The Probability of Nuclear War: A Rigorous Quantitative Assessment of an Existential Risk

Israel Flores

Independent Researcher

El Paso, Texas

2026

Abstract

The dominant paradigm in nuclear policy—Mutual Assured Destruction—rests on the claim that rational deterrence renders nuclear war effectively impossible. This paper demonstrates that claim to be false on multiple independent grounds: game theory itself refutes it, empirical history falsifies it, and the epistemological framework invoked to avoid quantifying the risk is itself incoherent. In its place, we develop a rigorous quantitative framework—the temporal risk model $c(p, t, T) = 1 - (1 - p)^{t/T}$—grounded in the Fundamental Theorem of Epistemology and Probability (FTEP), which proves that rational agents cannot escape probabilistic commitment and that the expected value of an agent's uncertainty distribution is the probability for rational belief and action. We present three independent lines of empirical evidence: expert elicitation via the 2005 Lugar Survey (projecting approximately 97% probability of nuclear war within 100 years), systematic analysis of 15 documented historical close calls (yielding approximately 85%), and historical weapons-usage base rates (exceeding 90%). A formal Bayesian confidence analysis demonstrates that $P(c > 0.5) > 0.99$—that is, with greater than 99% confidence, nuclear war is more likely than not within a human lifetime. This finding is robust to extreme skepticism scenarios, survives the halving of all input probabilities, and is reinforced rather than weakened by every plausible research extension. We neutralize seven persistent fallacies that obstruct engagement with these estimates, including the frequentist fallacy and the variance-nonexistence-zero fallacy. A cost-benefit analysis reveals a benefit-cost ratio exceeding 5,000 to 1 in favor of disarmament, rising to effectively infinite when avoided arsenal modernization costs are included. We conclude with a five-step treaty framework designed to drive cumulative nuclear war probability toward zero through verified, proportional disarmament; multiple independent inspection bodies; structured whistleblower incentives; international safe zones; and economically enforced compliance. The house is on fire. This paper provides the calorimeter readings and the phone number.

Keywords: nuclear war probability, Mutual Assured Destruction, temporal risk model, expected value, Bayesian confidence, close calls, nuclear disarmament, existential risk, Fundamental Theorem of Epistemology and Probability

Contents

1. Introduction

1.1 The Problem

For more than half a century, nuclear weapons have posed an existential threat to human civilization. Yet rigorous quantitative risk assessment—of the kind that is mandatory for pharmaceuticals, nuclear reactors, commercial aircraft, and virtually every other hazardous technology—has never been applied to nuclear war. The doctrine of Mutual Assured Destruction has been treated as a sufficient guarantee of safety, and the probability of nuclear war has been implicitly assumed to be negligible. This assumption has never been formally tested against the historical record of close calls, the expert consensus of national security professionals, or the base rates of weapons use in other contexts. This paper demonstrates that when it is tested, it fails—and that the cost-benefit arithmetic of the status quo reveals a macroeconomic blind spot of staggering proportions. Disarmament is not a cost. It is the most profitable investment humanity can make, as Section 11.3 demonstrates.

1.2 Prior Work

Several researchers have addressed nuclear risk qualitatively or with limited quantification. Sagan (1993) documented organizational failures in nuclear command systems, arguing that "normal accidents" make nuclear war inevitable over sufficient time horizons. Schlosser (2013) revealed numerous near-accidents involving nuclear weapons, demonstrating systematic safety vulnerabilities. These analyses remain primarily historical and qualitative.

Quantitative attempts exist but are limited. Hellman (2008) estimated approximately 1% annual probability based on expert judgment, implying roughly 63% per century if compounded. Baum et al. (2013) surveyed nuclear risk experts, finding median estimates of approximately 1.2% annual probability. The 2005 Lugar Survey found average expert estimates of 16.4% within 5 years and 29.2% within 10 years—figures this paper subjects to formal projection for the first time. Ord (2020) estimated 3% per century, a significant outlier from every other assessment.

None of these studies developed a formal mathematical model with an explicit three-parameter structure, systematically analyzed all documented close calls with individual probability estimates, incorporated explicit terms for unknown events, provided rigorous Bayesian confidence intervals, or proved convergence across independent methodologies with a formal confidence result. This paper provides all five contributions.

1.3 The Argument in Brief

The probability that nuclear war will occur within a human lifetime is not a matter of speculation. It is a quantity that can be rigorously estimated, formally bounded, and subjected to the same standards of evidence that govern every other domain of risk assessment. This paper does exactly that—and finds that the probability is alarmingly, indefensibly high.

Every other domain of hazardous technology operates under a simple regulatory principle: the burden of proof rests on the proponent of the hazard to demonstrate safety with high confidence, not on the public to prove danger. A new drug, a new reactor design, a new aircraft—none of these are deployed on the assumption of safety; safety must be shown. Nuclear weapons are the conspicuous exception. In over eight decades of the nuclear age, no government, no defender of the status quo, has ever specified an acceptable probability threshold for nuclear war, justified that threshold, and demonstrated with high confidence that the actual risk falls below it. The closest anyone has come to meeting that burden is Mutual Assured Destruction—the doctrine that rational actors will never initiate a nuclear exchange because the assured retaliation would annihilate their own country. This paper completely disproves MAD, on both logical and empirical grounds, in Section 2. Game theory itself, properly understood, does not require that a rational actor value the survival of others, their own state, or humanity—a result that yields several logically sufficient scenarios in which a perfectly rational leader chooses nuclear war, detailed in full in Section 2.1. And the empirical record supplies its own refutation: fifteen documented close calls between 1949 and 2025, including a Cuban Missile Crisis that multiple independent analysts place at 30–40% escalation probability, demonstrate that competent, rationality-respecting decision-makers have repeatedly brought the world to the brink despite MAD's prediction that they should not. With MAD eliminated, the status quo is left having never met its own burden of proof.

This paper does not stop at clearing that bar. It inverts it. Where the defenders of the status quo have never shown that the probability of nuclear war falls below any stated threshold, this paper shows the opposite with rigor exceeding anything previously offered in either direction: that the probability exceeds 50%—itself already an extraordinary finding—with greater than 99% confidence. And even that understates the result, because the 50% mark is not the threshold that matters. As Section 11.3 establishes through formal cost-benefit analysis, given the astronomical expected cost of nuclear war against the comparatively trivial cost of the disarmament framework proposed in Section 11.2, the actual threshold for action is far below 50%—closer to a probability indistinguishable from zero. We do not merely clear that real threshold; we clear it by an enormous margin, because the actual estimated probability is not hovering near 50% at all. It sits in the 84–97% range, drawn from three independent lines of empirical evidence detailed in Sections 6 through 8.

The confidence behind that range does not rest on the empirical evidence alone. It is reinforced by an independent epistemological result, developed in Section 4: the Expected Value Identity Theorem (EVIT), the central result of the Fundamental Theorem of Epistemology and Probability (FTEP). The EVIT proves, by two independent arguments, that the expected value of an agent's uncertainty distribution over a probability parameter is the probability for rational belief and action—and as an immediate corollary, that claimed ignorance is not a blank check but a precise commitment to $P = 0.5$. This framework does two things for the paper. First, it forecloses the most common evasion in nuclear policy discourse—the appeal to "deep uncertainty" as a reason to avoid quantification—by showing that such an appeal is itself a probabilistic commitment, and an indefensible one given the evidence. Second, it supplies the formal justification for treating every parameter in this paper's empirical model—each historical close call's probability, the allowance for unknown incidents, the projected century-scale risk—as the rational credence the available evidence demands, not as an arbitrary guess open to dismissal on the grounds of subjectivity.

When the empirical evidence is assembled under that epistemological discipline, it converges with remarkable consistency. The 2005 Lugar Survey of 79 national security professionals—selected for expertise in nuclear weapons policy and arms control—yields an implied 100-year probability of approximately 97%, with internal consistency between the 5-year and 10-year estimates confirming that respondents held coherent mental models rather than providing random guesses. Historical analysis of the 15 documented close calls, using decision-tree estimates for each event's escalation probability and an explicit parameter for unmodeled incidents, yields approximately 85%. Historical weapons-usage base rates—the empirical observation that every major weapons system developed and deployed since 1800 has eventually been used bilaterally in armed conflict between two opposing forces—yield a prior of approximately 90%, which rises toward certainty after Bayesian updating with the close-call evidence. These three methods are genuinely independent: they draw on different data sources, employ different methodologies, and make different structural assumptions. Their convergence on the 84–97% range is not a coincidence. It is the signal.

Formal Bayesian confidence analysis in Section 9 translates this convergence into a precise statement: $P(c > 0.5) > 0.99$. This finding is robust under extreme skepticism scenarios—halving every input probability, adopting the most conservative parameter estimates, and assuming systematic expert overestimation all leave the century-scale probability well above 50%. The joint probability of all errors required to push the estimate below 50% is less than 10^{-14}.

Section 5 catalogs seven persistent fallacies that obstruct engagement with these findings, building on the demolition of the MAD fallacy already completed in Section 2. The frequentist fallacy—that the absence of nuclear war to date proves the probability is near zero—is refuted by the same logic that shows a Russian roulette survivor has not demonstrated the game is safe. The variance-nonexistence-zero fallacy—the slide from "the probability is uncertain" to "the probability is probably low"—is shown to be a mathematical error:

high variance means a spread distribution, not a distribution concentrated near zero. The waiting-for-more-information fallacy is shown to be a disguised form of inaction whose hidden probabilistic commitments make it indefensible when expected probability and confidence are both high. The blue whale fallacy—dismissing a conclusion because the supporting analysis is not perfectly precise—is refuted by noting that the gap between the estimated probability and any reasonable action threshold is so vast that no plausible inaccuracy in the analysis can close it.

The economic analysis of Section 11.3 places the policy failure in starkest relief. A standard cost-benefit analysis, using a conservative century-scale risk of 80% and a conservative casualty estimate of 1 billion deaths, reveals an expected economic loss of $6.48 quadrillion in discounted global GDP. The cost of the five-step treaty framework proposed in Section 11.2—verified, proportional disarmament; multiple independent international inspection bodies; structured whistleblower incentives of approximately $1 billion per verified violation; international safe zones for leadership accountability; and economically enforced compliance through collective sanctions—lies between $467 billion and $1.25 trillion in present value terms. The benefit-cost ratio exceeds 5,000 to 1, rising to effectively infinite when the multi-trillion-dollar avoided cost of arsenal modernization is included. This is not a close call. It is the largest misallocation of expected value in the history of human civilization.

The remainder of this paper proceeds as follows. Section 2 demolishes MAD on theoretical and empirical grounds. Section 3 develops the temporal risk model and its mathematical properties. Section 4 establishes the epistemological framework, proving the EVIT and its immediate corollaries. Section 5 neutralizes seven remaining fallacies, cross-referencing the MAD demolition of Section 2 rather than repeating it. Sections 6, 7, and 8 present the three independent lines of empirical evidence. Section 9 synthesizes those lines into a formal confidence result. Section 10 maps the research agenda, showing that every plausible extension increases the estimated probability. Section 11 presents the five-step treaty framework, the cost-benefit analysis, and the central message. The burden of proof has never been met by those defending the status quo. This paper meets the opposite burden in full, and provides the framework to act on it.

---

2. Why MAD Cannot Eliminate Nuclear Risk

Mutual Assured Destruction is not a historical curiosity. It is the intellectual foundation of the nuclear posture of every major atomic power, and it has been for more than half a century. The doctrine's appeal lies in its apparent logical simplicity: no rational actor would deliberately initiate a nuclear war, because the assured retaliation would annihilate their own country. So long as leaders remain rational, the probability of nuclear war is effectively zero, and any resources spent on reducing that probability further are wasted.

That reasoning is not merely flawed. It is catastrophically wrong, and it can be shown to be wrong on multiple independent grounds before a single equation appears. The MAD doctrine depends on game theory, and game theory depends on the assumption that all relevant players are rational actors. Both dependencies fail spectacularly—and each fails on several independently fatal fronts. Indeed, game theory itself, properly understood, proves that MAD is fallacious even under the idealized assumption of perfect rationality. And the empirical world, as we will see, has already supplied more than a dozen demonstrations that the doctrine does not describe reality.

This section will establish those points. By its end, it should be clear that MAD is not a safety guarantee. It is a psychological comfort blanket that has been wrapped around the nuclear enterprise for decades—and the blanket is on fire.

2.1 The Theoretical Bankruptcy of MAD

MAD proponents borrow a result from classical game theory: in a one-shot, perfect-information game with symmetrical payoffs, rational players will always avoid mutually destructive outcomes. From this they conclude that nuclear war is irrational and therefore will not occur. The argument is rendered invalid even by game theory itself. The specific preferences that maximize a utility function for a rational player in game theory are not objective. They are subjective. Therefore a rational player in game theory does not need to value other human's lives whatsoever. Hence, a rational player can be a literal psychopath.

Game Theory Itself Refutes MAD

Even if every leader on Earth were a perfectly rational agent in the strict game-theoretic sense (an assumption that we will later show is wrong), MAD would still not be a guarantee of safety. Game theory does not dictate what a player's utility function must contain. It only dictates that a rational player will choose actions that maximize expected utility given their personal preferences. Therefore a rational player doesn't necessarily have to be benevolent or altruistic—they can be a literal serial killer. Nowhere in the von Neumann–Morgenstern axioms is there a requirement that a player must value the welfare of their own citizens, or the survival of their own country, or even the non-extinction of humanity.

This is not a subtle theoretical point. It is the death of MAD as a deductive guarantee. MAD claims that nuclear war is not possible because all leaders are rational players, and therefore the probability of nuclear war is essentially zero. To disprove that claim, one does not need to show that nuclear war is likely (which we will do later), or even show that not all leaders are rational players (which we will also do later). One only needs to identify a single logically possible scenario—any scenario at all—in which a rational player could choose to initiate it. Once a scenario is identified, MAD is false. The four scenarios that follow each satisfy this condition (and then some).

