Proponents of the program argue that virtually all lifetime extension modifications can now be validated with nonnuclear aboveground tests and that unique design components that can only be tested in an underground test have not undergone significant change.

Nevertheless, some scientists have been warning of looming problems with increasing urgency, as weapons continue to evolve away from original configurations. As recently as 2019, LANL technical experts concluded that "it has not been demonstrated that SSP-based results are, or will be, sufficient to supplant nuclear tests as a source of information that is indispensable for assessing the nuclear performance of the weapons in today's stockpile in a credible and trustworthy manner."[97] With the continued evolution of system configuration, the situation only grows worse. The attitude of many of the most knowledgeable scientists was pithily summed up by Merri Wood of LANL who suggested a stewardship program without testing "was a religious exercise, not science."[98]

The annual certification process is also problematic. Before the Stockpile Stewardship Program, the great majority of scientists, including the laboratory directors of the era, probably opposed a testing moratorium. After inception of the program, with its large investment in new experimental, fabrication, and computing facilities, the three laboratory directors have all been supportive. Without cynically assuming a direct connection, it is clear such a certification process must be inherently compromised because of a conflict of interest, as the laboratories' financial viability significantly depends on their directors' judgment calls that all is well. Still, the sufficiency of the Stockpile Stewardship Program is a highly specialized technical argument, and it is hard for nonscientists, or even scientifically trained nonspecialists, to evaluate its claims.

In any event, the Department of Energy's designated experts, the three nuclear laboratory directors, have certified the program as providing the confidence required without needing to test.

Underground tests will inevitably create health risks to civilian populations. Health risks from radioactive contamination may manifest either when radioactive products vent into the atmosphere or when highly radioactive residue leaks into the hydrological system, contaminating the regional environment.[99] Various studies have documented excess cancer deaths and other health hazards from

---

[97] Hopkins and Sharp, "Scientific Foundation Eroding," 23.

[98] Quoted in Glanz, "Testing the Aging Stockpile."

[99] OTA, *Containment of Explosions*.

prolonged exposure to relatively low levels of radiation.[100] Resuming underground testing could well result in increased civilian deaths.

While some theoretical flow models suggest a low probability of aquifer contamination due to flow out of the Nevada Test Site,[101] a *New York Times* article expressed concerns, noting[102]

> Studies in recent years have found that radioactive particles like long-lived plutonium 239 can travel with water, and that water is flowing more rapidly beneath the site than was once believed. Scientists now agree that contaminated plumes have the potential to flow beyond the borders of the 1,573 square-mile test site in south-central Nevada, toward populated areas. The trouble is that no one knows how big the plumes are, where they have already traveled or what exactly they contain. Scientists from the United States Geological Survey and the University of Nevada say that a witch's brew of radionuclides could take as little as a decade to reach well water in Beatty, a town of 1,500 people in the Oasis Valley about 25 miles from the heavily contaminated northwest corner of the test site. "Could it show up there in the next 10 years?" Randell Laczniak, a Geological Survey hydrologist and a coauthor of a 1996 report on ground water at the test site, said in an interview. "There's that possibility. Will it show up at a dangerous level? I don't know."

Those wishing to allay concerns over atmospheric venting point to the more than eight hundred underground nuclear tests conducted to date, from which we have gained much experience. An Office of Technology and Assessment (OTA) report offers a perspective on the health risk between 1970 (when Baneberry vented) and 1989 (the date of its report): "If the same person had been standing at the boundary of the Nevada Test Site in the area of maximum concentration of radioactivity for every test since Baneberry (1970), that person's total exposure would be equivalent to 32 extra minutes of normal background exposure (or the equivalent of 1/1000 of a single chest x-ray)."[103]

But post-Baneberry experience is hardly the whole underground testing story. Before Baneberry, a number of containment failures released considerable amounts of radioactivity into the atmosphere. While the total release after Baneberry amounted to 54,000 curies, before 1970, 12,300,000 curies had vented, obviously posing a risk to human health.[104] As well, the disappearance over time of expert personnel experienced in the arcana of containment of underground nuclear explosions, as much an art as a science,[105] does not increase confidence that such tests, after a hiatus of more than thirty years, will be immediately well contained. The risk of containment failure—either immediate or through longer-term hydrology—with its inevitable health and environmental consequences, cannot be ignored and can be avoided by continuing adherence to a testing moratorium.

---

[100] IPPNW International Commission and IEER, *Radioactive Heaven and Earth*.

[101] See, for example, US Department of Energy, *Regional Groundwater Flow*.

[102] Forstenzer, "Concerns Arise over Aquifer."

[103] OTA, *Containment of Explosions*, executive summary.

[104] OTA, *Containment of Explosions*, 4.

[105] The "artistic" component of nuclear testing is discussed in greater detail in this report's section on test readiness.

Our Bottom Line: To Test or Not to Test?

We have laid out what we believe are the most compelling arguments advanced to support or oppose resumption of nuclear testing. We reviewed a sixfold taxonomy of technical objectives for which, historically, the United States has conducted nuclear tests. Before offering our analysis and bottom-line judgment, it seems fair to also inquire: If the US government, persuaded perhaps by one or more of the pro-testing arguments, were to decide to resume nuclear testing, what sort of tests might it conduct?

We dismiss immediately that testing might be resumed for three of the six historical technical reasons. First, there has been no serious call for the resumption of the unsuccessful and environmentally fraught Plowshare Program. Second, the science of long-range test monitoring has advanced considerably beyond any urgent need for new test calibration data, and none of the arguments for resumed testing cited above mention it. Finally, absent some sudden discovery of a presently unperceived common-mode failure vulnerability, safety tests would also not seem relevant to the decision under deliberation.

After due consideration, we reject, as well, a fourth technical reason for testing. We assess that we can live with current uncertainties in weapon effects, the largest of which concern electromagnetic pulse and secondary nuclear effects such as fire.[106] These effects were never (despite a number of attempts) incorporated into the VNTK target damage methodology[107] employed by US Strategic Command strike planners and are unlikely, or are simply not able, to be resolved by underground nuclear testing in any event.[108]

We are thus left to consider the final two historical purposes of testing: maintaining the stockpile and supporting the development of new weapons. Not surprisingly, these two purposes are at the core of the two most important contending arguments for and against resumption of nuclear testing, and we will discuss them next in the context of those arguments. This will be followed by our analysis of the issue of potential cheating and CTBT ambiguity. We then address the remaining arguments—some of which are also the weightiest arguments—against resumption of nuclear testing. Finally, we will present our answer to the first question posed in this report: Should the United States resume nuclear testing?

Maintaining the stockpile. The argument that testing is needed to ensure the reliability of aging US warheads presumes that high—or even exquisite—reliability is necessary or extremely desirable, as maintained by Schlesinger and others. Yet, if we are aware of—or suspect—weapon system reliability issues, operational measures may be taken to mitigate their impact. For example, we could assign weapons judged more reliable or of higher yield to targets deemed more critical. Or we could double up on highly critical targets. For targets that are neither sufficiently urgent nor important, we could use information from post-strike nuclear detonation detection systems to allocate a second weapon, if necessary, to restrike the target. Such mitigation measures have been taken in the past when we uncovered reliability issues in the stockpile.

Moreover, while high reliability in our stockpile is better than low reliability, very little analysis has been published on how much reliability is enough. Can we settle for, say, 70 percent reliability that a

---

[106] Frankel, Scouras, and Ullrich, *Uncertain Consequences*.

[107] Binninger, Castleberry, and McGrady, *Mathematical Background and Programming Aids*.

[108] Other speculative effects, such as the suggested possibility of a nuclear winter or ozone depletion, are not amenable to resolution by an underground test and would not make a very compelling rationale for resumption of testing even if they were.

particular weapon in our arsenal is reliable, or do we need 90 percent reliability or even 99 percent reliability? Do we need the same reliability in all weapon types? And by what logic are reliability requirements determined? Also, it seems plausible that our adversaries will have significant uncertainty in our arsenal's reliability and a propensity to err on the side of caution, which also provides some leeway in being able to accept imperfect reliability. Finally, without clandestine testing, our nuclear adversaries will also experience increasing concerns with their own arsenals' reliabilities. We need to ask whether an increase in our own weapon reliabilities due to a resumption of testing is worth a concomitant increase in reliability in the weapons of our adversaries.

The Stockpile Stewardship Program has provided invaluable insights into the dynamics of nuclear explosions and stockpile reliability, and we do not believe we are at the point where it has failed in its goals.

Our conclusion is not that reliability is unimportant, but rather that the virtues of exquisite nuclear weapon reliability and correctly assessing that reliability should not be presumed and should not trump other considerations in addressing the question of whether or not the United States should resume nuclear testing.

At the same time, on balance we reject the technical arguments presented in favor of the Stockpile Stewardship Program being sufficient to forever maintain confidence in stockpile performance in the absence of any nuclear testing. The history of real-world experience of testing surprises is simply too much to overcome. Too many things have gone too wrong too often to trust Lucy with the football[109] one more time. Moreover, we are uncomfortable with the stockpile certification process. It seems improbable to us that the laboratory directors, in making their determinations, would be immune to inevitable pressures to certify the stockpile as reliable. Even the perception of mixed motives is enough to warrant a revised certification process.

In sum, the Stockpile Stewardship Program has provided invaluable insights into the dynamics of nuclear explosions and stockpile reliability, and we do not believe we are at the point where it has failed in its goals or must be supplemented by nuclear testing, although we also believe that eventually that point is likely to be reached. We do suggest that if the United States ever resumes nuclear testing, it consider a new technical objective: validating Stockpile Stewardship Program modeling. Many measurements of the initial phases of a nuclear explosion that might have been obtained during the test era to strengthen confidence in the program were simply not taken. Past tests too often were rushed to qualify intricate device design details but with limited diagnostics, insufficient to fine-tune computational codes without resorting to "fudge factors" of questionable predictive legitimacy. Should testing resume for some reason, each test should be designed in a way to also resolve any outstanding physics issues, thereby increasing confidence in our computational tools, helping to minimize the need for further tests, and preparing for the day when the country might again impose a moratorium on nuclear testing.

Supporting the development of new nuclear weapons. We share the concerns expressed about emergent threats, principally from Russia, composed largely of very low-yield tactical nuclear weapons with reduced collateral effects, perhaps some with specialized outputs.[110] The ability to respond in kind to nuclear aggression that employs these kinds of weapons without escalating to

---

[109] A reference to Charles Schulz's comic strip *Peanuts* (see https://www.comicartfans.com/gallerypiece.asp?piece=995507).

[110] Frankel, Scouras, and Ullrich, *Nonstrategic Nuclear Weapons*.

high-yield weapons provides an essential contribution to the credibility of deterrence. Currently, however, the United States maintains neither sufficient numbers nor types of tactical weapons on survivable platforms and bases to adequately counter and deter the escalate-to-deescalate, or E2D, ladder of (alleged) Russian military doctrine.[111]

But we are not convinced that the United States would need to mirror the Russian arsenal in this regard to maintain an effective deterrent. In fact, some lower-yield warheads are now coming into the US arsenal. A low-yield version of the W76 warhead, designated the W76-2, has been mated with the Trident II D5 missile and is being deployed on *Ohio*-class ballistic missile submarines. The B61-12 bomb, a lower-yield and more accurate progeny of its predecessor B61 variants, soon will be deployed on the dual-capable version of the F-35 aircraft with forward basing in several NATO countries. While these two are not specialized output weapons, we believe they will partially close the gap in US retaliatory options. Moreover, we do not think very many such weapons would be required. After a round or two of nuclear exchanges confined to a theater, it seems that the strategic arsenal might well be called on to respond to further aggression.

The dangers of loss of credibility of protection under the US nuclear umbrella can motivate our allies to develop their own nuclear arsenals. As former secretary Schlesinger testified, given the United States' unique geopolitical status and NATO's and other allies' reliance on US constancy, "If confidence in the reliability of the U.S. nuclear deterrent were to decline, other nations that have been content to rely on American protection might feel impelled to seek their own protection."[112] The biggest need we see to buttress extended deterrence is modest numbers of weapons with sub-kiloton yield deployed on a survivable platform that is not based in the continental United States, such as sea-launched cruise missiles on the US fleet of attack submarines. However, while the latest Nuclear Posture Review[113] argues for such a weapon, consensus is lacking; US attack submarines are currently not configured to carry such a weapon; and the weapon itself has not been designed. So, at this point, we are a long way from a potential need for testing. And advocates for testing such a warhead would have to explain how it is that Russia seems capable of deploying such weapons without testing.

We view the loss of credibility of the nuclear umbrella as a more dangerous threat to stability than lowering the nuclear threshold.

On the other hand, we reject the argument that introduction of new, more "usable" weapons is inherently destabilizing because it lowers the threshold for nuclear first use. We have already addressed the problem of deterring adversary nuclear first use with low-yield nuclear weapons. This argument, by contrast, addresses US nuclear first use. Would the United States be more likely to undertake a first strike with nuclear weapons if it had lower-yield weapons with reduced collateral effects? Perhaps so, in some scenarios. Of course, the decision to use such weapons would be up to the United States. It could just say no. Thus, the source of concern with more "usable" weapons in the US nuclear arsenal can be isolated to not trusting US leadership to wisely use—or not use—the options such weapons would provide. However, with US overall military superiority, US leadership would have many options to draw from other than resorting to nuclear first use.

---

[111] Ryan, "Is 'Escalate to Deescalate' Part of Russia's Nuclear Toolbox?"

[112] *Safety and Reliability: Hearing 105–267*, 7–8 (testimony of James Schlesinger).

[113] US Department of Defense, *2018 Nuclear Posture Review*.

Thus, in summary, we view the loss of credibility of the nuclear umbrella as a more dangerous threat to stability than lowering the nuclear threshold. The United States risks being perceived as self-deterred from actual employment of the present stockpile, which is skewed toward large-yield weapons, should adversaries threaten allies who have (well-founded?) waning confidence that the United States would risk initiating a strategic exchange on their behalf. The larger threat of employment of such "usable" weapons comes from adversaries already possessing them to the United States' disadvantage when it has no equivalent option to counter.

Implication of CTBT ambiguity and limited monitoring capability. While it is disconcerting that Russia does not adhere to the US definition of zero yield in the CTBT, we can hardly believe this provides adequate justification by itself or in combination with other considerations for resuming kiloton-scale nuclear testing. Importantly, it is also unclear that Russia could achieve any military advantage over the United States by exploiting nuclear testing up to the level of its definition of zero yield. And were the United States so concerned with that possibility, all it would need to do is adopt Russia's definition and conduct hydronuclear testing accordingly.

Rather, the argument that the United States should resume testing because of its limited ability to monitor adversary activity lies squarely with concern that the country is presently unable to verify whether Russia or China may be testing clandestinely above the kiloton level. It is argued that such a level of test activity, accomplished through active measures that might mask the long-range detection of a nuclear signal, is sufficient to support the development of new thermonuclear weapon designs that would disadvantage us on the battlefield. It was principally for this reason, an inability to adequately monitor Russian test activity, and technical disputes about the ability to discriminate between earthquakes and nuclear test signals from long range, that negotiations over what eventually resulted in the Limited Test Ban Treaty of 1963 dragged on for over five years.

Nevertheless, the argument that we should start testing now because of a clandestine activity that *might* be taking place is a difficult one to make. To the National Academy's assessment of the difficulty of high-confidence concealment at the kiloton test level, we might add the observation that if tests were being carried out for the purpose of new weapon development, the equivalent US experience indicates that more than a single clandestine test would be needed. Before new designs entered the US stockpile, typically many underground tests were required to validate their effectiveness and safety, and the confidence that a series of tests could be executed clandestinely would plummet accordingly. While certainly much physics could be learned or proven by a single "small" thermonuclear test, it is uncertain how far this would advance the development of an operational new weapon. It is for these reasons we believe the as yet unproven concern of clandestine testing is quite insufficient to justify the United States' resumption of nuclear testing at present.

We conclude that, at present, the United States should not resume nuclear testing. For now at least, let the sleeping dragon lie.

In any event, even if the United States can live with different US and Russian definitions of zero yield, we wonder why this issue could not have been resolved during negotiations on the treaty language. We understand that the State Department has claimed that US negotiators were aware of this unresolved definitional discrepancy but decided to accept the ambiguity, possibly presuming that it could be worked out in the implementation process or that it would not be a big deal. At the least, an independent body should look into what, if anything, went wrong and take the appropriate lessons for future negotiations.

Additional arguments. We are in agreement with the remaining three arguments against resuming nuclear testing: (1) other states will inevitably also resume testing, making the world more dangerous; (2) US nonproliferation leadership will be undermined, condemnation by the rest of the world provoked, and US bipartisan support for nuclear policy threatened; and (3) underground tests will inevitably create health risks to civilian populations. Regarding this last argument, while we believe that renewed health risks are likely, we believe this consideration pales in significance compared with others analyzed above, and in any event, only supports our conclusion based on those more significant arguments. By contrast, we cannot dismiss so easily the other additional arguments against resuming nuclear testing. While all forecasts are speculative, we think there is a reasonable likelihood that nuclear proliferation would increase and, possibly, also a nuclear race will be reignited.

With this analysis, our logic for coming to a determination is relatively straightforward. We assess the three arguments in favor of resuming nuclear testing as unpersuasive. We are not convinced that stockpile maintenance needs nuclear testing at this time, and while it well might at some point in the future, we cannot reliably predict when that might be. Similarly, while we support the development of low-yield nuclear weapons, simpler designs (or those based on previously tested but not weaponized designs) likely can be developed without testing, and even specialized new designs would have no need to be tested for years. Finally, we believe that the ambiguity in definitions of zero yield do not seem exploitable to achieve military advantage.

Thus, we conclude that, at present, the United States should not resume nuclear testing. For now at least, let the sleeping dragon lie. Once it is awoken, there is no plan to put it back to sleep and there will always be the opportunity to awaken it in the future.

What Might Change Our Decision?

In addition to the arguments presented above, we should consider a set of conditions that do not presently pertain but might in the future, and that would perhaps be sufficient to weight our decision calculus in a different direction.