Scenario 1: A terminally ill leader chooses nuclear suicide. A leader receives a diagnosis of an incurable disease with months to live. They face a painful decline, a legacy already stained by scandal, and political enemies who will dismantle their life's work the moment they are gone. If their utility function places no value on the survival of strangers (as with psychopaths), initiating nuclear war is simply a grandiose form of suicide with less physical suffering than a natural cancer. The leader dies on their own terms, takes his enemies with them, and ensures that their name is not tarnished for the rest of history (or echoes through history if there's any survivors).

Scenario 2: A leader chooses nuclear suicide over political ruin or disgrace. A leader is confronted with imminent, irreversible personal catastrophe. A coup is closing in, and the rebels have made clear they will execute the leader and their family. A foreign intelligence service possesses blackmail material whose release means life in prison. A corruption investigation has reached the inner circle, and the evidence is irrefutable. In each case, the leader's alternative to nuclear war is not a long, peaceful retirement. It is imprisonment, humiliation, or even a torturous death. If human extinction means little or nothing to them (as with many psychopaths), and nuclear war offers even a marginally preferable outcome—revenge, control, the satisfaction of destroying one's enemies—then initiating it is utility-maximizing (i.e. in accordance with being a rational player in game theory). No terminal illness is required to end the world—just politics as usual.

Scenario 3: The leader facing certain ruin survives in a nuclear bunker. The same precipitating catastrophes—coup, prosecution, humiliation, death—confront a leader who additionally has access to a secure underground facility capable of sustaining life for decades. The choice becomes: certain destruction above ground, or a personal future in the bunker while the world burns. For a purely self-interested utility function (as with many psychopaths), the calculation is straightforward. The leader survives; the rest of humanity does not. Again, this is not a breakdown of rationality as defined by game theory. It is in perfect accordance with being a rational player in game theory. Game theory is completely indifferent to whether a rational player believes in any sort of God(s) or not. A rational player in game theory can believe in heaven, hell, karma, or nothing at all.

Scenario 4: The luxurious bunker lure (no imminent death or political ruin required). A psychopathic leader—or a psychopathic oligarch who controls the leader—need not be facing any catastrophe at all. They may simply calculate that the destruction of civilization maximizes their personal utility because it creates a world in which they are the sole center of power, wealth, and attention. A luxurious bunker, stocked for decades and staffed with a carefully selected group of companions (like attractive models), is not a refuge from catastrophe. It is the goal. The catastrophe is the mechanism that eliminates rivals, dissolves legal constraints, and transforms the leader into the absolute ruler of the only remaining habitable world. Inviting attractive partners for a "tour" of the facility—coincidentally timed with a nuclear crisis—is one vivid version of the scenario. The underlying logic does not depend on the specific lure. It depends only on a utility function that prefers a tiny, controlled world to a large, uncontrolled one.

Sub-variations and unknown unknowns. These four scenarios are not an exhaustive catalog. They are a sketch of a vastly larger space of possible utility functions and triggering circumstances. A terminally ill leader can also employ the supermodel lure to live out their final months in luxury. A leader facing a coup can be a religious fanatic who believes they are fulfilling prophecy. A leader whose spouse is threatening to leave them can decide to remove the habitable world rather than face abandonment. A leader can be motivated by a doomsday cult, a misinterpreted intelligence report, a medication-induced psychosis, or a simple, bottomless hatred of their enemies. Each of the four categories branches into countless sub-variations, and the sub-variations can combine, nest, and permute with one another. The combinations are endless. And beyond the variations we can imagine lie the variations we cannot—the unknown unknowns that no bounded imagination can exhaust. History repeatedly shows that engineered systems fail in ways their designers never anticipated, and the system at the center of nuclear deterrence is a human being with launch authority.

And all of the countless possibilities involving psychopaths above are amplified by an uncomfortable empirical regularity: individuals with psychopathic traits are significantly overrepresented in positions of political and corporate leadership. What is considered "rational" in the corporate cutthroat world of the oligarchs (ultra wealthy donors) often entails not placing a high value on the well-being of even their own workers, much less on humanity as a whole. Moreover, people from the corporate world tend to think in short term quarterly reports, not in long term existential events. Human survival in the nuclear age depends on leadership that thinks in the long term and values the well being of others. Corporate leadership often tends to think in the short term while placing little value on others—a recipe for disaster. Ultra wealthy and corporate donors are the largest contributors to election campaigns. The selection mechanisms that oligarchs filter for high office—ruthlessness, charisma, grandiosity, a willingness to make decisions that harm others for personal gain—are precisely the traits that correlate with a utility function that places little weight on the welfare of strangers. MAD proponents fallaciously assume that all leaders with nuclear arsenals care deeply about the survival of humanity. The actual evidence proves that this is nowhere near a safe assumption, and that the people most likely to control nuclear arsenals are, as a class, less likely to satisfy that assumption than the general population.

Amplifying the disturbing scenarios above further, many of the oligarchs who install and protect these leaders have acquired personal nuclear bunkers. This severs the link between their personal survival and the survival of the societies they govern. A psychopathic oligarch who can retreat to a subterranean compound while the surface burns does not face the same deterrence calculus that MAD assumes. The doctrine treats "the state" as a unitary actor with a single utility function that includes the survival of its population. In reality, the people who steer the state may have already purchased a ticket out, and their decisions will reflect that.

Each of the aforementioned scenarios is, on its own, sufficient to disprove the claim that MAD makes nuclear war impossible. Together, they render the doctrine almost comically fallacious. The existence of countless such sub-variations, each with its own non-zero probability, does not merely refute MAD. It makes the doctrine ridiculous. It is a safety guarantee that depends on the assumption that no leader, in any nuclear-armed state, across the entire span of human history, will ever have a utility function that favors catastrophe—and that assumption is not merely unsupported. It is refuted directly by game theory and everything we know about human psychology, political power, and the selection pressures that determine who rises to leadership.

MAD is now officially toppled by the very logic of the game theoretic foundation that it pretended to stand on. However, MAD is incoherent on so many different levels that we can still take a sledgehammer and shatter the remaining pieces scattered on the ground. And we will.

Not All Leaders Are Rational or Competent

Even if it were magically true that by having all leaders be rational players, it would somehow make MAD correct in theory (which isn't true as we already proved), MAD would STILL be incorrect because in the real world, not all leaders are rational or even competent. To assert that all leaders are rational and competent in the real world is wildly at odds with observed reality.

Political leadership in nuclear-armed states is not selected by a dispassionate process that filters for strategic wisdom, emotional stability, and cognitive fitness. Leaders are often installed by oligarchs who prioritize personal loyalty over competence, and the resulting selection pressure favors individuals who will serve the interests of their patrons rather than the interests of their populations—or of humanity. Some leaders have exhibited clear signs of cognitive decline, including dementia, while retaining sole authority over nuclear launch. Some are simply not very intelligent. Some are impulsive, paranoid, or detached from reality. The claim that “all nuclear-armed leaders are rational actors” is not a neutral default assumption; it is an extraordinary assertion about human psychology that would require extraordinary evidence, and that evidence does not exist.

The Nixon administration’s “madman theory” provides a fitting historical coda. The theory was a deliberate strategic posture: convince adversaries that the American president was volatile and irrational enough to escalate unpredictably, thereby extracting concessions. The performance, however, was not always clearly distinguishable from reality. Nixon’s documented drinking, his late-night rants, and his own staff’s efforts to circumvent his orders during moments of crisis suggest a line between theater and genuine instability that was vanishingly thin. If the commander of the most powerful nuclear arsenal in history could not reliably be distinguished from a madman, the assumption that all nuclear-armed leaders will always be rational, altruistic, and perfectly informed is not merely optimistic. It is a bet against human nature itself.

MAD Assumes Perfect Information and Infallible Technology

Even if we falsely assumed that every nuclear-armed leader is not only rational but deeply altruistic, genuinely committed to avoiding nuclear war, and immune to the psychological vulnerabilities described above. MAD still fails, because it assumes that these benevolent leaders will receive accurate, timely, and unambiguous information about the world during a crisis. They will not.

The nuclear command-and-control environment is not a seminar room. It is a system of radars, satellites, and communication links that can fail—and has failed, repeatedly—in ways that mimic an incoming attack. When a screen lights up with what appears to be a full-scale missile launch, a leader has minutes, sometimes seconds, to decide whether to believe it. They cannot check the sensor. They cannot phone the adversary to ask whether it is a glitch. They must act on the information in front of them, and that information has been catastrophically wrong on multiple occasions.

- In 1983, a Soviet satellite warning system reported five inbound American missiles with the highest level of confidence. The protocol demanded immediate notification of the Soviet leadership, which would likely have triggered a retaliatory launch. The system was wrong. Stanislav Petrov, the duty officer, guessed correctly—but he was guessing.
- In 1995, a Norwegian scientific rocket was detected by Russian early-warning radar and misidentified as a submarine-launched ballistic missile whose trajectory matched a decapitation strike against Moscow. The nuclear briefcase was activated and placed in front of President Yeltsin.
- In 1979, a technician at NORAD inadvertently loaded a training tape into the opera-

tional computer, causing screens across the United States to display a full-scale Soviet missile attack. Retaliatory bombers were scrambled before the error was discovered.

None of these incidents required an irrational leader. They required only a sensor malfunction and a decision clock that expired before the truth could be known. MAD, as a formal game-theoretic argument, assumes the players can observe the state of the world without error. The real world does not cooperate.

2.2 The Empirical Refutation

If MAD were correct—if the probability of nuclear war were genuinely held near zero by the structure of deterrence—then that probability would be a constant. It would not fluctuate with geopolitics. It would not spike during crises. The Doomsday Clock would never move. Expert assessments would cluster tightly around a number indistinguishable from zero. And close calls—incidents in which the world was saved not by the system but by individual conscience, luck, or sheer accident—would not exist.

The historical record is the opposite.

What follows is not a curated selection of the most dramatic moments. It is a systematic listing of the documented incidents in which nuclear war was a live possibility—incidents that, by MAD's own logic, should not have occurred. The list spans seven decades, multiple continents, and every type of nuclear weapons state. In each case, the key decision-makers were, by any reasonable standard, attempting to avoid catastrophe. Their rationality did not prevent the probabilities from being alarmingly high.

1. **The 1961 Goldsboro B-52 Crash.** A B-52 carrying two Mark 39 thermonuclear bombs broke up over North Carolina. One bomb's arming sequence activated; a single low-voltage switch, later found to be defective, prevented detonation. The weapon was 3.8 megatons—roughly 250 times the yield of the Hiroshima bomb.

2. **The Cuban Missile Crisis (1962).** For thirteen days, the United States and the Soviet Union stood at the brink. A U-2 was shot down. A Soviet submarine, cut off from communications and believing war had already begun, prepared to launch a nuclear torpedo; only Vasili Arkhipov's refusal to consent prevented it. Multiple analyses place the crisis-wide probability of escalation at 30–40%.

3. **The 1966 Palomares Incident.** A B-52 collided with a refueling tanker over Spain, releasing four hydrogen bombs. Two detonated conventionally on impact, scattering plutonium over a wide area. Nuclear detonation did not occur, but the safety mechanisms were pushed to their limits.

4. **The 1968 Thule Crash.** A B-52 with four hydrogen bombs crashed into the ice off

Greenland. The high explosives in all four weapons detonated; the nuclear cores were recovered, but only after a massive contamination event.

5. **The 1973 Yom Kippur War.** The United States and Soviet Union engaged in a tense naval standoff. U.S. forces went to DEFCON 3. Soviet ships were ordered to break the blockade. Miscommunication and miscalculation could easily have escalated.

6. **The 1960 NORAD False Alarm.** Early-warning radars mistakenly interpreted the moon rising over Norway as a massive Soviet ICBM attack. Bomber crews began engine-start procedures before the error was identified.

7. **The 1979 NORAD False Alarm.** A technician inadvertently loaded a training tape into the operational computer, causing screens across the United States to display a full-scale Soviet missile attack. Bombers were scrambled; missile crews went on high alert. The error was discovered only after minutes of genuine terror.

8. **The Petrov Incident (1983).** A Soviet early-warning satellite reported five inbound American missiles with maximum confidence. Protocol demanded immediate notification of the leadership, which would likely have triggered a retaliatory launch. Stanislav Petrov decided it was a false alarm and violated procedure. His probability of being wrong—and of being responsible for a missed genuine attack—was non-zero. So was the probability that he would be ignored.

9. **Able Archer 83 (1983).** A NATO command-post exercise simulated nuclear release procedures with unprecedented realism. Soviet intelligence interpreted the exercise as possible cover for a genuine first strike. Soviet nuclear forces were placed on heightened alert. The world came closer to war than was understood at the time.

10. **The 1991 Soviet Coup.** During the attempted coup against Mikhail Gorbachev, the Soviet nuclear briefcase was briefly unaccounted for. Control over launch authority was ambiguous for a period of hours.

11. **The Norwegian Rocket Incident (1995).** A scientific rocket launched to study the aurora was detected by Russian radar and misidentified as a submarine-launched ballistic missile on a trajectory consistent with a decapitation strike. The Russian nuclear briefcase was activated; President Yeltsin had minutes to decide.

12. **The 1999 Kargil Conflict.** India and Pakistan, both nuclear-armed, engaged in intense conventional fighting. Both sides mobilized; the risk of escalation to a nuclear exchange was non-zero and explicitly contemplated by analysts at the time.

13. **The 2001–2002 India-Pakistan Standoff.** A terrorist attack on the Indian parliament brought the two countries to the brink of war. Troops massed on the border.

Nuclear signaling—including missile tests—occurred. The probability of a nuclear exchange was assessed by some as the highest since the Cuban Missile Crisis.

14. **The 2007 Bent Spear Incident.** A B-52 bomber was mistakenly loaded with six nuclear-armed cruise missiles and flown across the United States without anyone realizing the weapons were live. The incident went undetected for 36 hours.
15. **The 2014-Present Ukraine Crisis.** A nuclear-armed Russia has explicitly threatened the use of nuclear weapons in the context of a conventional war. The rhetoric has been more direct than at any time since the Cold War, and the probability of escalation has fluctuated in response to battlefield events.

This list is not exhaustive. It omits lower-profile incidents, ambiguous signals that were never declassified, and the unknown unknowns that by definition cannot be cataloged. But even this incomplete roster makes the point unanswerable: the world has repeatedly come to the brink of nuclear war despite the presence of rational actors on all sides. MAD predicted that such incidents would not occur. They occurred. The doctrine is empirically falsified.

2.3 The Fallacy Behind the Fallacy

Before moving forward, it is worth briefly foreshadowing a fallback position that often arises at this stage. "But nuclear war hasn't happened yet," the defender of MAD will say. "We have had close calls, yes, but we are still here. Doesn't that suggest the probability is low, or even zero?"

This is the frequentist fallacy, and it will be dismantled in full detail in Section 5.1. For now, a single analogy will suffice. A person who plays Russian roulette five times and survives has not demonstrated that the game is safe. The probability of death on each round was one in six, regardless of the outcome. The fact that the bullet has not yet landed in the chamber is evidence of luck, not evidence of a low probability. The same logic applies to the 15 incidents listed above. The fact that the worst has not yet occurred is not evidence that the risk is low. It is evidence that, so far, we have been fortunate. And as the catalog makes painfully clear, the margins of that fortune have sometimes been terrifyingly thin.

2.4 The Takeaway

MAD is not a proof that nuclear war cannot happen. It is an incoherent psychological coping mechanism masquerading as game theory. Its theoretical foundations—perfect rationality, universal altruism, perfect information, infallible technology, transparent utility functions, and leader selection mechanisms that filter for strategic wisdom—are individually unrealistic and collectively absurd. Its empirical predictions are falsified by a historical record that contains more than a dozen near-catastrophes, each of which represents a non-zero probability of nuclear war that the doctrine cannot explain. And the most common fallback argument—that the absence of nuclear war proves the probability is low—rests on a misunderstanding of probability so fundamental that it would be laughed out of any other domain of risk assessment.

Before moving on to the formal mathematics, it is worth noting that even without a single equation, the logic of non-zero probabilities intuitively suggests that nuclear war is not just possible, but highly likely over time. "Murphy's Law" is a heuristic widely respected in engineering: if something can go wrong, eventually it will. Nuclear weapons, early-warning radars, and command bunkers are engineered hardware. Launch protocols and human decision-making chains are engineered processes. The entire apparatus of deterrence is an engineered system with a human at its center—and as history repeatedly proves, such systems fail in ways their designers never anticipated. Murphy's Law is not a formal theorem, but it captures an intuition that every actuary and systems engineer knows: over a long enough time horizon, with enough trials, events with a non-zero baseline probability are practically guaranteed to occur. When we combine the theoretical risk of the psychopathic utility scenarios above with the empirical reality of the 15 historical near-misses—where sensor glitches, lost briefcases, and crashed bombers brought us inches from annihilation—the baseline probability of failure on any given day is manifestly greater than zero. Over decades, across multiple nuclear-armed states, facing both systemic hardware failures and the opaque utility functions of flawed leaders, the cumulative probability that at least one fatal scenario materializes approaches 100% (unless the structural risks are drastically reduced via policy changes).

The remainder of this paper is dedicated to calculating the probability of nuclear war in a rigorous fashion, and it will be shown that those calculations demand immediate and far-reaching action. The next section introduces a formal model for projecting cumulative risk over time. The one after that establishes the epistemological foundation that makes all subsequent projections coherent. Together, they will accomplish in a quantitative way what this section has demonstrated qualitatively: the house is on fire. The psychopath scenarios prove the fire is possible. The fifteen close calls prove it has nearly started, repeatedly, for decades. And we have been standing in the smoke, telling ourselves the alarm must be faulty.

---

3. The Temporal Risk Model—From the Dart Analogy

Before developing the full mathematical framework, we solve an analogous made-up problem to introduce the main equations we will subsequently use and to illustrate the core logic in an intuitive way.

3.1 The Camouflage Dart Game

Consider a reckless regime that plays the following game:

- A blindfolded man throws a dart at a wall.
- After the throw, he sees that the wall is painted with army camouflage. The black regions—which would trigger a nuclear war if hit—cover 5% of the total wall area.
- Probability per throw: 0.05 (5%).

3.2 Cumulative Probability Over a Fixed Period

Suppose the regime repeats this game 15 times at random moments over a 76-year period. Each throw is independent in the sense that the outcome of one throw does not change the fraction of black on the wall.

- Probability of avoiding black on all 15 throws: $(0.95)^{15} \approx 0.4633$.
- Therefore, the cumulative probability of at least one nuclear war over those 76 years is:
$$p = 1 - (0.95)^{15} \approx 1 - 0.4633 = 0.5367 \approx 53.7\%$$

Already we see that after 15 throws, the chance of triggering nuclear war exceeds 50%—it is more likely than not over a 76-year period.

3.3 Projecting to Any Future Time Horizon

If the regime continues this pattern indefinitely (15 throws every 76 years), what is the probability of at least one black hit within the next t years?

- Number of 76-year periods in t years is $t/76$.
- The probability of surviving all those periods without a black hit is $(1-p)^{t/76}$.
- Hence the cumulative probability is: $c(p, t, 76) = 1 - (1-p)^{t/76}$.

With $p = 0.5367$ (the baseline, ignoring unknown throws), we obtain:

- $c(0.5367, 76, 76) = p = 53.7\%$
- $c(0.5367, 100, 76) = 1 - (0.4633)^{100/76} \approx 1 - 0.363 = 63.7\%$

3.4 Relationship to the Continuous Hazard Model

The discrete compounding formula can be rewritten in continuous form. Let $\lambda = -\ln(1-p)/T$. Then:

$$c(p, t, T) = 1 - e^{-\lambda t}$$

This is the standard survival-hazard equation, where λ is the constant annual hazard rate. With $p = 0.5367$ and $T = 76$, $\lambda = -\ln(0.4633)/76 \approx 0.0101$ (about 1.01% per year). This equivalence shows that our discrete model is not ad-hoc; it aligns with the widely used continuous hazard framework.

3.5 Generalizing to the Three-Parameter Formula $c(p, t, T)$

The projection depends on three quantities: the cumulative risk p over a reference period T, and the target horizon t. We therefore write it as a function of three explicit parameters:

$$\boxed{c(p, t, T) = 1 - (1-p)^{t/T}}$$

This is the primary equation we will use throughout the paper. For the dart analogy baseline ($p = 0.5367$, $T = 76$), we can compute risk over several policy-relevant time horizons:

- 25 years (one generation): $c(0.5367, 25, 76) = 1 - (0.4633)^{25/76} \approx 1 - 0.776 = 22.4\%$
- 50 years (half a century): $c(0.5367, 50, 76) = 1 - (0.4633)^{50/76} \approx 1 - 0.603 = 39.7\%$
- 100 years (a human lifetime): $c(0.5367, 100, 76) \approx 63.7\%$ (as computed above)

Even these baseline numbers (ignoring unknown incidents) already exceed 50% over the 76-year period and reach 63.7% over a century. If the real probability of nuclear war were anywhere near these figures, rational society would demand immediate, drastic risk reduction—comparable to evacuating a building with a 50%+ chance of collapse in the next hour.

When we later replace the analogy with actual historical data (15+ real close calls, expert surveys, weapons usage base rates), the estimated 100-year probability will be substantially higher—around 84–97%. This makes the situation far more urgent. The remainder of the paper is dedicated to presenting that rigorous evidence and the mathematical framework that forces us to act on it.

3.6 The Problem of Unknown Incidents (Introducing P_a)

Our calculation so far assumed that we know about all 15 dart throws. But what if the regime also conducted additional, unreported throws—perhaps because they were embarrassed by near-misses or because the incidents were classified? The wall's black area is still 5%, and any additional throw adds more risk.

Let P_a represent the cumulative additional probability from these unknown or unreported dart incidents. The true cumulative probability over the 76 years would then be:

$$p_{\text{true}} = 1 - (1 - P_a) \cdot (0.95)^{15}$$

A naive approach might treat P_a as a fixed number, but classical probability theory struggles with such "unknown unknowns" because there is no frequency data and no well-defined distribution. How can we rationally assign a probability to something we have never observed and cannot even enumerate?

This is exactly the type of problem that the Expected Value Identity Theorem (presented in the next section) resolves. As we will prove, any bounded agent with uncertainty about a parameter necessarily has a probability distribution over that parameter, and the expected value of that distribution is the rational probability to use. This framework allows us to handle P_a rigorously—using expert judgment, declassification rates, and other indirect evidence—rather than merely guessing.

3.7 General Equation for p

The dart example is a specific case of a general principle. Suppose an event may be triggered by any of N independent incident types, each with probability P_i of occurring within a reference period T. The probability that at least one trigger occurs is

$$p = 1 - \prod_{i=1}^{N}(1 - P_i)$$

If, in addition, there is a residual contribution from unknown or unmodeled incidents (denoted P_a), the full cumulative probability becomes

$$p = 1 - (1 - P_a) \cdot \prod_{i=1}^{N}(1 - P_i)$$

The term $(1 - P_a)$ represents the probability that no unknown incident triggers the event. This decomposition is the fundamental formula for combining independent risk contributions, and it will be used throughout the paper to incorporate both identified close calls (P_i) and the allowance for unreported or unimagined incidents (P_a). The Expected Value Identity Theorem of later sections will provide the rigorous justification for treating P_a and each P_i

as expected values of their respective uncertainty distributions.

3.8 Mathematical Properties

The model $c(p, t, T) = 1 - (1 - p)^{t/T}$ exhibits the following properties, confirming it is not ad-hoc:

Property 1 (Boundary Conditions):

- $c(p, 0, T) = 0$ for all p, T (no risk over zero time).
- $c(p, T, T) = p$ for all p, T (risk over the reference period equals the reference probability).

Property 2 (Monotonicity):

- $c(p, t, T)$ increases monotonically with t (risk accumulates over time).
- $c(p, t, T)$ increases monotonically with p (higher baseline risk yields higher projected risk).

Property 3 (Temporal Symmetry): The model satisfies a symmetry relation. Given $c_1 = c(p, t_1, T)$ and $c_2 = c(p, t_2, T)$:

$$c_2 = 1 - (1 - c_1)^{t_2/t_1}$$

Derivation: From c_1 we have $1 - c_1 = (1 - p)^{t_1/T}$, so $1 - p = (1 - c_1)^{T/t_1}$. Substituting:

$$c_2 = 1 - [(1 - c_1)^{T/t_1}]^{t_2/T} = 1 - (1 - c_1)^{t_2/t_1}$$

Significance: To project from one time horizon to another we only need c_1, t_1, and t_2; the reference period T cancels out. This bidirectional relationship demonstrates the model's internal consistency and mathematical elegance.

---

4. Foundational Epistemology: The Expected Value Identity Theorem and the Logic of Ignorance

Before presenting the risk model, we establish a fundamental principle underlying all quantitative probability assessment under uncertainty. This section proves that rational agents cannot escape probabilistic commitment, that claimed ignorance is itself a precise probability assignment, and that the standard objections—Knightian uncertainty, parameterization paradoxes, and partition dependence—are not barriers to quantification but consequences of failing to specify what is actually being asked.

4.1 The Problem of Probability Under Uncertainty

A common objection to probability estimates is: "We're uncertain about the probabilities themselves, so estimates are unreliable." This appears reasonable but is epistemologically confused. When uncertain about probability P, we represent that uncertainty with a distribution $\pi(p)$ over possible values. The question becomes: What single probability value should we use for rational belief and decision-making?

4.2 Theorem 4.1 (Expected Value Identity Theorem)

Theorem 4.1 (Expected Value Identity Theorem)**.** *Let P be a probability about which we have uncertainty, represented by distribution $\pi(p)$ over possible values $p \in [0, 1]$. Then the expected value*

$$\mathbb{E}[P] = \int_0^1 p \cdot \pi(p)\, dp$$

is the probability P for purposes of rational belief, decision-making, and action.

We provide two independent proofs. The first derives the result directly from the probability calculus itself, requiring no decision-theoretic axioms. The second shows that any departure from $\mathbb{E}[P]$ leads to incoherence between an agent's beliefs and their choices.

Proof 1 (From the Law of Total Probability). By the Law of Total Probability, we may condition on the possible values of the probability P:

$$P(E) = \int_0^1 P(E \mid \text{probability is } p) \cdot \pi(p)\, dp$$

The term $P(E \mid \text{probability is } p)$ is, by the very meaning of the conditioning statement, the agent's credence for E under the hypothesis that their credence equals p. That credence is simply p. Hence

$$P(E) = \int_0^1 p \cdot \pi(p)\, dp = \mathbb{E}[P]$$

This establishes the identity directly from the standard rules of probability, without invoking preferences, utilities, or coherence requirements. It shows that $\mathbb{E}[P]$ is not an approximation or a pragmatic substitute; it is the marginal probability that the agent's own distribution over probabilities entails. □

Proof 2 (From Coherence of Rational Belief). Let an agent hold beliefs about an uncertain probability P, represented by distribution $\pi(p)$. Define $\mathbb{E}[P] = \int_0^1 p \cdot \pi(p)\, dp$. The agent must use some probability value P^* for decision-making. We show that $P^* = \mathbb{E}[P]$ is required by coherence.

Suppose $P^* \neq \mathbb{E}[P]$. Consider a decision with outcomes dependent on an event with probability P:

- Utility U_1 if the event occurs (probability P)
- Utility U_2 if the event does not occur (probability $1 - P$)

Expected utility using P^*: $\mathrm{EU}(P^*) = P^* \cdot U_1 + (1 - P^*) \cdot U_2$.

From the agent's own perspective (uncertain about P), the true expected utility is

$$\begin{aligned}
\mathrm{EU}_{\mathrm{true}} &= \int_0^1 \big[p \cdot U_1 + (1-p) \cdot U_2\big] \cdot \pi(p)\, dp \\
&= \left[\int_0^1 p \cdot \pi(p)\, dp\right] U_1 + \left[\int_0^1 (1-p) \cdot \pi(p)\, dp\right] U_2 \\
&= \mathbb{E}[P] \cdot U_1 + (1 - \mathbb{E}[P]) \cdot U_2 = \mathrm{EU}(\mathbb{E}[P])
\end{aligned}$$

Thus using any $P^* \neq \mathbb{E}[P]$ systematically misrepresents the expected utility that the agent's own beliefs imply. The agent's beliefs and actions would be internally inconsistent—an epistemic incoherence. Rationality therefore requires $P^* = \mathbb{E}[P]$. $\square$

Remark. This theorem establishes that $\mathbb{E}[P]$ is not merely "an estimate" or "best guess" about some unknown "true" probability. Rather, $\mathbb{E}[P]$ is the probability for rational belief. Given the agent's information state, there is no other probability value that supersedes or is "more true" than $\mathbb{E}[P]$.