Nuclear war or crisis. The unthinkable is not the impossible. Despite our managing to avoid nuclear war since the end of World War II, we must recognize that it remains a possibility. The probability of nuclear war is, of course, anyone's guess.[114] Crises short of nuclear war, in which nuclear weapons and associated doctrines play a major role in determining the outcome, are of course also to be expected. There have been perhaps a score of such crises since the dawn of the nuclear age.[115]

If a nuclear war or crisis occurs in the future, it is reasonable to presume that both nuclear and nonnuclear states will subsequently evaluate the adequacies of their arsenals, or lack thereof, to support their national security imperatives. How these states will respond is speculative, but a new arms race supported by a renewal of nuclear testing is a distinct possibility.

A moratorium breakout by Russia or China. There are presently no legal impediments to a resumption of underground testing by Russia or China, nor by the United States.[116] A similar situation pertained in 1958 when the then Soviet Union announced its adherence to an atmospheric testing moratorium and the United States followed suit. Several years thereafter, in 1961, evidently to the complete surprise of the Kennedy administration, the Soviet Union broke out of this self-imposed moratorium

---

[114] Scouras, "Global Catastrophic Risk."

[115] Brecher et al., International Crisis Data Set.

[116] Having signed the draft CTBT in 1997, the US Congress to date has refused to ratify it. Some—e.g., Rogoff, "International Legal Obligations"—have argued that our unratified signature alone still imposes certain legal constraints. See also Bradley, "Unratified Treaties."

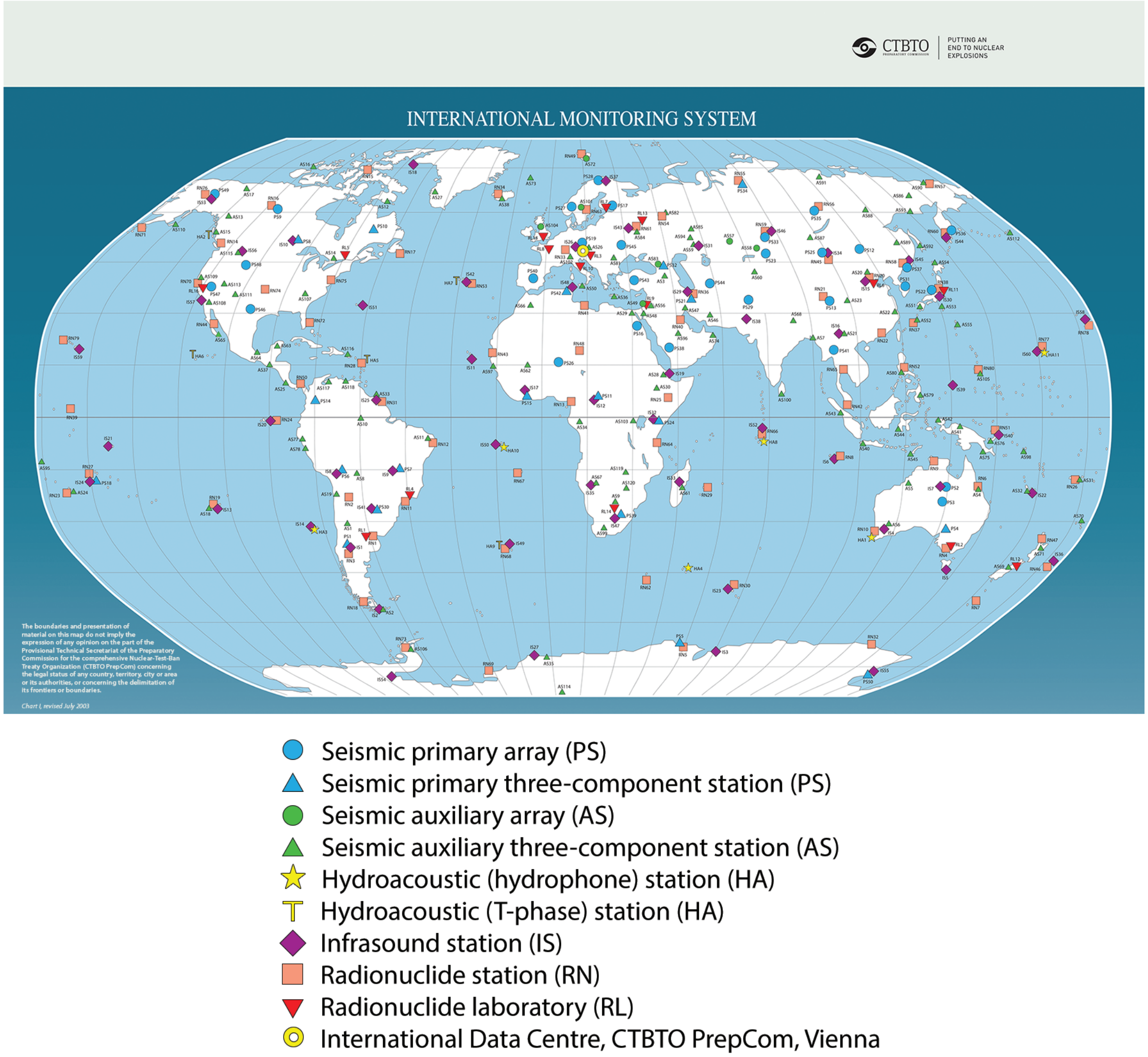


Figure 12. The IMS for Verifying Compliance with the CTBT

with an atmospheric nuclear test.[117] It (and then the United States) continued with a series of atmospheric tests until the 1963 Limited Test Ban Treaty entered into force. Thus, there is both precedent and legal opportunity for that history to repeat with any of the great powers even today, with restraints imposed mostly by political considerations, which are always subject to miscalculation and change.

Unambiguous discovery of Russia and/or China cheating. With the extensive IMS in place (see Figure 12), cheating is not easy to conceal, and in some respects, is not easy to define, even when "cheating" takes place in front of our eyes. Cheating through a clandestine underground explosion of substantial yield—say, on the order of a kiloton or more—is likely detectable by the IMS should it be carried out anywhere on Earth, even in remote, less geologically characterized, regions.[118] The National

---

[117] Tsar Bomba—at fifty megatons, the largest nuclear detonation ever to occur—was part of this breakout series of tests.

[118] Which are more seismically difficult to interpret. While the IMS claims detectability down to much lower yields—tens to hundreds of pounds—in much of the world (see, e.g., Richards and Kim, "Advances in Monitoring"), active masking efforts such as shock wave decoupling or hiding signatures in mine blasts call that capability into some question.

Academy's 2012 review of the technical issues associated with nuclear test monitoring explicitly dismissed the notion that any state not already well experienced with nuclear testing could ever hope to conceal a clandestine test of a few kilotons by employing decoupling or masking techniques. However, it implicitly acknowledged that Russia or China might be able to do so,[119] albeit with some difficulty. Then there are the hydronuclear tests,[120] which in fact produce nuclear yield. While the National Academy's 2012 review asserts that "the largest fission release was less than 0.5×10^{-8} kilotons (0.01 pounds)."[121] Russia has never accepted the US definition of a so-called zero-yield test,[122] and declarations from Russian officials support the notion that they have tested yields greater than one hundred kilograms in hydronuclear explosions and consider the hydronuclear test regime to extend up to one metric ton—although it is unknown whether they have tested to this level. This opens the way for tests reaching tens or even hundreds of pounds, which the United States has deemed to raise compliance concerns[123] regarding Russian adherence to the notification and verification protocols of the Threshold Test Ban Treaty. It is unknown whether Russia has tested to this high level, or for that matter, to any US-designated noncompliant level. According to the Department of State, "The United States assesses that Russia has conducted nuclear weapons-related experiments that have created nuclear yield. The United States does not know how many, if any, supercritical or self-sustaining nuclear experiments Russia conducted in 2019."[124]

Table 1. Range of Yields for Various Test Objectives

Yield Range	Test Objectives
<0.25 kilograms	Criticality
<1.8 kilograms (4 pounds)	Safety, plutonium equation of state
A few to hundreds of kilograms	Better signal-to-noise ratio and better margin for errors in projecting to higher yields
A few to tens of tons	Validated advanced pure-fission designs with improved yield-to-weight ratios
100–200 tons	Fusion phenomena and D-T boosting
>1 kiloton	Two-stage thermonuclear designs

Data source: Quirk, *Low-Yield Nuclear Testing*.

At some point, these differences in definitions of what is and is not allowable testing may come to a head and, failing resolution, may compel the United States to resume nuclear testing, at least within the constraints of Russia's more permissible definition of what is allowed. Unambiguous

---

[119] NRC, *Technical Issues*, Appendix E.

[120] According to the US National Academy of Sciences, the United States has historically considered the hydronuclear energy release region to top out at two kilograms, while the Soviet Union has considered it to extend to one hundred kilograms. The same US National Academy of Sciences report notes that academician Viktor Mikhailov has suggested that Russian hydronuclear could extend up to one metric ton. At the low end of the energy release spectrum, some experts make a sharp distinction between subcritical and supercritical energy release regimes, with only the latter identified as hydronuclear. Others apply the term to subcritical testing as well. Others may reserve the term *subcritical* for nonnuclear experiments, where the fissile material in a model has been replaced by something more benign, like tungsten. Such semantic niceties matter since the (unratified) CTBT and adherence to the nuclear test moratorium permit hydronuclear tests but not nuclear explosions, with similar consequential ambiguities applying to a claimed consensus understanding of the CTBT as a "zero-yield" treaty.