The Expected Value Identity Theorem is Part II of the Fundamental Theorem of Epistemology and Probability (FTEP). The full FTEP paper provides several additional independent proofs (Dutch Book, von Neumann–Morgenstern representation, Bayesian squared-error minimization, and proper scoring rules) and establishes the complete epistemic architecture: Part 0 (the Socratean Theorem, which proves that no bounded agent can possess absolute certainty) and Part I (the universal existence of probability distributions for all rational agents). Readers seeking the full foundational framework are referred to that work; here we rely on the two proofs above, which are fully sufficient for our purposes.

4.3 The Logic of Complete Ignorance

The Expected Value Identity Theorem supplies everything needed to dissolve the most common objections to probabilistic reasoning. The first and most important is the claim that an agent can be "completely ignorant" about a probability without thereby committing to any specific value.

Corollary 4.2 (Claimed Ignorance Implies Specific Commitment)**.** *If an agent claims "complete ignorance" about a probability P over the interval $[0, 1]$, this claim logically entails the commitment $P = 0.5$.*

Proof. The claim "complete ignorance" means no value of $p \in [0, 1]$ is privileged over any other. In the discrete case of a binary state space $\{A, \neg A\}$ with no evidence favoring either, the Principle of Indifference forces $P(A) = P(\neg A) = 0.5$. Extending this to a continuous probability parameter, complete ignorance demands that this lack of privilege holds uniformly for all p. If $\pi(p_1) > \pi(p_2)$ for any two values, the agent would be revealing a hidden belief that some probabilities are more likely than others, contradicting the claimed ignorance. A symmetric but non-uniform distribution (e.g., a Gaussian centered at 0.5) would still concentrate belief near 0.5, which also constitutes knowledge. The only distribution consistent with pure unconstrained ignorance is the uniform density $\pi(p) = 1$ for all $p \in [0, 1]$.

Given $\pi(p) = 1$, we have

$$\mathbb{E}[P] = \int_0^1 p \cdot 1 \, dp = \left[\frac{p^2}{2}\right]_0^1 = 0.5$$

By Theorem 4.1, this expected value is the probability. Hence the agent who claims complete ignorance is mathematically committed to $P = 0.5$. □

Why the Uniform Distribution Is Uniquely Correct

A deeper symmetry argument reinforces this result. If an agent truly has no information about p, then for any possible value p, the agent has no more reason to believe the true probability is p than to believe it is $1 - p$. The two are perfectly symmetric. Conditional on the true value being one of these two possibilities, the agent's expected probability is simply $(p+(1-p))/2 = 0.5$. Since this holds for every possible p, and every pair $\{p, 1-p\}$ is equally represented in the interval $[0, 1]$, the overall expected value across the entire distribution must be 0.5. No decision-theoretic axioms are required—only the Law of Total Probability applied to the symmetry constraint that the ignorance claim itself imposes.

This symmetry argument has an immediate corollary: any parameterization other than the uniform distribution over p breaks the complement symmetry. A uniform distribution over p^2, for example, concentrates probability mass toward smaller values of p and away from larger values, meaning the agent implicitly believes that p is more likely to be below 0.5 than above it. But that is a substantive belief, not ignorance. The only continuous distribution on $[0, 1]$ that respects the complement symmetry for all possible pairs simultaneously is the uniform distribution over p. Any other choice encodes knowledge.

Addressing the Reparameterization and Bertrand Objections

A sophisticated objection asks why the uniform distribution over p is preferred over, say, a uniform distribution over $\log(p)$ or p^2, which would yield different expected values. The objection often invokes Bertrand's Paradox—the observation that a "random chord" on a circle can be generated by three different mechanisms, each yielding a different probability that the chord is longer than the inscribed equilateral triangle's side. Critics claim that because "randomness" appears to be ill-defined in continuous spaces, the uniform distribution over p is arbitrary.

This objection mistakes an underspecified question for a paradox. Bertrand's three answers are not contradictory; they are answers to three different questions, each corresponding to a different generative mechanism. The "paradox" arises only because the original problem statement—"choose a random chord"—fails to specify the mechanism, just as asking for the average weight of potatoes in "a house on this block" without specifying which house makes the question unanswerable. Once the constraint structure is specified, the Principle of Indifference operates correctly within that structure, and a unique answer follows. There is no paradox—only an initial ambiguity that specification resolves.

The application to nuclear risk is direct. The agent is being asked about the probability p itself, not about p^2 or $\log(p)$. In any decision-theoretic or betting framework, the natural parameter is the probability—bets are priced at p dollars, not at p^2 or $\log(p)$. The parameter is already specified by the elicitation context. Choosing a different parameterization would encode additional knowledge about the structure of the problem, which contradicts the claimed ignorance. An agent who insists on a different parameterization is no longer "completely ignorant"; they are asserting something about how uncertainty is organized.

For completeness, we note that even if an agent were genuinely uncertain about which parameterization is appropriate, the Expected Value Identity Theorem provides a recursive resolution: the agent would place a uniform distribution over the possible parameterizations (applying the Principle of Indifference at the meta-level), compute the expected probability under each, and then compute the overall expected value. The recursion eventually bottoms out in the agent's actual evidence. This is not required to dissolve the Bertrand objection—the objection was already dissolved by recognizing that the parameter is specified—but it demonstrates that the framework handles meta-level uncertainty with the same rigor as ground-level uncertainty.

Partition Dependence

A related objection observes that different partitionings of possibility space yield different probability assignments under the Principle of Indifference. For example, the partition {Life on Mars, No Life on Mars} yields $P(\text{No Life}) = 0.5$, while the finer partition {Mammalian Life, Non-Mammalian Life, No Life} yields $P(\text{No Life}) = 1/3$. Critics argue that the Principle of Indifference is therefore inconsistent.

The response is that different partitions encode different knowledge states. The binary partition encodes the knowledge that the relevant distinction is between life and non-life; the ternary partition encodes the knowledge that the mammalian distinction matters. These are fundamentally different epistemic situations, and they correctly produce different probabilities. The Principle of Indifference operates within a specified partition, and the specification of the partition is itself a knowledge claim about which distinctions are relevant. An agent who is uncertain about which partition correctly represents reality simply applies the Principle of Indifference recursively—placing a distribution over partitions and computing the expected probability.

As with Bertrand, the appearance of a paradox vanishes once the knowledge claims implicit in the question's framing are made explicit. The Principle of Indifference, far from being refuted by partition-dependence or Bertrand-type puzzles, is validated by them: they demonstrate that the principle yields unique, coherent answers precisely when the question is well-specified, and that apparent inconsistencies arise only when the specification is left ambiguous.

From Ignorance to Precision: The Binary Search Elicitation Algorithm

The result that complete ignorance commits the agent to $P = 0.5$ is not merely a theoretical curiosity. It is the starting point for a constructive, operational procedure that can extract an agent's implicit probability with arbitrary precision.

The method proceeds by comparing the uncertain event to a series of reference lotteries with known probabilities:

1. Begin with a reference lottery that has a known probability of 0.5 (e.g., drawing a red ball from an urn with 50% red balls).
2. Ask the agent: "Is nuclear war more likely than drawing a red ball from this urn?"
3. If yes, the agent's probability is above 0.5; if no, it is below. The interval is bisected.
4. Repeat with a new reference lottery at the midpoint of the remaining interval (e.g., 0.75 or 0.25), and continue until the agent is indifferent between the event and the lottery.

The indifference point is the agent's subjective probability. If the agent expresses uncertainty about where exactly they would be indifferent, that uncertainty is represented by a distribution $\pi(p)$ around that point, and the EVIT identifies $\mathbb{E}[P]$ as the rational probability to use. The procedure is operational, repeatable, and grounded in the same von Neumann–Morgenstern framework that underlies the EVIT. It demonstrates that even in situations that feel deeply uncertain, coherent probabilities can be elicited, and the resulting values are not arbitrary inventions—they are revealed properties of the agent's own judgment.

4.4 Immediate Implications

Corollary 4.3 (Elimination of Vague Uncertainty Objections)**.** *Objections to probability estimates based on "uncertainty" are invalid unless accompanied by specification of an alternative distribution $\pi'(p)$ with $\mathbb{E}'[P] \neq \mathbb{E}[P]$, and a justification for why $\pi'(p)$ is superior to the original distribution.*

Proof. By Theorem 4.1, uncertainty is already captured by the distribution $\pi(p)$, and $\mathbb{E}[P]$ is the probability. A mere assertion of "uncertainty" without specifying an alternative distribution with a different expected value is epistemically vacuous—it does not challenge the original estimate, because the original estimate already incorporates all uncertainty represented by $\pi(p)$. □

The material developed in this section completely eliminates the "deep uncertainty" exemption that is so frequently invoked in nuclear policy discussions. An agent who claims complete ignorance is committed to $P = 0.5$. An agent who claims uncertainty without specifying an alternative distribution has offered no challenge at all. An agent who invokes Bertrand or partition-dependence has failed to recognize that the elicitation framework already specifies the parameter and the relevant distinctions. The binary search procedure shows that the movement from ignorance to a precise credence is not a theoretical abstraction—it is an operationally straightforward process that can be performed by any agent willing to make comparative judgments. The only way to avoid these conclusions is to refuse to make any probabilistic claim at all—which, as the FTEP paper establishes, is itself a decision with probabilistic consequences.

4.5 Application to Nuclear Risk Assessment

When we estimate P_i (the probability that historical event i would have caused nuclear war), we are finding $\mathbb{E}[P_i]$ given the available historical evidence, participant testimony, and expert analysis. By Theorem 4.1, this expected value is the probability P_i. Similarly, when we estimate P_a (the cumulative probability from unknown events) or p (the cumulative probability over a reference period), we calculate expected values which are the probabilities.

Therefore, our model parameters are not arbitrary or "subjective" in any pejorative sense—

they are the probabilities that rational belief requires given the available information. Objections based on vague "uncertainty" must either specify alternative distributions with different expected values and justify why those distributions are better supported by evidence, or they fall silent under Corollary 4.2. And the policy maker who claims to have "no idea" about the probability of nuclear war is, whether they recognize it or not, committed to treating it as equally likely as not. If that commitment is unacceptable, the only coherent response is to engage with the evidence and derive a better distribution.

4.6 The Death of Knightian Uncertainty and a Look Ahead

Perhaps the most consequential fallacy in policy discussions is the appeal to "deep uncertainty" or "Knightian uncertainty" as a reason to avoid assigning probabilities and postpone action. This stance is directly contradicted by the Expected Value Identity Theorem.

The previous subsections established that any bounded agent who claims uncertainty about a probability parameter P possesses a distribution $\pi(p)$ over possible values. The claim of "complete ignorance" over the interval $[0, 1]$ is not a blank check; it mathematically entails $\pi(p) = 1$ and thus $P = \mathbb{E}[P] = 0.5$. A policy maker who asserts "we simply don't know" is, whether they realize it or not, committed to treating the event as equally likely as not. If that commitment is uncomfortable, the only coherent escape is to provide evidence that supports a non-uniform distribution—which is precisely the act of probabilistic risk assessment.

Moreover, the movement from complete ignorance to a more precise credence need not be arbitrary or mysterious. A simple, constructive procedure—the binary search elicitation algorithm—can extract an agent's implicit probability with as much precision as desired. The method proceeds by comparing the uncertain event to a series of reference lotteries with known probabilities:

1. Begin with a reference lottery that has a known probability of 0.5 (e.g., drawing a red ball from an urn with 50% red balls).
2. Ask the agent: "Is nuclear war more likely than drawing a red ball from this urn?"
3. If yes, the agent's probability is above 0.5; if no, it is below. The interval is bisected.
4. Repeat with a new reference lottery at the midpoint of the remaining interval (e.g., 0.75 or 0.25), and continue until the agent is indifferent between the event and the lottery.

The indifference point is the agent's subjective probability. If the agent expresses uncertainty about where exactly they would be indifferent, that uncertainty is represented by a distribution $\pi(p)$ around that point, and the EVIT identifies $\mathbb{E}[P]$ as the rational probability to use. The procedure is operational, repeatable, and has deep foundations in von

Neumann–Morgenstern decision theory. It demonstrates that even in situations that feel deeply uncertain, coherent probabilities can be elicited, and the resulting values are not arbitrary inventions—they are revealed properties of the agent's own judgment.

This eliminates the "deep uncertainty" exemption. If an agent is willing to make any comparison between the uncertain event and a reference lottery, their probability is revealed. If they refuse to make any comparison at all, they are effectively declining to engage in rational choice—an option that is itself a choice with probabilistic consequences, as Corollary 4.1 and the inaction result of Section 4 demonstrate.

The practical upshot is that "the uncertainty is too great to act" is a disguised form of inaction that implicitly assumes the expected utility of the status quo exceeds that of alternative actions. Given that our analysis—and multiple independent lines of evidence—place $\mathbb{E}[P]$ for nuclear war in the range of 84%–97% per century, the expected utility calculus overwhelmingly favors immediate risk reduction. The appeal to uncertainty does not justify delay; it only confirms that the agent has not yet performed the elicitation that their own beliefs demand.

This mechanism proves that we are never trapped in a false dichotomy between a blind guess and a perfect, low-variance dataset. Subjective information states can always be rigorously mapped to probability bounds, rendering the excuse of Knightian uncertainty obsolete and demanding that expected-value calculations be mathematically addressed rather than philosophically dismissed.

Any remaining appeal to Knightian uncertainty is therefore fallacious. Speaking of fallacies, the next section catalogues several additional errors that routinely obstruct serious engagement with nuclear risk estimates—errors that persist even after the logic of ignorance has been clarified.

---

5. Neutralizing Common Fallacies in Nuclear Risk Assessment

The epistemological framework established in Section 4 proves that every rational agent possesses a probability for any well-defined event, that the expected value of the agent's uncertainty distribution is that probability, and that claimed ignorance commits the claimant to $P = 0.5$. These results are not merely technical niceties; they directly dismantle a series of fallacies that routinely obstruct serious engagement with nuclear risk estimates. The most significant of these—the claim that Mutual Assured Destruction eliminates nuclear risk—has already been refuted at length in Section 2, and we will not repeat that demolition here. In this section we address the remaining fallacies, showing that each collapses under the weight of either probability theory or empirical observation—or both.

5.1 The Frequentist Fallacy: "Zero Wars Have Occurred, Therefore the Probability Is ≈ 0"

A naïve frequentist approach tallies observed outcomes and treats the sample frequency as the probability: because no nuclear war has occurred, the probability must be near zero. This reasoning is not simply inaccurate for the nuclear case; it is fundamentally inapplicable and dangerously misleading.

The fallacy is easily exposed by analogy. Consider a revolver with six chambers, one containing a live round. A person plays Russian roulette five times and survives every round. A strict frequentist would compute the observed death frequency as $0/5 = 0\%$ and conclude that the probability of death is approximately zero. Everyone immediately recognizes this as absurd: the probability of death on each round is $1/6$, determined by the revolver's physical mechanism, not by the history of survival. The survivor got lucky; the probability was never zero.

The same logic applies to nuclear close calls. We have documented over 15 incidents that could have escalated to nuclear war (Cuban Missile Crisis, Petrov incident, Norwegian rocket false alarm, and many others), yet nuclear war has not occurred. The frequentist frequency is $0/15 = 0\%$. But just as with the revolver, we are not forced to infer probabilities from outcome frequencies. We can analyze the mechanism: decision trees, escalation pathways, false-alarm protocols, and human judgment under crisis pressure. These mechanistic analyses yield non-zero probabilities P_i for each historical event. The fact that the worst outcome did not materialize reflects luck, not a probability of zero.

The dart analogy of Section 3 makes the same point visually. After five throws have missed the black regions, the frequentist says the observed hit rate is 0%. But we can measure the black area on the wall (25%) and compute the true per-throw probability directly. The frequentist estimate is not just less accurate; it is completely uninformative for the actual risk.

For one-off or rare events where the underlying causal structure can be examined, mechanistic (Bayesian) analysis supersedes frequentist estimation. Nuclear war falls squarely into this category: we cannot re-run history, but we can dissect the moments when the world came close. The frequentist fallacy, far from undermining our probability estimates, simply highlights that the wrong tool is being applied to the problem.

5.2 The "Conflicting Probabilities Prove Arbitrariness" Fallacy

A common reaction to the existence of differing probability estimates—for instance, our estimate of approximately 85% for nuclear war in 100 years versus the 3% offered by Ord (2020)—is to conclude that probability estimation is inherently arbitrary and that no estimate can be taken seriously. This reaction fails to recognize a fundamental feature of Bayesian probability: probabilities are information-relative.

Consider three people observing the same Texas Hold'em hand. Player A holds two aces: her probability that an ace appears on the river, conditioned on her information, is roughly 4.3% (2 aces remaining out of 46 unknown cards). Player B holds no aces: his probability, conditioned on his different information, is 8.7% (4 aces remaining out of 46). The dealer, who has already burned and looked at the river card, has probability either 0 or 1. All three probabilities are simultaneously correct given their respective information sets. It would be irrational for Player B to adopt Player A's 4.3%—B has different evidence and should use the probability that reflects that evidence.

Applied to nuclear risk: different analysts condition on different information and employ different modeling choices. Ord's estimate reflects his synthesis of expert judgment at the time of his writing. Our estimate incorporates a larger set of declassified close calls, a formal temporal model, and explicit Bayesian confidence bounds. The divergence between the estimates does not demonstrate that probability is meaningless; it demonstrates that information and methodology matter. The relevant questions are: whose information set is more comprehensive, and whose reasoning from that information is more rigorous? We contend that our analysis is more thorough on both counts, for reasons detailed throughout this paper. Crucially, even if a reader disagrees, the proper response is not to dismiss probability estimation altogether but to specify an alternative information set and justify an alternative distribution. Mere observation that estimates differ is not a counter-argument; it is an invitation to examine whose conditioning is superior.

5.3 The Variance-Nonexistence-Zero Fallacy

A particularly insidious fallacy—common far beyond nuclear risk assessment—proceeds through three distinct but mutually reinforcing cognitive steps:

1. *Variance confused with nonexistence.* An agent observes that their uncertainty about a probability P is large (the distribution $\pi(p)$ has high variance) and concludes that the probability itself "does not exist" or is "unknowable" in principle.

2. *Nonexistence confused with zero.* The agent then conflates the claim "no probability exists" with the claim that the probability is effectively zero. If one cannot assign a number, the reasoning goes, then the risk is negligible.

3. *Zero implies no action.* From the false premise that $P \approx 0$, the agent concludes that no costly risk-reduction measures are warranted.

This cascade is appealing because it moves from a genuine feeling of ignorance to a comforting policy conclusion without engaging the evidence. But it is invalid at every step.

The first step fails because variance is a property of a distribution, not evidence of its absence. The Expected Value Identity Theorem (Section 4.2) has already established that an agent's uncertainty about P is represented by a well-defined probability distribution $\pi(p)$. The spread of that distribution—its variance—measures how imprecise the agent's information is, not whether a probability exists. A wide distribution is still a distribution.

The second step commits a categorical error. "The probability is unknown or uncertain" is an epistemic statement about the agent's information state; "the probability is nearly zero" is a factual statement about the event's likelihood. The former describes properties of $\pi(p)$; the latter claims that $\pi(p)$ is concentrated near zero. These are fundamentally different claims, and nothing about high variance implies concentration at zero. Indeed, the opposite is often true: high variance simply means the agent's credence is spread broadly across the $[0, 1]$ interval, potentially assigning substantial mass to high probabilities.

The formal refutation follows directly from the definition of the expected value. Suppose an agent has a distribution $\pi(p)$ with support on $[a, b]$ where $b > 0$. Then $\mathbb{E}[P] \geq a > 0$. High variance does not force the expected value toward zero; it reflects uncertainty about where in the interval the probability lies. For example, a uniform distribution over $[0.3, 0.9]$ has high variance (0.03) but an expected value of 0.6. An agent who claims "high uncertainty therefore probability is probably low" is making a mathematical mistake unless they can justify why the distribution is actually skewed toward zero.

In the nuclear context, this fallacy arises frequently. An analyst may say: "The probability of nuclear war is highly uncertain—some put it at 5%, others at 95%. Therefore we really have no idea, and it could be essentially zero. So maybe we shouldn't spend trillions on risk reduction." But the very spread of expert estimates cited by the analyst implies a distribution with substantial mass at high values. The proper response is to formalize that distribution and compute its expected value—which, as our triangulation shows, is uncomfortably large. High uncertainty about exactly when a catastrophe will occur does not license treating the probability as negligible; it licenses working harder to quantify the risk and act on the expected value.

This fallacy is not restricted to nuclear policy. It appears whenever someone dismisses a frightening possibility with "well, nobody really knows, so it might be zero." The appropriate reply is: "Uncertainty means we have a distribution, and the expectation of that distribution may still demand action." In the next sections we apply this exact principle to the available data and show that the expected probability of nuclear war over a human lifetime is far above zero—and above any reasonable threshold for inaction.

5.4 The "Waiting for More Information" Fallacy

A common response to quantitative risk estimates is: "We don't know enough yet. We should gather more data and refine our models before taking costly action." This sounds like responsible caution. In reality, it is an active decision carrying hidden probabilistic commitments—and often a far larger expected cost than its advocates recognize.

The formal problem is captured by Corollary 9.5 of the FTEP paper (Incoherence of Epistemic Abstention). Choosing to delay a decision in order to gather information is itself an action, revealing an implicit expected-utility comparison: the agent believes the expected utility of waiting—factoring in the value of future information, the probability of events occurring during the delay, and the costs of delay itself—exceeds the expected utility of acting now. This is not a neutral stance; it is a prediction about the future wrapped in a claim of ignorance. The agent who says "we should wait for more data" is asserting a probability distribution over what the new data will show, a set of utilities for different states of the world, and a model of how long the delay will last. Almost never are these commitments made explicit.

The fallacy becomes indefensible when the current state of knowledge already supports both a high probability of catastrophe and high confidence in that assessment. Consider a coastal community warned that a massive tsunami is crossing the Pacific. Models indicate with over 99% confidence that the wave has passed the halfway point. The exact location, height, and arrival time are still uncertain. A resident objects: "We don't know exactly where the wave is. Let's wait for a more precise location on where the wave is before evacuating." Everyone recognizes this as absurd. The fact that the wave's precise coordinates are unknown does not change the overwhelming probability that a destructive event is approaching. Waiting merely accumulates additional risk of death.

As the subsequent sections of this paper will demonstrate through rigorous Bayesian analysis, the nuclear situation is analogous. The probability that nuclear war is more likely than not within a human lifetime exceeds 99% confidence. Expert surveys, historical close-call analysis, and weapons-usage base rates converge on a per-century probability of 84%–97%, with formal bounds showing $P(\text{cumulative probability} > 0.5) > 0.99$. We are already in the position of the coastal community: the wave has passed the halfway point, even if we cannot see its exact crest.

When probability and confidence are both high, the calculus flips decisively against delay. Even a conservatively low annual probability of 1% yields a roughly 5% chance of nuclear war during a five-year waiting period—expected fatalities in the tens of millions. The new information would need to be extraordinarily valuable to offset that expected cost, and the proponent of delay must demonstrate that value explicitly. In typical policy discussions, no such demonstration is ever attempted. The appeal to "more data" functions as a conversation-stopper, not a rational calculation.

The epistemological framework of Section 4 already provides methods for extracting actionable probabilities from current knowledge. The binary-search elicitation procedure (Section 4.3.4) converts expert judgment into a well-defined distribution $\pi(p)$ right now. If the objection is that the resulting distribution has too much variance, one can compute the expected value of perfect information and compare it against the expected cost of delay. But the burden of proof rests on the person advocating inaction. In the nuclear domain, where risks compound relentlessly and the stakes are existential, the default must be to act on the best available probability estimate, not to demand precision while the clock runs.

In summary, waiting for more information is not a costless posture—it is a costly action that must be justified by the numbers, not invoked as a rhetorical escape. In situations of high probability and high confidence, exactly the situation our empirical analysis will establish for nuclear war, the fallacy ceases to be an intellectual mistake and becomes a catastrophic failure of reasoning. One does not wait to see the tsunami from the shore before running.

5.5 The "Events Are Not Independent" Fallacy

A frequent objection to the compound probability formula $p = 1 - \prod_{i=1}^{N}(1 - P_i)$ is that nuclear close calls are not strictly independent events. Each crisis, the argument goes, alters the geopolitical landscape—perhaps inducing caution or normalizing brinkmanship—so the probability of a future crisis escalating to war cannot be independent of what came before. From this, critics conclude that multiplying probabilities is invalid and the resulting cumulative estimate is unsound.

This objection fails on two independent levels, the second of which renders the entire independence debate entirely irrelevant.

First layer: Causal dependence is not statistical dependence. The distinction is familiar from elementary probability. When rolling a die twice, the second roll's initial physical state causally depends on the first roll's final resting position—the die does not magically reset to some metaphysical initial configuration. Yet we correctly calculate the probability of snake-eyes on the second roll as 1/36, entirely independent of the first roll's outcome. The causal chain is real but provides no useful predictive information; the complex tumbling dynamics scramble any trace of the prior state beyond recovery. Nuclear crises, separated by years or decades and driven by different leaders, technologies, and triggering circumstances, exhibit the same effective scrambling. What matters for the independence assumption is not the absence of causal connection, but the absence of systematic predictive information that would make $P(\text{Event}_{i+1}$ triggers war $\mid$ Event_i occurred) materially different from $P(\text{Event}_{i+1}$ triggers war $\mid$ Event_i did not occur). The burden of proof lies on the objector to demonstrate such a signal, and the historical record—with close calls distributed across distinct geopolitical eras without obvious clustering—offers no evidence for one.

Second layer: Any dependence is already baked into the P_i. Even if we set aside the first argument, the objection collapses at a deeper level. It treats the P_i as static, context-free numbers drawn from an unchanging urn—as if we are naively using the same base probability for every close call regardless of what came before. But that is not how the P_i estimates actually work. Each P_i is a dynamic, event-specific probability that is conditioned on the full state of the world at that moment. That state already includes every prior close call, every policy change, and every geopolitical shift that has occurred.

An analogy makes the point vividly. Suppose a mixed-martial-arts champion wins a fight but suffers a shoulder injury that never heals properly. A bookmaker setting odds for his next fight does not simply reuse his pre-injury probability and then append a "dependence correction factor." The bookmaker conditions on the fighter's current state—the damaged shoulder, the lost mobility, the altered training—and sets a new probability that reflects all of it. If the bookmaker now gives him a 30% chance of winning, that 30% already has the injury baked in. The causal dependence between the first fight and the second is not ignored; it is fully expressed in the revised probability. No further adjustment is required.

The same logic applies directly to nuclear close calls. When we estimate P_i for the Petrov incident of 1983, we are estimating the probability of nuclear war given the Soviet early-warning technology of that era, the specific protocols in place, the heightened Cold War tensions that had accumulated over decades, and the institutional memory of crises such as the Cuban Missile Crisis. If the Cuban Missile Crisis had never occurred, the world of 1983 would have been different, and our P_i for Petrov would presumably be different too. But we are estimating probabilities in the actual world, not a counterfactual one. Every scrap of dependence—causal, statistical, geopolitical—is already absorbed into the inside-view estimate of Petrov's escalation risk. There is no residual dependence lurking outside the P_i values waiting to be accounted for. It is already in them.

This second argument does not merely make independence a defensible default; it makes the entire independence debate moot. The P_i variables are not static quantities like the faces of a die. They are dynamic, like the odds set by a Vegas bookmaker who updates them continuously as new information arrives. If a prior crisis demonstrably increased the probability of escalation in a later crisis, that increase would already be reflected in the later P_i. If the objector believes a particular P_i has been estimated incorrectly—that it fails to capture the way a prior event altered the risk landscape—then the proper response is to propose a revised P_i with a justification for why the new number better reflects the evidence.