[121] NRC, *Technical Issues*. However, Thorn and Westervelt (in the report *Hydronuclear Experiments*) assert that an LANL hydronuclear test achieved a 0.4-pound nuclear energy release. In any event, it is clear that "zero-yield" does not mean zero, but rather a small, if ill-defined, number.

[122] See footnote 2.

[123] AVC, *Adherence and Compliance*, executive summary. Any nuclear energy release by physical breach of the explosive container requires notification and opportunity to conduct verification activities in accordance with treaty protocol. Testing at levels up to hundreds of pounds would raise such concerns of a physical breach requiring notification.

[124] AVC, *Adherence and Compliance*.

"true cheating"—the violation of the announced self-imposed adherence to even the one-ton limit by testing clandestinely at up to a few hundred tons,[125] a level (as indicated in Table 1)[126] sufficient to support development of advanced single-stage weapons—may drive political and military pressures to respond in kind.

Discovery of common-mode arsenal failure issues. It has happened before. Problems, some associated with one-point safety, some with aging components or other aspects, have, in a historical retrospective, been all too frequent. In 1996 congressional testimony, the Department of Energy identified over 1,200 "significant findings" of a defect or failure in a weapon system over the course of US nuclear testing. Of these, over 120 required redesign of US stockpile elements.[127] After the atmospheric testing moratorium of 1958, it was discovered that the B43 high-yield bomb was not one-point safe[128] and a number of other systems were similarly questionable. Senior scientific officers at LLNL report that fifteen of the United States' weapon systems employed in 1970 required *post-deployment* nuclear testing to identify or resolve problems.[129] In congressional testimony in 1997,[130] Secretary Schlesinger testified that the W80 and W88, whose initial designs still form the basis for two legs of the strategic triad, the air-launched cruise missile and the Trident, respectively, still elicit "safety concerns." Secretary of Defense Caspar Weinberger stated that "over one-third of all nuclear weapon designs introduced into our stockpile since 1958 have encountered reliability problems, and 75% were discovered and subsequently corrected thanks to actual explosive testing."[131] Because of certain commonalities of design or components, such common-mode failures could risk the integrity and reliability of the stockpile as a whole.[132] The Stockpile Stewardship Program (untested by definition and design) is responsible for preventing such a black swan in the future, but history provides neither comfort nor confidence. Should a common-mode failure be discovered in the future, there may be a national security imperative to resolve it through resumption of nuclear testing.

Failure of certification. Since the requirement was established in 1995, in the words of the Department of State:[133]

> The Directors of the three DOE nuclear weapons laboratories—Los Alamos National Laboratory (LANL), Lawrence Livermore National Laboratory (LLNL), and Sandia National Laboratories (SNL)—are required to complete annual assessments of the safety, reliability, and performance of each weapon type in the nuclear weapons

---

[125] Arguably undetectable if masking strategies are used in ill-characterized geologies.

[126] Quirk, *Low-Yield Nuclear Testing*.

[127] *Nuclear Weapons: Status of Stockpile Surveillance Program* (testimony of Victor S. Rezendes, director of energy, resources, and science issues at the Resources, Community, and Economic Development Division).

[128] It was discovered that ignition characteristics of a single point depended on its location on the explosive driver.

[129] Miller, Brown, and Alonso, *Report to Congress*, 19.

[130] *Safety and Reliability: Hearing 105–267* (testimony of James Schlesinger).

[131] Kidder, *Maintaining the U.S. Stockpile*.

[132] In particular, nuclear weapon performance would seem to comprise a classic example of what the late Charles Perrow defined in his book *Normal Accidents*, whereby complex systems that were also "tightly bound" would inevitably fail over time, as complex and unforeseen feedback mechanisms would emerge unexpectedly in unanticipated environments. The Goldsboro nuclear accident of 1961 is a great example of Perrow's prescience. In this event, an airplane in flight over North Carolina accidentally dropped two nuclear weapons, one of which failed to detonate only after three of the four safety switch interlocks had failed. The fourth safety switch interlock, which prevented a megaton-scale nuclear explosion in North Carolina, had failed numerous times in other circumstances. Two authors of this report (Frankel and Scouras) conferred with Professor Perrow about aspects of nuclear weapon phenomena as exemplary of *Normal Accident* theory.

[133] AVC, "Annual Assessment."

> stockpile. In addition, the Commander of U.S. Strategic Command provides an assessment of the military effectiveness of the stockpile. These assessments also include a determination as to whether it is necessary to conduct an underground nuclear test to resolve any identified issues.

By law these assessments are included—unchanged—in the annual report of the secretaries of energy and defense to the president of the United States.

As described previously, knowledgeable experts continue to express concerns over the long-term viability of the assessment and certification process. Certification failure may be realistically contemplated in both explicit and implicit scenarios. Explicit failure would manifest should one or more of the nuclear laboratory directors withhold their required concurrence in some future year because they lack technical confidence that the Stockpile Stewardship Program continues to ensure arsenal reliability. Implicit failure may not be as abrupt but would be no less injurious to the certification process. Such a scenario would be realized should the laboratory directors' independence come to be perceived as fatally compromised by their inherent conflict of interest over laboratory funding. In such circumstances their concurrence is likely to be widely recognized as pro forma, and the annual certification an unpersuasive political exercise.

Should certification not be achieved through this process, resumption of underground testing is explicitly cited as an option to resolve issues.

Failure of the NPT. The NPT recognizes five nuclear weapon states—the United States, Russia, the United Kingdom, France, and China. Of the total of 191 parties, all others participate as nonnuclear weapon states. The NPT's central bargain is that nonnuclear states will not seek to acquire nuclear weapons, while the nuclear states will share the benefits of peaceful nuclear technology and pursue general and complete disarmament (including nuclear disarmament). The few, but important, nonsignatories include Israel, which maintains official ambiguity about its nuclear capability but is widely believed to possess a nuclear arsenal; India and Pakistan, which have significant nuclear arsenals on the order of a hundred weapons each; and North Korea, which withdrew from the NPT in 2003 and has a nascent but growing arsenal. The treaty was extended indefinitely in 1995.

It is simply not credible that the current US nuclear arsenal, configured to meet Cold War exigencies, will forever serve its needs.

By preventing the anticipated wide-scale global proliferation of nuclear weapons in the 1960s, the NPT is considered a pillar of international security, but its future viability remains uncertain. Nonnuclear states complain that the nuclear states are not living up to their end of the bargain by not vigorously pursuing nuclear disarmament. Resumption of nuclear testing by any of the five nuclear weapon states would demonstrably support this position and could lead to a general breakdown of the treaty. Alternatively, if the treaty ceases to be a significant obstacle to nuclear proliferation, the nuclear weapon states may feel one less impediment to resuming nuclear testing.

Other forces at play would undermine the NPT. In particular, US allies around the world rely on the United States' nuclear umbrella to deter Russian and Chinese nuclear and conventional aggression and to enable them to stand up to implicit and explicit threats. Because so much is at stake, the credibility of the US nuclear umbrella is a constant and serious source of concern. If the efficacy of the arsenal on which US extended deterrence ultimately relies comes into significant doubt, at least some US allies may decide they need their own arsenals. This thinking had some part in the British and French decisions to develop independent arsenals. While US arsenal reliability is likely insufficient by itself to trigger a similar response by additional allies such as Turkey, Japan, South Korea, and Germany, it could contribute to future allies' decisions to develop their own nuclear capabilities. And, of course, if these states develop nuclear weapons, others might follow in response. In any event, eroding confidence in the reliability of the US stockpile will surely contribute to an increasingly fragile credibility of extended deterrence.

Finally, it is evident that regional rivalries can motivate nuclear proliferation. India and Pakistan provide a clear example of this dynamic. The possibility of an Israeli arsenal is a thorn in the side of other Middle Eastern states, such as Iran, that consequently aspire to acquire their own arsenals. Of course, proliferation is not a one-way street. Several states, including Brazil, Argentina, South Africa, South Korea, and Taiwan, have abandoned nuclear programs. Nevertheless, without the NPT, one can easily imagine a world with a score or more of nuclear states, at long last validating the predictions of the 1960s. And in such a world, it is also not difficult to imagine a resumption of nuclear testing by the United States or any other of the nuclear weapon states.

Emergence of a new design imperative. It is simply not credible that the current US nuclear arsenal, configured to meet Cold War exigencies, will forever serve its needs. The arsenal is dominated by thermonuclear weapons with high yields (typically one hundred kilotons or higher), and the United States is only beginning to supplement these with modified weapons of significantly lower yields (on the order of ten kilotons or lower). Even so, these weapons are few, their yields are not as low as might be desirable, and some are not deployed on platforms survivable in the more worrisome scenarios. Yet, emerging threats, increasingly composed of weapons with lower and lower yields and with minimal unwanted secondary effects (i.e., fallout), may not be deterred by the threat of retaliation with disproportionately large and dirty weapons.

We already face an increasing need to deploy an arsenal with flexibility to match the escalatory ladder of Russia's escalate-to-deescalate doctrine. Deterring or defeating Russia might commend the urgent development of capabilities such as low-yield nuclear penetrators or specialized weapons with tailored outputs that minimize collateral damage, address targets such as biological threats, or suppress electronics over a wide area.

As the threat continues to evolve to emphasize smaller and specialized weapons, the imperative to develop a new warhead (or warheads) for our arsenal may become irresistible. For example, unlike the United States, Russia has already invested heavily in development of high fusion-fraction weapons, demonstrating a specially designed fifteen-kiloton "device" with 98 percent fusion output during the course of its Peaceful Uses of Nuclear Energy program.[134] Should Russia actually field a weapon with near-pure fusion device capability, the tactical, strategic, and political advantages conferred could mandate some response by the United States. In addition, the advantages of a warhead with long-lived stability, low-cost maintenance, ease of modification, and presumed lowered stress on the certification process, such as embodied in development of the Reliable Replacement Warhead,[135] or the tactical advantages of a nuclear penetrator, may commend themselves. Development of such weapons, which would deviate significantly from current designs, may not be feasible without testing. Thus, these or other developmental black swan imperatives may point to a need to resume underground testing.

---

[134] Nordyke, *Soviet Program*.

[135] Canceled by the Obama administration in 2009.

Test Readiness

In the run-up to the 1963 Limited Test Ban Treaty, the Joint Chiefs of Staff, as price for their public support, demanded and received a series of assurances that the US government would pursue four safeguards:

(A) The United States would continue to pursue a "comprehensive, aggressive, and continuing" underground nuclear test program.

(B) The United States would maintain human and laboratory resources to ensure continued progress in nuclear technology.

(C) Should it be deemed necessary for national security, or should the Soviet Union abrogate terms of the treaty, the United States would maintain facilities and capabilities required to "promptly" resume atmospheric testing.

(D) The United States would continue to improve, within feasible and practical limits, its ability to monitor the Soviet Union's and China's treaty compliance and maintain knowledge of their nuclear activity.

These guarantees, deemed "unqualified and unequivocal assurances," were provided to the Senate as part of a presidential letter dated September 10, 1963, and were instrumental in securing Senate ratification two weeks later.[136]

In 1976, the Ford administration updated Safeguard C, relaxing the requirement for "prompt" resumption of atmospheric testing and replacing it with a standard that ensured "the maintenance of the basic capability to resume nuclear testing in the atmosphere."[137] This relaxation occurred in the context of the signing of the Peaceful Nuclear Explosions Treaty with the Soviet Union that year and was driven by the satisfactory experience with the underground test program to date and a desire to reduce the costs of maintaining a "prompt" testing posture.[138] President Bush's 1990 letter to the Senate on the Peaceful Nuclear Explosions Treaty and the Threshold Test Ban Treaty[139] further modified Safeguard C to omit reference to atmospheric testing.

In 1994, after the start of the still-continuing voluntary nuclear test moratorium, the last official modification to the safeguards[140] converted Safeguard C into an assurance of readiness to conduct underground tests only. The new language explicitly barred the use of any funds "to maintain the capability of the United States to conduct atmospheric testing of a nuclear weapon." In negotiations as part of the debates over attempts to ratify the CTBT during and after the Clinton administration, the issue of safeguards remained important to the Senate and the Joint Chiefs. Various tweaks to its language have been proposed, including the explicit addition of a new safeguard to ensure conduct of a stockpile stewardship program. But the CTBT has not, as of this writing, been ratified, and the 1993 version of the safeguards, including Safeguard C, which mandates US preparation to resume underground nuclear testing, remains the current legal standard.

Complementing the legislative language embodied in the 1994 public law, President Clinton issued an implementing presidential directive (PDD-15,

---

[136] Kennedy, letter to Senate leaders.

[137] US Department of Energy and US Department of Defense, "Memorandum of Understanding," B-1.

[138] In 1990, in a sign of the times as the Soviet Union neared collapse, the US Senate finally ratified the Peaceful Nuclear Explosions Treaty after the administration consented to an updated Safeguard A that replaced the 1963 language assuring the conduct of a "comprehensive, aggressive, and continuing" underground test program with the less-aggressive sounding assurance of "the conduct, within the constraints of treaties on nuclear testing, of effective and continuing underground nuclear test programs."

[139] Bush, letter to the US Senate.

[140] National Defense Authorization Act for Fiscal Year 1994, Pub. L. No.103-160, § 3137. The same act also created the Stockpile Stewardship Program.

"Stockpile Stewardship")[141] calling for the Department of Energy to maintain a capability to perform an underground test within twenty-four to thirty-six months, should such testing be deemed necessary. Presently, official requirements for test readiness consist of the following:

- Six to ten months for a "simple test" with minimal diagnostics and environmental and safety procedural waivers
- Two to three years for a fully instrumented stockpile stewardship test
- Five years for a test to develop new capabilities

Since the United States last conducted an underground nuclear test in 1992 at the Nevada Test Site,[142] the test teams have long since dispersed and the associated firsthand knowledge base has atrophied. Moreover, most of the equipment, facilities, and supporting infrastructure have long since fallen into disuse and would have to be reconstituted. In light of that, it is fair to ask whether the timelines articulated above in the 2017 Stockpile Stewardship and Management Plan are realistic.

While testing a nuclear weapon underground is an extreme exercise of big science, it is also in part an art. And it is not only the device designers who are part of the art but also other uniquely accomplished technical specialists who may not be as familiar to the public. Every individual underground test is unique in terms of geology, undetected rock faults, unexpected vagaries of weapon performance, containment challenges, stemming and grouting, grounding and shielding, data acquisition design, and various emplacement issues. Although the experienced and expert national laboratory personnel conducted over eight hundred underground nuclear tests, they did not all contend with containment failures or data acquisition failures to the same degree. Some tests, such as the early Baneberry vertical line-of-sight event conducted by the Atomic Energy Commission (Figure 13) or the 1975 Department of Defense Des Moines horizontal line-of-sight effects test, vented catastrophically, whereas lesser radioactive contamination events or data loss occurred more often.

Figure 13. 1970 Baneberry Event Venting through Undetected Rock Fissure

At its peak during the Cold War, there were over seven thousand personnel on-site at the Nevada Test Site and over one hundred thousand personnel as part of the supporting industrial infrastructure nationwide. These are mostly gone. According to the NNSA, much, if not most, of the equipment and technology required for nuclear testing in the past has not been adequately maintained, is obsolete, or has been sold or salvaged. More important, the knowledge needed to conduct a nuclear test, which comes only from testing experience, is all

---

[141] White House, *U.S. Policy on Stockpile Stewardship*.

[142] Since renamed the Nevada National Security Site.

but gone too.[143] In the words of John C. Hopkins, retired associate director of LANL, "In sum, there is essentially no test readiness. The whole testing process—whether to conduct one test or many—would in essence have to be *reinvented*, not simply resumed."[144] Given that assessment, some have questioned our current capability to satisfy the two- to three-year timeline mandated by Presidential Decision Directive 15.

While we recommend against resumption of nuclear testing, we also believe it would be prudent to take steps to enhance test readiness.

As noted by a former director of the Defense Nuclear Agency, "The tens of thousands of active nuclear weapons scientists in our three nuclear weapons labs have never designed, tested and built a nuclear weapon."[145] With the dissipation of testing expertise and infrastructure, reacquiring such capability would take time. In the event of a military crisis or discovery of a common-mode failure affecting all weapons of a particular design—which has happened before—the US nuclear arsenal could be compromised at a time it is most needed.

Recommendations

The primary recommendation we have advanced in this report is that we should not resume nuclear testing at this time. However, while we recommend against resumption of nuclear testing, we also believe it would be prudent to take steps to enhance test readiness. We think it reasonably plausible that one or more of the future conditions discussed above could occur, which might then weigh the decision in favor of resumption of nuclear testing. We offer the following possibilities to further the debate on testing and hedging; each of these deserves a more complete analysis before an informed decision can be made.

First, the United States should consider relaxing its interpretation of the CTBT limits to be consistent with the Russian definition that allows tests ("experiments") of very low yields. If Russia is operating under the assumption that tests with yields possibly as much as a ton are permitted, what would be the ramifications should the United States also adopt that position?[146] The United States should develop a testing plan that would exploit such a policy change and, *inter alia*, determine the extent to which such testing would reduce the need for larger-scale nuclear testing. An additional benefit of this reinterpretation is that it would remove the United States' constant accusations of Russian cheating, an irritant to both the United States and Russia. However, if the United States does take this step, it will be important to couch it in terms of resolving conflicting interpretations of the CTBT rather than opening the door to additional testing.