What the objector cannot legitimately do is claim that the entire multiplicative framework is invalid because of an unspecified, unquantified "dependence" that the framework already accommodates by design.

In combination, these two arguments leave the independence objection with no ground to stand on. Either the causal connections between crises are too scrambled to provide predictive traction, or any predictive traction they do provide is already incorporated into the event-specific P_i estimates. In neither case is the product formula invalidated. The multiplication of probabilities remains a principled and fully defensible method for aggregating the risk contributions of individual close calls into a cumulative assessment.

5.6 The "Future Probabilities Will Change" Fallacy

A final objection holds that projecting nuclear risk from historical data is unreliable because the future may differ from the past. Geopolitical conditions shift, new technologies emerge, proliferation continues, and diplomatic breakthroughs remain possible. Since the probability of nuclear war might change over time, the argument goes, any current estimate is inherently untrustworthy and should not serve as a basis for action.

This objection misunderstands what probability means in a Bayesian framework—and it does so in a way that would paralyze every form of risk assessment ever conducted.

Probability is always conditional on current information. When we state that the per-century probability of nuclear war is approximately 84%–97%, we mean: given the information available now—including the historical record of close calls, expert surveys, weapons-usage patterns, and the current geopolitical environment—the rational credence is in that range. If the information changes, the probability changes. That is not a weakness; it is how probability works.

Two analogies make the point unanswerable.

The coin flip. Before a fair coin is tossed, the probability of heads is 1/2. After the toss, when the outcome is observed, the probability becomes either 1 or 0. The fact that the probability will change in the future does not make the pre-toss estimate of 1/2 incorrect, unreliable, or useless. The pre-toss probability was the correct credence given the information available at that time. The post-toss probability is the correct credence given the new information. Both are correct for their respective information states. No one would refuse to call a coin toss on the grounds that "the probability might change after we see the result."

The injured fighter. Before his first fight, an MMA champion is given a 90% probability of winning. He wins but suffers a shoulder injury that never heals properly. Bookmakers now set his probability of winning his next fight at 30%. The 30% does not retroactively invalidate the earlier 90%. The 90% was correct given what was known before the injury.

The 30% is correct given what is known now. If the fighter undergoes a miraculous recovery tomorrow, the odds will shift again—and that new number will be correct for that new information state. A spectator who objects that "the probability keeps changing, so the whole thing is unreliable" has simply failed to understand what a probability is.

The same logic applies directly to nuclear risk. If a breakthrough in verification technology enables verified disarmament, the probability of nuclear war will fall—and our framework is explicitly designed to accommodate that update. If a new nuclear power emerges with unstable command-and-control systems, the probability will rise—and again, the framework updates. The temporal model $c(p, t, T)$ can be recomputed whenever new information arrives: as the reference period T expands and new events are observed or avoided, the cumulative probability p is recalculated, and the projection updates accordingly. This is not a flaw; it is the model functioning exactly as intended.

The objector's demand—that we should not assign any probability now because the probability might change later—would, if taken seriously, destroy every quantitative risk assessment in every domain. Climate models would be deemed unreliable because future emissions trajectories remain uncertain. Pandemic risk assessments would be dismissed because surveillance technology might improve. Financial risk models would be abandoned because market conditions evolve. The alternative to "assigning the best probability given current information" is not "waiting for perfect information." It is paralysis dressed as prudence.

By the Expected Value Identity Theorem (Section 4.2), $\mathbb{E}[P \mid \text{current information}]$ is the probability now. Future information will yield $\mathbb{E}[P \mid \text{future information}]$, which may differ. Both are correct for their respective information states. There is no separate "true" probability floating independently of information, waiting to be discovered. The objector who claims that probability estimates are unreliable because they might change must either specify an alternative distribution with a different expected value (Corollary 4.2) or concede that they are simply expressing discomfort with the fact that knowledge evolves—a fact that applies to every scientific claim ever made.

5.7 The Blue Whale Fallacy: Dismissing Overwhelming Evidence with Minor Inaccuracies

Another possible failure in thinking can arise when someone dismisses a finding due to minor inaccuracies—I'll call it the blue whale fallacy. It simply points to a minor unaccounted detail—something the model omitted, a minor math error, some imperfection in the data—and claims that because the analysis is not flawless, its conclusions cannot be trusted. This objection confuses any imperfection with fatal imperfection, and it is decisively refuted by considering the scale of the conclusion relative to the threshold for action.

Imagine a biologist who sets out to prove that a fully grown blue whale is larger than a great white shark to a critic who's completely ignorant of marine animals. She presents a mathematical model with extensive evidence: direct measurements, comparative anatomy, photographic documentation. But the critic objects: "You didn't subtract the contents of the whale's stomach from its total weight." Or "You falsely assumed the whale's heart density is homogeneous." Or "You made an arithmetic mistake when adding its brain and skull." Although these pedantic observations may be true, it would be fallacious for the critic to then say "Your measurements are imprecise. Therefore your conclusion that a blue whale is larger than a great white shark is unreliable."

Even the most generous estimate of stomach contents cannot close the gap between a blue whale and a great white shark. The conclusion that the whale is larger than the great white shark is not sensitive to these pedantic details. Even after taking these petty details into account, the probability that a blue whale is larger than a shark is still overwhelmingly over 50%. Similarly, any pedantic inaccuracies that a critic may later find in our mathematical model do not nullify its ultimate conclusion—that the probability of nuclear war is overwhelmingly over 50% and therefore immediate action is required to lower the probability.

Moreover, the threshold for action—the point above which the risk becomes intolerable and entails policy changes—is not the 50% mark. It is far below 50%. If the true probability of 84–97% is the blue whale, and the 50% probability mark is the shark, then the true probability threshold for action is something like a guppy. As Section 11.3 will later demonstrate, given the astronomically large expected cost of nuclear war and the comparatively trivial cost of the five-step treaty, the probability threshold at which inaction could be justified is effectively indistinguishable from zero.

To demonstrate this fallacy further, let's consider a specific area of our analysis to which an ignorant critic might fall victim—the Bayesian confidence proof of upcoming Section 9. A critic might argue that the three lines of evidence—expert elicitation, historical close-call analysis, and weapons-usage base rates—are not fully independent, and therefore the combined confidence estimate is unreliable. The objection is analogous to the following scenario. Imagine a Major League Baseball all-star team is scheduled to play a Little League team three times in a row. An analyst concludes the MLB team has a higher than 50% probability of winning at least one game, and with a higher than 99% confidence level. He comes to this conclusion by estimating the actual probability using a model of the form $1-p^3$, where p is the probability of the Little League winning any given game. The analyst therefore treated the games as independent. But then an ignorant critic objects that the games are not technically independent: a close loss might boost the Little League team's confidence, or an MLB team's player might break a leg, affecting later games. The dependence effects are real. They are also orders of magnitude too small to change the conclusion. The probability that the MLB team wins at least one game remains overwhelmingly over 50% under any plausible dependence structure. The critic who dismisses the conclusion because the games aren't perfectly independent is committing the same error as the critic who dismisses the whale-guppy conclusion because of stomach contents. In the nuclear case, the three lines of evidence converge on the same range, and the combined confidence exceeds 99% even under conservative assumptions about correlation (Section 9.6). The dependence objection, even if granted, does not change the result.

The formal refutation of this fallacy is straightforward. Let C be the paper's central conclusion—that immediate action is required to reduce the probability of nuclear war—and let δ be the maximum conceivable effect of the unaccounted detail on the estimated probability. The objection is valid only if the detail could, in principle, reverse C. The burden of proof rests on the objector to demonstrate that δ is large enough to do so. Pointing to the existence of an imperfection without quantifying its effect is not an argument; it is a rhetorical gesture that exploits the fact that no model is perfect.

In combination with the extreme skepticism scenarios of Section 9, the response to this fallacy is decisive. Section 9 will later show that halving every P_i, adopting the most conservative P_a estimate, and assuming systematic expert overestimation still leaves the century-scale probability far above any reasonable action threshold. Any unaccounted detail that an objector can name is, with overwhelming probability, smaller in magnitude than the halving of every estimate. The whale is definitely larger than the shark (the 50% mark), and the whale is most definitely larger than the guppy (the threshold for action). The stomach contents are a distraction.

These seven fallacies account for the vast majority of objections raised against quantitative nuclear risk assessment. Each has been shown to rest on a misunderstanding of probability, a misapplication of decision theory, or an empirical claim contradicted by the historical record. Together with the positive epistemological framework of Section 4 and the MAD demolition of Section 2, they provide a solid foundation for the empirical analyses that follow.

---

6. Evidence I: Expert Elicitation (Lugar Survey)

We begin our empirical analysis with expert consensus, as this provides the most immediately credible evidence to skeptical readers. Moreover, a powerful statistical principle—the wisdom of crowds—provides strong reason to take these aggregated judgments seriously. We foreshadow that principle now and return to it in detail at the end of the section.

6.1 Survey Design

The 2005 Lugar Survey, commissioned by the U.S. Senate Foreign Relations Committee under Senator Richard Lugar, surveyed 79 national security professionals with expertise in nuclear weapons policy, arms control, and national security strategy (Lugar, 2005).

Two key questions:

1. "What is the probability of a nuclear attack somewhere in the world in the next 5 years?" Average response: 16.4%
2. "What is the probability of a nuclear attack somewhere in the world in the next 10 years?" Average response: 29.2%

6.2 Internal Consistency Check

If experts held coherent probability models with constant underlying risk rates, the 5-year and 10-year estimates should satisfy a consistency relationship.

From 5-year estimate: Using our model, if $p = 0.164$ over $T = 5$ years, the implied 10-year probability is:

$$c(0.164, 10, 5) = 1 - (1 - 0.164)^{10/5} = 1 - (0.836)^2 = 1 - 0.699 = 30.1\%$$

Actual 10-year estimate: 29.2%. Difference: 0.9 percentage points (less than 1%).

Interpretation: The experts demonstrated remarkable internal consistency. Their 10-year estimate nearly perfectly matches what their 5-year estimate implies under constant risk assumption. This suggests they were not providing random guesses but had coherent mental models of nuclear risk accumulation over time.

6.3 Projections to 100 Years

From 5-year estimate ($p = 0.164$, $T = 5$):

$$c(0.164, 100, 5) = 1 - (0.836)^{100/5} = 1 - (0.836)^{20} = 1 - 0.0278 = 97.2\%$$

From 10-year estimate ($p = 0.292$, $T = 10$):

$$c(0.292, 100, 10) = 1 - (0.708)^{100/10} = 1 - (0.708)^{10} = 1 - 0.0316 = 96.8\%$$

Average: $(97.2\% + 96.8\%)/2 = \mathbf{97.0}\%$

Expert consensus (2005): Approximately 97% probability of nuclear war or attack within 100 years.

Table 1: Lugar Survey Projections Using Model $c(p, t, T)$

Time Horizon	$\boldsymbol{c(0.164, t, 5)}$	$\boldsymbol{c(0.292, t, 10)}$	**Average**
5 years	16.4%	—	16.4%
10 years	30.1%*	29.2%	29.7%
25 years	59.2%	57.8%	58.5%
50 years	83.3%	82.2%	82.8%
100 years	97.2%	96.8%	**97.0%**

*Implied from 5-year estimate; actual survey response was 29.2% (0.9 p.p. difference, confirming consistency).

6.4 Robustness: Trimming High Responses

Potential objection: "Perhaps some experts were overly pessimistic, skewing the average upward."

Test: Remove highest responses and recalculate. For the 10-year question, 7 of 79 respondents (8.9%) answered between 90–100%.

Trimming analysis: Assume all 7 answered 100% for simplicity, and remove them:

$$\text{Adjusted average} = \frac{0.292 \times 79 - 1.00 \times 7}{79 - 7} = \frac{16.068}{72} = 0.223 \text{ (22.3\%)}$$

Recalculation with trimmed data:

$$c(0.223, 100, 10) = 1 - (0.777)^{10} = 1 - 0.0805 = 0.920 = 92.0\%$$

Result: Even after removing the most pessimistic ∼9% of experts, the 100-year probability remains 92%—still well above 50%.

Conclusion: Even aggressive trimming of "pessimistic" experts leaves probability well above 50%. The finding is robust to outlier concerns.

6.5 Context: Risk Has Increased Since 2005

The Lugar Survey was conducted in 2005. Has risk changed since then?

Doomsday Clock (Bulletin of Atomic Scientists):

- 2005: 7 minutes to midnight

- 2024–2025: 90 seconds to midnight (closest position in the Clock's history)

Factors increasing risk since 2005:

- North Korean nuclear advancement and ICBM development
- Deterioration of US-Russia arms control (INF Treaty collapse 2019, New START extensions uncertain)
- Russia-Ukraine war with explicit Russian nuclear rhetoric (2022–present)
- India-Pakistan tensions remain elevated
- Cyber vulnerabilities in nuclear command and control systems increasingly recognized
- Multipolar nuclear order (9 states) vs. bipolar Cold War

Implication: The 2005 Lugar Survey estimates likely represent a conservative lower bound on current risk. If anything, contemporary expert surveys would likely yield higher estimates given the deteriorated security environment. Using 2005 data provides a conservative baseline for our analysis. Current risk may be substantially higher.

6.6 The Wisdom of Crowds: Why Expert Aggregation Is a Feature, Not a Bug

A natural objection to any expert survey is that "experts are often wrong," and that averaging opinions does not produce knowledge. This objection misunderstands a well-established phenomenon: under the right conditions, the aggregated judgment of a diverse group of independent assessors is not merely "as good as" the typical expert—it is better, often dramatically so.

The classic demonstration comes from Francis Galton's 1906 observation at a country fair, where nearly 800 villagers guessed the weight of an ox. No individual guessed the true weight exactly, yet the average of all guesses was accurate to within a single pound. The errors canceled: those who overestimated and those who underestimated did so in roughly equal measure, leaving the collective judgment remarkably close to the truth (Galton, 1907).

The conditions for the wisdom of crowds to operate are:

1. *Diversity of judgment* – assessors bring different information, models, and perspectives.
2. *Independence* – assessors form their judgments without coordinating or influencing one another.
3. *Decentralization* – assessors draw on local knowledge, not a single shared source.

When these conditions are met, the aggregate estimate is often more accurate than any individual estimate, including those of the most credentialed experts (Surowiecki, 2005).

The mathematical reason is straightforward: individual errors, if they are uncorrelated and mean-zero, cancel in the average.

The Lugar Survey satisfies these conditions remarkably well. The 79 respondents were drawn from senior positions across government, military, intelligence, and academia. They had access to different classified and open-source information, different analytic frameworks, and different professional experiences. They responded independently to a confidential survey; there was no opportunity for coordination or groupthink. Each expert provided a private, considered judgment. The diversity of their backgrounds and the independence of their responses make the aggregated result a textbook candidate for the wisdom-of-crowds effect.

One might worry that experts in national security share certain biases—institutional or ideological—that could lead to correlated errors. Even if some correlation exists, however, the central tendency would still be pulled toward the group mean. The trimmed mean, which discards the most extreme responses, provides additional protection against outliers (Section 6.4). Furthermore, the near-perfect internal consistency between the 5-year and 10-year responses (Section 6.2) suggests that the experts were not merely guessing but applying coherent underlying models—another hallmark of meaningful judgment (Tetlock and Gardner, 2015).

Thus, the Lugar Survey is not merely an opinion poll; it is a structured information-aggregation mechanism whose output is supported by a rigorous statistical principle. The fact that its central estimates project to approximately 97% over a century means that the collected judgment of the people best informed about nuclear threats is that catastrophe is not a remote possibility but a near-inevitability over policy-relevant time horizons.

This finding stands on its own, but it does not stand alone. In the sections that follow, we will see that two entirely independent methodologies—historical close-call analysis and weapons-usage base rates—converge on the same uncomfortable conclusion. The Lugar Survey is the first leg of a triangulation that makes the cumulative case extraordinarily difficult to dismiss.

---

7. Evidence II: Historical Close-Calls Analysis

Having established expert consensus pointing to $\sim$97% century-probability, we now examine whether systematic analysis of historical close calls supports this assessment.

7.1 Methodology

We identified close-call events during 1949–2025 ($T = 76$ years) using criteria: (1) event involved nuclear-armed states or potential nuclear weapons use, (2) credible historical documentation exists (declassified documents, scholarly consensus, participant testimony), (3) event posed non-trivial risk ($P_i > 0.01$).

For each event, we estimate P_i using decision tree analysis, historical scholarship integration, and participant retrospective assessments. By Theorem 4.1, each $P_i = \mathbb{E}[P_i]$ given available evidence, which IS the probability.

Detailed event documentation with primary sources and P_i estimation methodology is provided in Table 2 below, with key citations given in the brief justification column.

7.2 Summary of Identified Events

Table 2 presents all 15 identified events or event categories with probability estimates and brief justifications.

7.3 Calculation of p

Using the general equation with $P_a = 0.20$ and the P_i values from Table 2:

Step 1: Calculate product of survival probabilities for known events:

$$\prod_{i=1}^{15}(1 - P_i) = (0.60)(0.85)(0.92)(0.93)(0.94)(0.95)(0.96)(0.97) \times (0.97)(0.975)(0.98)(0.98)(0.985)(0.985)(0.95) = 0.3038$$

Step 2: Incorporate unknown events parameter:

$$(1 - P_a) \times \prod(1 - P_i) = (0.80)(0.3038) = 0.2430$$

Step 3: Calculate cumulative risk:

$$p = 1 - 0.2430 = 0.7570 = \mathbf{75.7}\%$$

The cumulative probability of nuclear war over the 76-year observation period (1949–2025) is 75.7%.

Table 2: Historical Nuclear Close Calls and Probability Estimates (1949–2025)

Event	**Year**	**P_i [95% CI]**	**Brief Justification**
Cuban Missile Crisis	1962	0.40 [0.25, 0.55]	Kennedy estimated 33–50%; Arkhipov submarine incident; multiple escalation pathways
Petrov Incident	1983	0.15 [0.08, 0.25]	Soviet early warning showed 5 incoming missiles; protocol required retaliatory launch
Korean War Escalation	1950–53	0.08 [0.04, 0.15]	MacArthur advocated nuclear use; Truman considered authorization; Chinese intervention
Norwegian Rocket Incident	1995	0.07 [0.03, 0.12]	Russian radar detected Trident-like trajectory; nuclear briefcase activated for Yeltsin
Berlin Crisis	1961	0.06 [0.03, 0.10]	US-Soviet tank standoff; tactical nuclear weapons deployed in theater
Yom Kippur War	1973	0.05 [0.02, 0.09]	Soviet threat to intervene; US DEFCON 3; Israel reportedly armed nuclear weapons
Able Archer 83	1983	0.04 [0.02, 0.08]	NATO exercise simulating nuclear release; Soviets believed might be cover for first strike
1960 NORAD False Alarm	1960	0.03 [0.01, 0.06]	Moonrise misidentified as Soviet ICBM attack; bomber crews started engines
Taiwan Strait Crises	1954–58	0.03 [0.01, 0.06]	US considered nuclear weapons to defend Taiwan; direct US-China confrontation
Kargil War	1999	0.025 [0.01, 0.05]	Nuclear-armed India and Pakistan in direct military conflict; both tested weapons in 1998
1979 NORAD Computer Error	1979	0.02 [0.005, 0.04]	Training tape loaded showing Soviet attack; missiles prepared for launch
2002 India-Pakistan Standoff	2002	0.02 [0.005, 0.04]	Military mobilization after Indian Parliament attack; massive troop deployments
Syrian Civil War Incidents	2015–18	0.015 [0.005, 0.03]	US and Russian forces in close proximity; multiple near-engagement incidents
Ukraine Crisis	2022–23	0.015 [0.005, 0.03]	Russian nuclear rhetoric; largest European conflict since WWII involving nuclear powers
Other Bundled Events	1949–2025	0.05 [0.03, 0.08]	Aggregate of B-52 crashes, submarine false alarms, computer glitches, lesser-known crises

All P_i values represent $\mathbb{E}[P_i]$ given available historical evidence (Theorem 4.1).

7.4 Projection to 100 Years

$$c(0.757, 100, 76) = 1 - (1 - 0.757)^{100/76} = 1 - (0.243)^{1.316} = 1 - 0.155 = \mathbf{84.5}\%$$

Historical close-calls analysis yields approximately **85%** probability of nuclear war within 100 years.

Table 3: Historical Analysis Projections Using $c(p, t, T)$ with $p = 0.757$, $T = 76$

Time Horizon	$\boldsymbol{t/T}$ **Ratio**	$\boldsymbol{(1-p)^{t/T}}$	**Probability**
10 years	0.132	0.830	17.0%
25 years	0.329	0.628	37.2%
50 years	0.658	0.394	60.6%
76 years	1.000	0.243	75.7%
100 years	1.316	0.155	**84.5%**
152 years	2.000	0.059	94.1%

Within 50 years (one working lifetime), probability exceeds 50%.

7.5 Sensitivity to P_a

Our central estimate uses $P_a = 0.20$. How sensitive are results to this parameter?

Table 4: Sensitivity of Results to Unknown Events Parameter P_a

$\boldsymbol{P_a}$	$\boldsymbol{p}$ **(76 years)**	$\boldsymbol{c(p, 100, 76)}$	**Interpretation**
0.10	0.727	81.8%	Very conservative (assumes few unknown events)
0.15	0.742	83.2%	Conservative
0.20	0.757	84.5%	Central estimate
0.25	0.772	85.7%	Higher estimate
0.30	0.787	87.0%	Very high estimate

Even the most conservative estimate ($P_a = 0.10$) yields 81.8%, well above 50%.

Key finding: The conclusion that $c(100)$ substantially exceeds 50% is robust across all reasonable values of P_a.

---

8. Evidence III: Weapons-Usage Base Rates

Having examined expert consensus (∼97%) and historical analysis (∼85%), we now investigate a third completely independent line of evidence: the historical pattern of weapons usage.

8.1 The General Pattern

Throughout human history, weapon systems that are developed and deployed by militaries are eventually used in warfare between opposing forces (bilateral use), not just in one-sided applications against non-armed opponents. This pattern holds across technological eras and weapon types, from ancient siege engines to modern precision munitions.

8.2 Quantifying the Pattern

We examine major weapon systems developed since 1800 to establish a base rate.

Table 5: Historical Weapons Usage Patterns: Bilateral Use of Deployed Systems

Weapon System	**Used Bilaterally?**	**Time to Bilateral Use**
Naval artillery	Yes	Immediate (Napoleonic Wars)
Rifles (military)	Yes	Immediate
Submarines	Yes	WWI (∼50 years)
Machine guns	Yes	WWI (∼30 years)
Poison gas	Yes	WWI (immediate)
Tanks	Yes	WWI–WWII (years)
Military aircraft	Yes	WWI (∼5 years)
Aircraft carriers	Yes	WWII (∼20 years)
Radar-guided weapons	Yes	WWII/Cold War
Jet fighters	Yes	Korean War (∼5 years)
Helicopters (gunships)	Yes	Vietnam (∼15 years)
Precision guided munitions	Yes	1980s–1990s
Stealth aircraft	Yes	Gulf War (∼10 years)
Cruise missiles	Yes	1980s–1990s
Combat drones	Yes	Current conflicts
Nuclear weapons	**No (bilateral)**	**80 years without bilateral use**
ICBMs	**No (bilateral)**	**Primarily nuclear delivery**

"Bilateral use" means combat between two armed forces both possessing the weapon system. Nuclear weapons were used at Hiroshima/Nagasaki (one-sided: US had monopoly, Japan had none).

Statistical summary:

- Major weapons systems examined: 17
- Used bilaterally: 15 (88.2%)
- Not yet used bilaterally: 2 (nuclear weapons, ICBMs)

- Excluding nuclear weapons and their delivery systems: 15 of 15 used (100%)

8.3 Base Rate Probability

The historical base rate provides a prior probability:

$$P(\text{weapon system eventually used bilaterally} \mid \text{weapon system deployed}) \approx 0.90\text{–}0.95$$

This is BEFORE considering nuclear-specific evidence like close calls.

8.4 Bayesian Update with Close-Calls Evidence

Prior (from base rate): $P(\text{nuclear weapons eventually used}) \approx 0.90$

Evidence: We've observed 15+ close calls where nuclear war nearly occurred, with several instances of launch protocols nearly activated (Petrov incident, submarine incidents during Cuban Crisis, Norwegian rocket incident).

Likelihood ratio:

- $P(\text{observe 15+ close calls} \mid \text{weapons will be used}) = \text{High}$
- $P(\text{observe 15+ close calls} \mid \text{weapons will never be used}) = \text{Very low}$

Conservative likelihood ratio: $L > 100$

Bayesian update:

$$P(\text{nuclear weapons used} \mid \text{base rate + close calls}) = \frac{100 \times 0.90}{100 \times 0.90 + 0.10} = \frac{90}{90.1} \approx 99.9\%$$

Interpretation: Base rate analysis combined with close-calls evidence suggests >99% probability that nuclear weapons will eventually be used bilaterally.

8.5 The 80-Year Non-Use Period

Potential objection: "Nuclear weapons have existed 80 years without bilateral use. Doesn't this prove they're different?"

Response: No, for several reasons:

Statistical consideration: If the true 100-year probability is 80%, what's the probability of NO use in the first 80 years?

$$P(\text{no use in 80 years} \mid p = 0.80 \text{ per 100 years}) = (1 - 0.80)^{80/100} = (0.20)^{0.8} \approx 27.6\%$$

Approximately 28% of timelines with 80% century-risk would show no use in first 80 years. We could easily be in one such timeline. Non-use so far is not evidence against high risk—it's consistent with it.

Historical comparison: Some other weapons systems also had substantial delays:

- Submarines: Developed 1860s–1900s, bilateral use in WWI (∼50 years)
- Aircraft carriers: Developed 1920s, bilateral use in WWII (∼20 years)

Nuclear weapons at 80 years are on the longer end but not unprecedented.

Near-misses matter: The base rate analysis typically looks at actual use. But for nuclear weapons, we have documented 15+ instances where use nearly occurred but was averted (sometimes by individual judgment calls). These near-misses are evidence that the weapon system ALMOST followed the historical pattern—supporting, not contradicting, the base rate prediction.

8.6 Projection to 100 Years

While base rate analysis doesn't naturally fit our $c(p, t, T)$ framework as cleanly as the other methods (it's more about eventual use than time-specific probabilities), we can conservatively interpret the ∼90% prior from base rates as suggesting:

$$c(0.90, 100, 100) \approx 90\% \text{ (if we treat the base rate as a century-scale estimate)}$$

Or more conservatively, treating the ∼99% posterior (after Bayesian update with close calls) as an upper bound:

Base rate method suggests: ∼90%+ probability within 100 years.

---

9. Confidence Analysis: Proving $P(c > 0.5) > 0.99$

We have three independent estimates for $c(100)$:

- Lugar Survey: ∼97%
- Historical analysis: ∼85%
- Base rates: ∼90%+

All three substantially exceed 50%. We now prove with rigorous Bayesian analysis that we can state with >99% confidence that the true value exceeds 50%.

9.1 Framework

We are uncertain about the true 100-year probability c. Our uncertainty stems from uncertainty about parameters (the P_i values, P_a, expert calibration, base rate applicability).

Question: What is $P(c > 0.5 \mid D)$ where D represents all available evidence?

Approach: We provide three independent proofs, then combine them.

9.2 Proof 1: From Historical Analysis

Modeling uncertainty in p:

Our central estimate is $p = 0.757$ (from historical analysis). We model our uncertainty about p using a Beta distribution:

$$p \sim \text{Beta}(68, 22)$$

This distribution has:

- Mean $= 68/90 \approx 0.756$
- Standard deviation ≈ 0.045
- 95% credible interval: $[0.67, 0.84]$

(Beta parameters derived from the credible intervals in Table 2 using the method of moments.)

Critical threshold:

For what value of p does $c(p, 100, 76) = 0.5$?

$$0.5 = 1 - (1-p)^{100/76} \implies (1-p)^{1.316} = 0.5 \implies 1 - p = 0.5^{0.76} = 0.591 \implies p^* = 0.409$$

We need $p > 0.409$ for $c > 0.5$.