Second, the United States should more openly acknowledge the limitations inherent in and the potential for failure of the stockpile stewardship program. Modeling and laboratory experimentation must eventually be validated through testing. The United States should develop plans to mitigate these limitations, as well as to respond to the (currently unlikely) event that the Department of Defense, the Department of Energy, or both no longer certify the arsenal. Such a plan might include nuclear weapon testing at some point in the future. Alternatively, or in addition, it could include the development of a low-maintenance replacement warhead with design margins sufficient to virtually guarantee its reliability. Finally, it could identify operational measures meant to accommodate weapons with uncertain reliabilities.

---

[143] Hopkins, "Nuclear Test Readiness."

[144] Hopkins, "Nuclear Test Readiness," 10.

[145] Monroe, "Nuclear Weapons May Not Work."

[146] Weapon designers are not in complete agreement about the benefits that might be derived from such testing.

Third, the United States should revamp its current annual nuclear stockpile certification process. This would include replacing the requirement that the secretaries of energy and defense, informed by the judgments of the nuclear weapon laboratory directors, annually certify the arsenal with something less susceptible to perceived conflicts of interest and political pressures. One option could be establishing a standing independent review body, under the auspices of the National Academies of Science and Engineering and in consultation with the Joint Atomic Energy Intelligence Committee. Additionally, the United States should develop procedures for eliciting and reporting dissenting viewpoints from knowledgeable individuals.

Fourth, the United States should take more seriously the possibility that it may choose to test in the future, possibly to validate a new weapon design or as a political response to Russia or China resuming testing. This involves increasing confidence in our ability to execute the PDD-15 requirement, if called on to do so. The United States should establish a planning activity to coordinate all aspects of test planning, including issues of location selection, site preparation, device yield and design, device emplacement, containment, data capture, seismic mitigation if necessary, and identification of critical infrastructure and personnel resources. Past experience with a moratorium breakout demonstrated the downsides of hasty test execution without sufficient prior thought to actual needs. Moreover, this planning activity should be linked to the national laboratory prototyping processes.

A Final Thought

Perspectives on whether to resume nuclear testing can be understood in the context of the United States' conflicting far- and near-term goals with respect to nuclear weapons. Since nuclear weapons uniquely pose a mortal threat to the United States, in the far term many aspire to a world without such weapons, even if that state remains ill defined and there is no clear vision on how to achieve it. By contrast, in the near term there appears no viable alternative[147] but to maintain the nuclear peace through deterrence, underwritten by an effective nuclear arsenal.

As a nation, we have not come to terms with balancing these goals. As the Chiles Commission observes:[148]

> It is thus imperative that the nation's long-term commitment to maintaining an effective, safe, and reliable deterrent be powerfully and clearly emphasized by the nation's leaders. Part of the challenge is to distinguish this commitment from goals or hopes stated by individuals in and out of government that nuclear weapons may be eliminated over the long term. The distinction between long-term political goals and nearer-term programmatic goals is a critical one to the sense of mission within the nuclear weapons program.

Failure to fully understand this distinction has resulted in unproductive arguments about the recapitalization of the nuclear triad, arms control, launch-on-warning policy, and so on. The same holds true for nuclear testing. The main issue with resuming nuclear testing is that it focuses on the near-term goal of ensuring the efficacy of deterrence while disregarding the aspirational far-term vision of a world without nuclear weapons. We believe, in general, that the emphasis *does* need to be on the near term, if only to increase the probability that we survive to have the luxury of contemplating the far term. But, in fact, resumption of nuclear testing is simply *not* necessary to ensure a "safe, secure and effective arsenal" under current circumstances. Thus, in this case we are in favor of

---

[147] Boyd and Scouras, "Escape from Nuclear Deterrence."

[148] Commission on Maintaining United States Nuclear Weapons Expertise, *Report*.

not undermining our long-term goal for the sake of a relatively minor and unnecessary contribution to our near-term goal. By contrast, hedging against the many possible future events that would weigh heavily in changing this decision seems prudent and may even in some cases lessen the likelihood of their occurrence. Don't tickle the sleeping dragon's tail unless absolutely necessary.

Bibliography

Arms Control Association. "The Nuclear Testing Tally." Fact sheet. Last reviewed July 2020. https://www.armscontrol.org/factsheets/nucleartesttally.

———. "Senate Rejects Comprehensive Test Ban Treaty; Clinton Vows to Continue Moratorium." *Arms Control Today*, September 1999. https://www.armscontrol.org/act/1999-09/press-releases/senate-rejects-comprehensive-test-ban-treaty-clinton-vows-continue.

Ashley, Robert P. "Russian and Chinese Nuclear Modernization Trends." Remarks at the Hudson Institute, Washington, DC, May 29, 2019. https://www.dia.mil/News/Speeches-and-Testimonies/Article-View/Article/1859890/russian-and-chinese-nuclear-modernization-tren.

Atomic Energy Act of 1946, 42 U.S.C., Title 42, Chap. 23 (2010). https://www.govinfo.gov/content/pkg/USCODE-2010-title42/html/USCODE-2010-title42-chap23-divsnA.htm.

AVC (Bureau of Arms Control, Verification, and Compliance). *Adherence to and Compliance with Arms Control, Nonproliferation, and Disarmament Agreements and Commitments*. Washington, DC: US Department of State, June 2020. https://www.state.gov/wp-content/uploads/2020/06/2020-Adherence-to-and-Compliance-with-Arms-Control-Nonproliferation-and-Disarmament-Agreements-and-Commitments-Compliance-Report-1.pdf.

———. "Annual Assessment of the U.S. Nuclear Weapons Stockpile Share." Fact Sheet. Washington, DC: US Department of State, 2012. https://2009-2017.state.gov/t/avc/rls/202013.htm.

———. "Scope of the Comprehensive Nuclear Test-Ban Treaty." Fact Sheet. Washington, DC: US Department of State, 2013. https://2009-2017.state.gov/t/avc/rls/212166.htm.

Binninger, Gilbert C., Paul J. Castleberry, and Patsy M. McGrady. *Mathematical Background and Programming Aids for the Physical Vulnerability System for Nuclear Weapons*. Washington, DC: Defense Intelligence Agency, November 1, 1974. https://apps.dtic.mil/sti/pdfs/ADB010375.pdf.

Blank, Stephen J., Ed. *The Russian Military in Contemporary Perspective*. Carlisle, PA: Strategic Studies Institute, US Army War College, 2019.

Boyd, Dallas, and James Scouras. "Escape from Nuclear Deterrence: Lessons for Global Zero from the Strategic Defense Initiative." *Nonproliferation Review* 20, no. 2 (2013): 339–360. https://doi.org/10.1080/10736700.2013.799822.

Bradley, Curtis A. "Unratified Treaties, Domestic Politics, and the U.S. Constitution." *Harvard International Law Journal* 48, no. 2 (2007): 307–336. https://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=2517&context=faculty_scholarship.

Brecher, Michael, Jonathan Wilkenfeld, Kyle Beardsley, Patrick James, and David Quinn. International Crisis Behavior Data Set, Version 13. 2020. https://sites.duke.edu/icbdata/.

Bush, George H. W. Letter to the US Senate. Congressional Record—Senate, September 25, 1990, 25655. https://www.govinfo.gov/content/pkg/GPO-CRECB-1990-pt18/pdf/GPO-CRECB-1990-pt18-2-2.pdf.

Caldwell, Tod. Letter to the NNSA. November 23, 2019. https://www.lasg.org/MPF2/documents/JASONs-pit-aging_23Nov2019.pdf.

CIA (Central Intelligence Agency), Office of Transnational Issues. "Evidence of Russian Development of New Subkiloton Nuclear Warheads." Intelligence Memorandum. August 30, 2000. https://www.cia.gov/library/readingroom/docs/DOC_0001260463.pdf.

Cochran, Thomas B., and Christopher E. Paine. *The Role of Hydronuclear Tests and Other Low-Yield Nuclear Explosions and Their Status under a Comprehensive Test Ban*. Washington, DC: Natural Resources Defense Council, Inc., April 1995. https://fas.org/nuke/cochran/nuc_04009501a_143.pdf.

Commission on Maintaining United States Nuclear Weapons Expertise. *Report of the Commission on Maintaining United States Nuclear Weapons Expertise: Report to the Congress and Secretary of Energy*. Washington, DC: Department of Energy, March 1, 1999. http://www.breckenridgeinstitute.com/1999-CHILES-COMMISSION-REPORT.pdf.

Conrad, Edward E., Gerald A. Gurtman, Glen Kweder, Myron J. Mandell, and Willard W. White. *Collateral Damage to Satellites from EMP Attack*. Technical Report DTRA-IR-10-22. Ft. Belvoir, VA: Defense Threat Reduction Agency, August 2010. https://apps.dtic.mil/dtic/tr/fulltext/u2/a531197.pdf.

CTBTO (Comprehensive Nuclear-Test-Ban Treaty Organization Preparatory Commission). "18 December 1970 – The Baneberry Incident." https://www.ctbto.org/specials/testing-times/18-december-1970-the-baneberry-incident/.

———. "Fifteenth Anniversary of France's Last Nuclear Test." https://www.ctbto.org/press-centre/news-stories/2011/fifteenth-anniversaryof-frances-last-nuclear-test/.

———. "16 October 1964 – First Chinese Nuclear Test." https://www.ctbto.org/specials/testing-times/16-october-1964-first-chinese-nuclear-test.

Drell, S., C. Callan, M. Cornwall, D. Eardley, J. Goodman, D. Hanuner, W. Happer, J. Kimble, S. Koonin, R. LeLeviet, C. Max, W. Panofsky, M. Rosenbluth, J. Sullivan, P. Wetnberger, H. Vonc, and F. Zachariasen. *Science Based Stockpile Stewardship*. JSR 94-345. McLean, VA: JASON Program Office, MITRE, November 1994. https://fas.org/irp/agency/dod/jason/sbss.pdf.

Drollinger, Harold, Robert C. Jones, and Thomas F. Bullard. *A Historical Evaluation of the U12t Tunnel, Nevada Test Site, Nye County, Nevada*. Vol. 5. DOE/NV/26383–109. Las Vegas: NNSA, February 2009. https://doi.org/10.2172/1010604.

DTRA (Defense Threat Reduction Agency). "Nuclear Test Personnel Review (NTPR)." https://www.dtra.mil/DTRA-Mission/Reference-Documents/NTPR-Info/NTPR-Fact-/.

Forstenzer, Martin. "Concerns Arise over Aquifer Near Nuclear Test Site." *New York Times*, March 21, 2000. https://www.nytimes.com/2000/03/21/science/concerns-arise-over-aquifer-near-nuclear-test-site.html.

Frankel, Michael, James Scouras, and George Ullrich. *Nonstrategic Nuclear Weapons at an Inflection Point*. National Security Perspective NSAD-R-17-024. Laurel, MD: Johns Hopkins University Applied Physics Laboratory, 2017, https://www.jhuapl.edu/Content/documents/NonstrategicNuclearWeapons.pdf.

———. *The Uncertain Consequences of Nuclear Weapon Use*. National Security Report NSAD-R-15-020. Laurel, MD: Johns Hopkins University Applied Physics Laboratory, 2015. https://www.jhuapl.edu/Content/documents/UncertainConsequencesofNuclearWeapons.pdf.

Fry, J. G., R. A. Stane, and W. H. Crutchfield Jr. "Preliminary Design Studies in a Nuclear Excavation—Project Carryall." *Highway Research Record* no. 50 (1964): 32–39. http://onlinepubs.trb.org/Onlinepubs/hrr/1964/50/50-003.pdf.

Garbe, Paul, Charles W. Miller, Joan Morrissey, Claudia Parvanta, Judy Qualters, J. Felix Rogers, et al. *Report on the Feasibility of a Study of the Health Consequences to the American Population from Nuclear Weapons Tests Conducted by the United States and Other Nations*. Washington, DC: Department of Health and Human Services, Centers for Disease Control and Prevention and the National Cancer Institute, May 2005. https://www.cdc.gov/nceh/radiation/fallout/default.htm.

Glanz, James. "Testing the Aging Stockpile in a Test Ban Era." *New York Times*, November 28, 2000. https://www.nytimes.com/2000/11/28/science/testing-the-aging-stockpile-in-a-test-ban-era.html.

Hemley, R. J., D. Meiron, L. Bildsten, J. Cornwall, F. Dyson, S. Drell, D. Eardley, D. Hammer, R. Jeanloz, J. Katz, M. Ruderman, R. Schwitters, and J. Sullivan. *Pit Lifetime*. JASON Report no. JSR-06-335. McLean, VA: MITRE, January 11, 2007. https://fas.org/irp/agency/dod/jason/pit.pdf.

Hopkins, John C. "Nuclear Test Readiness." *National Security Science*, December 2016, 9–16. https://www.lanl.gov/discover/publications/national-security-science/2016-december/_assets/docs/NSS-dec2016_nuclear-test-readiness.pdf.

Hopkins, John C., and David H. Sharp. "The Scientific Foundation for Assessing the Nuclear Performance of Weapons in the U.S. Stockpile is Eroding." *Issues in Science and Technology* 35, no. 2 (2019): 23–25. https://issues.org/the-scientific-foundation/.

Hudson, John, and Paul Sonne. "Trump Administration Discussed Conducting First U.S. Nuclear Test in Decades." *Washington Post*, May 22, 2020, https://www.washingtonpost.com/national-security/trump-administration-discussed-conducting-first-us-nuclear-test-in-decades/2020/05/22/a805c904-9c5b-11ea-b60c-3be060a4f8e1_story.html.

IPPNW (International Physicians for the Prevention of Nuclear War and the Institute for Energy and Environmental Research) International Commission to Investigate the Health and Environmental Effects of Nuclear Weapons Production and IEER (Institute for Energy and Environmental Research). *Heaven and Earth: The Health and Environmental Effects of Nuclear Weapons Testing in, on, and above the Earth*. New York: Apex Press, 1991. https://ieer.org/wp/wp-content/uploads/1991/06/RadioactiveHeavenEarth1991.pdf.

Kastetter, Anna. "The Destabilizing Implications of Deploying New Low-Yield Nuclear Weapons." *Georgetown Security Studies Review*, April 23, 2020. https://georgetownsecuritystudiesreview.org/2020/04/23/the-destabilizing-implications-of-deploying-new-low-yield-nuclear-weapons/.

Kennedy, John F. Letter to Senate Leaders Restating the Administration's Views on the Nuclear Test Ban Treaty, September 11, 1963. https://www.presidency.ucsb.edu/documents/letter-senate-leaders-restating-the-administrations-views-the-nuclear-test-ban-treaty.

Kennedy, W R. Jr. *Fallout Forecasting—1945 through 1962*. LA-1605-MS, UC-11. Los Alamos, NM: Los Alamos National Laboratory, 1986. https://doi.org/10.2172/5722009.

Kidder, R. E. *Maintaining the U.S. Stockpile of Nuclear Weapons During a Low-Threshold of Comprehensive Test Ban*. UCRL-53820. Livermore, CA: Lawrence Livermore National Laboratory, October 1987. https://fas.org/programs/ssp/nukes/testing/kidderucrl53820.pdf.

———. *Report to Congress: Assessment of the Safety of US Nuclear Weapons and Related Test Requirements*. UCRL-LR-107454. Livermore, CA: Lawrence Livermore National Laboratory, July 26, 1991. https://fas.org/programs/ssp/nukes/testing/kidderucrllr107454.pdf.

Konopinski, E., C. Marvin, and E. Teller. *Ignition of the Atmosphere with Nuclear Bombs*. LA-602. Los Alamos, NM: Los Alamos National Laboratory, 1946.

Kramer, David. "Concerns about Aging Plutonium Drive Need for New Weapon Cores." *Physics Today* 71, no. 7 (2018): 22–24. https://physicstoday.scitation.org/doi/10.1063/PT.3.3969.

Kunkle, Thomas, and Byron Ristvet. *Castle Bravo: Fifty Years of Legend and Lore, A Guide to Off-Site Radiation Exposures*. DTRIAC SR-12-001. Washington, DC: Defense Threat Reduction Agency, January 2013. https://blog.nuclearsecrecy.com/wp-content/uploads/2013/06/SR-12-001-CASTLE-BRAVO.pdf.

Medalia, Jonathan. *Comprehensive Nuclear-Test-Ban Treaty: Updated "Safeguards" and Net Assessments*. CRS Report R40612. Washington, DC: Congressional Research Service, June 3, 2009. https://fas.org/sgp/crs/nuke/R40612.pdf.

Miller, George H., Paul S. Brown, and Carol T. Alonso. *Report to Congress on Stockpile Reliability, Weapon Remanufacture, and Role of Nuclear Testing*. UCRL-53822. Livermore, CA: Lawrence Livermore National Laboratory, 1987. https://www.osti.gov/servlets/purl/6032983.

Monroe, Robert R. "America's Stockpile of Nuclear Weapons May Not Work." *Washington Times*, November 11, 2020. https://www.washingtontimes.com/news/2020/nov/11/americas-stockpile-of-nuclear-weapons-may-not-work/.

Murphy, John R., and Brian W. Barker. *A Comparative Analysis of the Seismic Characteristics of Cavity Decoupled Nuclear and Chemical Explosions*. PL-TR-95-2177/SSS-TR-95-14980. Hanscom Air Force Base, MA: US Air Force Phillips Laboratory, 1995. https://apps.dtic.mil/sti/pdfs/ADA304812.pdf.

———. "Plutonium Pit Production." https://www.energy.gov/nnsa/plutonium-pit-production.

NNSA (National Nuclear Security Administration) Nevada Field Office. *Nevada Test Site Guide*. DOE/NV-715 Rev 1. Las Vegas: NNSA, March 2005. https://www.nnss.gov/docs/docs_LibraryPublications/DOENV_715_Rev1.pdf.

———. *United States Nuclear Tests July 1945 through September 1992*. DOE/NV-REV 16. Washington, DC: US Department of Energy, September 2015. https://www.nnss.gov/docs/docs_LibraryPublications/DOE_NV-209_Rev16.pdf.

Nordyke, Milo D. *The Soviet Program for Peaceful Uses of Nuclear Explosions*. UCRL-ID-124410. Livermore, CA: Lawrence Livermore National Laboratory, July 24, 1996. https://inis.iaea.org/collection/NCLCollectionStore/_Public/28/038/28038223.pdf.

NPR. "For 50 Years, Nuclear Bomb Lost in Watery Grave." *Weekend Edition Sunday*, February 3, 2008. https://www.npr.org/templates/story/story.php?storyId=18587608.

NRC (National Research Council). *The Comprehensive Nuclear Test Ban Treaty: Technical Issues for the United States*. Washington, DC: National Academies Press, 2012. https://doi.org/10.17226/12849.

Nuclear Weapons: Status of DOE's Nuclear Stockpile Surveillance Program, Hearing Before the US Senate Subcommittee on Strategic Forces Committee on Armed Services, 104th Cong., 2nd sess. (1996).

OASD(NM) (Office of the Assistant Secretary of Defense for Nuclear Matters). *Nuclear Matters Handbook 2020*. Washington, DC: Department of Defense, 2020. https://fas.org/man/eprint/nmhb2020.pdf.

OSTI (Office of Science and Technical Information). *Executive Summary: Plowshare Program*. Washington, DC: Department of Energy, 1997. https://www.osti.gov/opennet/reports/plowshar.pdf.

OTA (Office of Technology Assessment). *The Containment of Underground Nuclear Explosions*. OTA-ISC-414. Washington, DC: US Congress, October 1989. http://www.princeton.edu/~ota/disk1/1989/8909_n.html.

Perrow, Charles. *Normal Accidents: Living with High-Risk Technologies*. Rev. ed. Princeton, NJ: Princeton University Press, 1999.

Quirk, William J. *What Can Proliferants Learn from Low-Yield Nuclear Testing?* Unclassified version of a paper provided to JASON. Livermore, CA: Lawrence Livermore National Laboratory, August, 29 1994.

Rademacher, Steven E. *Plutonium Exposures to Personnel Assigned to Johnston Atoll*. Kirtland Air Force Base, TX: Headquarters Air Force Safety Center, January 2016. https://apps.dtic.mil/sti/pdfs/AD1069451.pdf.

Reis, Victor H., Robert J. Hanrahan, and W. Kirk Levedahl. "The Big Science of Stockpile Stewardship." *Physics Today* 69, no. 8 (2016): 46–53. https://doi.org/10.1063/PT.3.3268.

Richards, Paul G., and Won-Young Kim. "Advances in Monitoring Nuclear Weapon Testing." *Scientific American*, March 1, 2009. https://www.scientificamerican.com/article/advances-in-monitoring-nuclear/.

Ristvet, Byron. "A Brief History of US Nuclear Weapons Testing during the Cold War: 1947–1992." Defense Threat Reduction Agency briefing, September 11, 2013.

Rogoff, Martin A. "The International Legal Obligations of Signatories to an Unratified Treaty." In *The Law of Treaties*, edited by Scott Davidson, chap. 6. London: Routledge, July 2017.

Ryan, Kevin. "Is 'Escalate to Deescalate' Part of Russia's Nuclear Toolbox?" *Russia Matters*, January 8, 2020. https://www.russiamatters.org/analysis/escalate-deescalate-part-russias-nuclear-toolbox.

Safety and Reliability of the U.S. Nuclear Deterrent: Hearing 105–267, Before the Senate Subcommittee on International Security, Proliferation, and Federal Services of the Committee on Governmental Affairs, 105 Cong., 1st sess. (1997). https://www.govinfo.gov/content/pkg/CHRG-105shrg44720/pdf/CHRG-105shrg44720.pdf.

Schlesinger, James R. "Clinton Defers a Necessity—Nuclear Testing." *Wall Street Journal*, July 12, 1993.

Scouras, James. “Nuclear War as a Global Catastrophic Risk.” *Journal of Benefit-Cost Analysis* 10, no. 2 (2019): 274–295. https://doi.org/10.1017/bca.2019.16.

Shelley, Mary F. Memo to Colonel Shattuck. August 7, 1969. Defense Atomic Support Agency, DNA1.941220.102. https://www.osti.gov/opennet/servlets/purl/16004710.pdf.

Shute, Nevil. *On the Beach*. London: William Heinemann, Ltd., 1957.

Thorn, Robert N., and Donald Westervelt. *Hydronuclear Experiments*. LA-10902-MS UC-2. Los Alamos, NM: Los Alamos National Laboratory, February 1987. https://www.osti.gov/servlets/purl/6646692/.

Toton, Edward, and James Scouras. *Trinity and Ivy Mike: The Risk of Human Extinction from the First Atomic and Thermonuclear Bomb Tests*. National Security Analysis Report. Laurel, MD: Johns Hopkins University Applied Physics Laboratory, forthcoming.

Treaty on the Non-Proliferation of Nuclear Weapons. March 5, 1970. 729 U.N.T.S. 161. https://www.un.org/disarmament/wmd/nuclear/npt/text.

US Department of Defense. *Nuclear Posture Review 2018*. Washington, DC: Office of the Secretary of Defense, February 2018. https://media.defense.gov/2018/Feb/02/2001872886/-1/-1/1/2018-NUCLEAR-POSTURE-REVIEW-FINAL-REPORT.PDF.

US Department of Energy, Nevada Operations Office. *Regional Groundwater Flow and Tritium Transport Modeling and Risk Assessment of the Underground Test Area, Nevada Test Site, Nevada*. DOE/NV-477. Washington, DC: Department of Energy, October 1997. https://www.osti.gov/servlets/purl/788792.

US Department of Energy and US Department of Defense. “Memorandum of Understanding between the Department of Energy and the Department of Defense for Planning and Support for Safeguard C and Conducting Nuclear Weapons Tests Outside North American Continental Limits.” September 1984.

White House. *Moratorium on Nuclear Testin*g. Presidential Decision Directive PDD/NSC-11. July 3, 1993. https://fas.org/irp/offdocs/pdd11.htm.

———. *U.S. Policy on Stockpile Stewardship under an Extended Moratorium and a Comprehensive Test Ban*. Presidential Decision Directive PDD/NSC-15. November 3, 1993. https://fas.org/irp/offdocs/pdd/pdd-15.pdf.

Woolf, Amy F., Paul K. Kerr, and Mary Beth D. Nikitin. *Arms Control and Non-Proliferation: A Catalog of Treaties and Agreements*. CRS Report RL33865. Washington, DC: Congressional Research Service, March 26, 2020. https://apps.dtic.mil/sti/pdfs/AD1099624.pdf.

Acknowledgments

We gratefully acknowledge the reviews of earlier drafts of this paper by Dennis Evans, Christine Fox, Richard Gullickson, Mitch Nikolich, G. Peter Nanos, Byron Ristvet, Matt Schaffer, Eugene Sevin, and Alan Shaffer. Of course, their contributions should not be construed as endorsements of our conclusions and recommendations, which remain solely the responsibility of the authors.

About the Authors

Michael Frankel, one of the nation's leading experts on the effects of nuclear weapons, is a consultant on national security matters. Formerly, he was a national security studies fellow at the Johns Hopkins University Applied Physics Laboratory where he focused on deterrence policy and physical consequences of nuclear use. As executive director of the Commission to Assess the Threat to the United States from Electromagnetic Pulse Attack, he led development of foreign nuclear threat projections and US infrastructure vulnerability assessments. In prior government service, he was associate director for advanced energetics and nuclear weapons in the Office of the Deputy Under Secretary of Defense (Science and Technology), chief scientist of the Defense Nuclear Agency's Nuclear Phenomenology Division, congressional fellow at the US Senate, and research physicist at the Naval Surface Weapons Center. Dr. Frankel earned his PhD in theoretical physics from New York University.

James Scouras is a senior scholar at the Johns Hopkins University Applied Physics Laboratory and the former chief scientist of the Defense Threat Reduction Agency's Advanced Systems and Concepts Office. His research focuses on nuclear strategy and global catastrophic risk analysis. Previously, he was program director for risk analysis at the Homeland Security Institute, held research positions at the Institute for Defense Analyses and the RAND Corporation, and lectured on nuclear policy in the University of Maryland's General Honors Program. Among his publications is the book *A New Nuclear Century: Strategic Stability and Arms Control* (Praeger, 2002), coauthored with Stephen Cimbala. Publication of his edited volume, *On Assessing the Risk of Nuclear War*, is anticipated in 2021. Dr. Scouras earned his PhD in physics from the University of Maryland.

George Ullrich is senior vice president for strategy development at Applied Research Associates and the former chief technology officer at Schafer Corporation. Previously, he was senior vice president at Science Applications International Corporation. As a former member of the federal Senior Executive Service, he has held several prestigious positions within the Department of Defense, including deputy director of the Defense Nuclear Agency and director for weapons systems in the Office of the Secretary of Defense, receiving the Secretary of Defense Distinguished Service Medal for innovative weapon initiatives. He currently serves as a special advisor to the US Strategic Command's Strategic Advisory Group and director on the boards for ARA Inc. and the Washington chapter of the National Defense Industrial Association, and he previously served as a member of the Air Force Scientific Advisory Board. Dr. Ullrich earned his PhD in theoretical physics from Drexel University.

APL

JOHNS HOPKINS
APPLIED PHYSICS LABORATORY