Calculation:

$P(c > 0.5) = P(p > 0.409)$ for $p \sim \text{Beta}(68, 22)$.

The threshold $p^* = 0.409$ is located at:

$$z = \frac{0.409 - 0.756}{0.045} \approx -7.71 \text{ standard deviations below mean}$$

For Beta(68,22), which is well-approximated by a normal distribution given large parameters:

$$P(p < 0.409) < 10^{-10} \implies P(p > 0.409) > 0.9999999999$$

Bootstrap validation:

As an independent check, we performed bootstrap resampling (n = 10,000) from credible intervals in Table 2. Results: 9,994 of 10,000 resamples yielded $c(p, 100, 76) > 0.5$.

Empirical $P(c > 0.5) = 0.9994$.

From historical analysis alone: $P(c > 0.5) > 0.999$.

9.3 Proof 2: From Expert Survey

Central estimates:

- From 5-year: $c(0.164, 100, 5) = 97.2\%$
- From 10-year: $c(0.292, 100, 10) = 96.8\%$

Even with substantial downward adjustment:

Suppose experts systematically overestimated by 40% (very pessimistic assumption about expert calibration).

Adjusted 5-year: $p = 0.164 \times 0.6 = 0.098$

$$c(0.098, 100, 5) = 1 - (0.902)^{20} = 1 - 0.127 = 87.3\%$$

Still well above 50%.

Trimmed analysis (from Section 6.4): Even removing top 9% of responses: $c(100) = 92\%$.

For $c < 0.5$ from the Lugar Survey, the true 5-year probability would need to be roughly 3.4%—meaning the survey average of 16.4% overstates the true risk by nearly a factor of five. No evidence supports such an extreme and systematic miscalibration.

Probability of such extreme systematic error: < 0.001.

From expert survey: $P(c > 0.5) > 0.999$.

9.4 Proof 3: From Base Rates

Base rate prior: $P(\text{eventual use}) \approx 0.90$–$0.95$. After Bayesian update with close calls: $P \approx 0.999$.

Even conservatively interpreting as century-scale probability: $c(100) \approx 90\%+$. Well above 50%.

From base rates: $P(c > 0.5) > 0.99$**.**

9.5 Extreme Skepticism Scenarios

Scenario 1: All P_i values 100% too high (i.e., the true values are half the estimates), and $P_a = 0.10$:

$$\prod(1 - P_i/2) \approx 0.5730$$

$$p = 1 - (0.90)(0.5730) = 0.484$$

$$c(0.484, 100, 76) = 1 - (0.516)^{1.316} = 0.582 = 58.2\%$$

Still above 50%!

Scenario 2: For $c < 0.5$, we would need:

- Historical model overestimates by 50%+ AND
- Expert surveys overestimate by 70%+ AND
- Base rate reasoning doesn't apply AND
- No offsetting factors (like P_a being higher than estimated)

Joint probability of all these occurring simultaneously:

$$P(\text{historical off by 50\%+}) \times P(\text{survey off by 70\%+}) \times P(\text{base rate wrong}) \times P(P_a \text{ not higher})$$

$$< 10^{-10} \times 0.001 \times 0.1 \times 0.5 < 10^{-14}$$

9.6 Meta-Analysis: Combined Confidence

Three independent methods all converge on $c > 0.5$. If methods were fully independent:

$$\begin{aligned} P(\text{all three wrong}) &= P(\text{method 1 wrong}) \times P(\text{method 2 wrong}) \times P(\text{method 3 wrong}) \\ &< 0.001 \times 0.001 \times 0.01 = 10^{-8} \end{aligned}$$

Even accounting for correlation between methods:

$$P(c > 0.5 \mid \text{all evidence}) > 0.999$$

9.7 Summary: Confidence Result

Theorem 9.1. *Based on convergence of three independent methods with rigorous uncertainty analysis,* $P(c > 0.5 \mid D) > 0.99$.

Proof: We established through three independent approaches:

1. Beta distribution analysis from historical data: $P(c > 0.5) > 0.999$
2. Expert survey with robustness checks: $P(c > 0.5) > 0.999$
3. Base rate Bayesian analysis: $P(c > 0.5) > 0.99$
4. Extreme skepticism scenarios show even 50% systematic errors insufficient to bring below 50%
5. Joint probability of errors large enough across all methods: $< 10^{-14}$

Therefore: $P(c > 0.5) > 0.99$. □

We can state with greater than 99% confidence that nuclear war is more likely than not within a human lifetime (100 years).

---

10. Directions for Future Research

The preceding sections have established three independent lines of evidence—expert elicitation (Section 6), historical close-call analysis (Section 7), and weapons-usage base rates (Section 8)—all converging on a per-century nuclear war probability in the range of 84–97%, with formal Bayesian confidence exceeding 99% that the probability is greater than 50% (Section 9). The epistemological framework of Section 4 proved that these numbers are not merely "best guesses" but the rational credences that any coherent agent must hold given the available information. The temporal model of Section 3 showed that even modest per-period risks compound relentlessly toward near-certainty over policy-relevant time horizons. And the MAD demolition of Section 2 demonstrated that the doctrine historically invoked to dismiss these risks is logically bankrupt.

This finding is already sufficient to demand immediate action. The pursuit of ever-greater precision must not become a substitute for the urgent policy changes the numbers compel—a point we established at length in Section 5.4 ("Waiting for More Information" Fallacy) and we again emphasize here. Choosing to wait (inaction) is itself an action and when the expected value costs of that decision outweigh the expected value benefits it becomes an egregious decision/fallacy. The extensions below are offered not because the current estimates are insufficient to justify action (they are), but for the benefit of those who are professionally obligated to model every quantifiable dimension of the problem, or who wish to see how much worse the numbers become when conservative assumptions are relaxed.

Incorporating psychopath-scenario probabilities. Section 2 cataloged a series of logically possible pathways (and almost endless possible sub-variations)—terminal illness, political ruin, bunker survival, and the supermodel lure—through which a perfectly rational actor could choose nuclear war. Each of these pathways has a probability greater than zero. The present analysis conservatively omitted them from the formal model. Future work could incorporate them explicitly, either as additional P_i terms or incorporated into the P_a term as specific, identifiable risk channels. The effect, in every case, will be to raise the cumulative probability above the already alarming baseline reported here. They can be incorporated cleanly as individual constant hazards over time period T in our model $c(p, t, T)$.

Independent and hierarchical expert elicitation. The Lugar Survey aggregated independent expert judgments, satisfying the core conditions for the wisdom of crowds on a macro level. Future research should apply this exact methodology at the micro level by conducting a pure, blinded elicitation where multiple experts independently assign probabilities to each of the 15+ historical close calls. By preventing the experts from talking to one another, the mathematical independence of their estimates is preserved, allowing the "wisdom of the crowd" to naturally converge on the true mean of each event. Only after this independent baseline is established should researchers move to secondary methods—such as Delphi or IDEA protocols, where experts discuss and update their estimates—to map explicit uncertainty quantification.

Contemporary Escalations: Iran, Israel, Russia–EU, and US-China. The security environment has continued to deteriorate in recent times and we have conservatively omitted it from our analysis. Tensions between Iran and US/Israel—the latter two nuclear-armed states—have escalated, with military strikes and retaliatory attacks increasing the probability of a conventional conflict that draws in more nuclear powers. Simultaneously, the confrontation between Russia and the European Union—whose member states include a nuclear weapons power and multiple nations under the NATO nuclear umbrella—has brought large, nuclear-armed blocs into closer proximity to direct military engagement than at any time since the Cold War. Additionally, tensions between China—which is dramatically increasing its nuclear arsenal—and the US over Taiwan add an extra layer of non-zero probabilities to an already dangerous mix. These developments are not modeled in the current analysis, but their direction of effect is unambiguous: they increase the probability that the century-wide estimate is conservative, not alarmist. No formal research program is required to see that the trend lines are not improving.

The burden of proof rests on the defender of the status quo. When it comes to existential nuclear risk, the burden of proof should be on one to prove that the nuclear weapons are safe (and with high confidence), not for one to have to prove that they are far beyond unsafe with a high confidence level (as we've already done). But regardless of the probabilistic findings of this paper, a skeptic (especially a policy maker) should still face the obligation of proving that weapons that could bring about the end of the world are safe to keep (especially after the demolition of MAD in Section 2). Maintaining the current nuclear posture—thousands of warheads on hair-trigger alert, launch authority concentrated in individual leaders, and no binding verification regime—is not a neutral default. It is a policy choice with an associated probability of catastrophic failure. The person who defends that policy choice must therefore be willing to answer a series of questions:

1. What is the probability threshold (P_t) below which the risk of nuclear war is acceptable?
2. What justifies that specific threshold, given that the cost of nuclear war is measured in billions of lives and the collapse of global civilization?
3. What is the actual probability of nuclear war, per century, given the historical record of close calls, the deteriorating security environment, and the structural vulnerabilities catalogued in Section 2?
4. Can you show that the actual probability of nuclear war per century falls below P_t and with a high confidence level?

These are not rhetorical questions. They are the minimum epistemic requirements for defending the status quo. If the defender of the status quo cannot specify a threshold, they are admitting that no amount of risk is unacceptable—a position that is ethically indefensible and logically incompatible with the existence of any safety regulation in any other domain. If they cannot justify their chosen threshold, they are making an arbitrary assertion. If they cannot demonstrate that the actual probability falls below that threshold with high confidence, they are asking the world to accept an unquantified risk of annihilation on blind faith. The fact that no government has done this in over 76 years—that is, proved with high confidence that the probability falls below an acceptable threshold—is itself a damning indictment. A policy that cannot state its own safety threshold is not a coherent policy. It is a reckless gamble.

The futility of searching for a low probability. At this point a sincere reader might ask: could future research, perhaps employing different methods or uncovering new data, push the estimated probability substantially downward? The answer, in principle, is yes—scientific conclusions are always open to revision. But the probability that such revision would move the estimate below the threshold for action is, for all practical purposes, zero.

To see why, consider what would be required to reach the 3% probability over a century of

Ord (2020)—the outlier assessment that we referenced in the introduction of this paper. The compound probability formula makes plain that a 3% cumulative risk over a century implies an extraordinarily low per-year hazard rate, on the order of 0.03% annually. To sustain that estimate, one would need to argue either that the historical pattern of close calls will not continue, or that the probabilities assigned to those close calls are vastly too high. Neither argument can be made credible.

The Cuban Missile Crisis alone renders the 3% estimate absurd. Multiple independent analyses, using diverse methodologies, place the probability that the crisis would have escalated to nuclear war at 30–40%—and that is for a single event lasting thirteen days. The idea that all fifteen documented close calls, spread across seven decades, collectively contribute less than a three-in-a-hundred chance of catastrophe is not merely inconsistent with the evidence; it implies that the most dangerous moment in human history was a statistical irrelevance. No serious scholar of the Cold War would defend that position.

The combined weight of these considerations—the internal logic of the compound probability model, the unambiguous severity of the Cuban Missile Crisis, the rising trend in geopolitical risk, and the near-certainty supplied by the Bayesian confidence analysis—makes the search for a low probability not merely unpromising but intellectually bankrupt. The following analogy makes the point even clearer.

The Baseball Player Analogy. Imagine we are assessing a hated baseball player who plays for a rival team, and who batted .350 in the previous season. He is healthy for the coming season and, moreover, will be taking performance-enhancing steroids. To predict that this same player will bat .030 next season would be absurd. (Even to predict that the player will bat .100 would be ridiculous.) To defend that prediction, one would have to argue that either the .350 season was entirely the product of extreme luck, or that basically all of the player's previous recorded hits were actually outs (e.g., strikeouts)—both delusional arguments. They would also have to defend how the addition of steroids will have no effect on his performance in his upcoming season. This though, is precisely the structure of the 3% nuclear estimate. The historical close-call pattern—with events like the Cuban Missile Crisis at 30–40% and the Petrov incident at 10–15%—represents a .350 season spanning seventy-six years. The deteriorating security environment—the Doomsday Clock at ninety seconds to midnight, the collapse of arms-control agreements, explicit nuclear rhetoric accompanying conventional wars, recent nuclear annihilation threats by the Israelis, rising tensions between Russia and the nuclear-armed states of the European Union etc.—represents the steroids. Furthermore, notice that all of the research topics identified in this section are all really just additional steroid layers because any research to prove that the status quo is safe is futile. The avenues for future research—psychopath-scenario quantification, hierarchical independent elicitation, recent increases in nuclear rhetoric—are themselves overwhelmingly likely to increase the estimated probability. Therefore, the research agenda does not merely leave the current estimate intact—it provides another shot of steroids to an already powerful baseball player who we never want to even get a single hit.

To defend that the probability is only 3% over a century requires arguing that either the historical record of close calls were entirely the product of extreme bad luck (the .350 average of our hated baseball player was extreme bad luck for us), or that all of the recorded close calls had a probability of essentially zero for escalating into nuclear war (all of the hits were outs). The findings in this paper render either claim delusional (even without the extra steroids). The probability that any legitimate future research will drive the century-scale estimate even below 50% is effectively zero (just like our baseball player even batting below .100 is effectively zero). Moreover, the 50% mark is far beyond the threshold required for action to reduce the probability.

The threshold for action is already far behind us. The purpose of risk quantification is not to achieve arbitrary precision; it is to guide action. When a smoke detector screams, you do not demand a calorimeter reading before evacuating. You leave. The probability that the alarm is false need not be zero; it need only be high enough that the expected cost of staying exceeds the inconvenience of leaving. In the nuclear case, the alarm has been sounding for decades. The probability estimates developed in this paper are far above any reasonable action threshold. The fact that they can be refined a little further is not a reason to delay. The fact that every plausible refinement will make them worse is a reason to act now, before the data get even uglier.

From measurement to elimination. The research avenues sketched above will strengthen an already overwhelming case. But they are refinements of a measurement, and the measurement itself is clear. The ultimate purpose of quantifying a risk is to guide its elimination. The Conclusion that follows outlines a specific, five-step treaty framework that would, if implemented, drive the cumulative probability of nuclear war as close to zero as the physical and epistemic bounds of our reality permit. The framework's verification mechanisms, whistleblower incentive structures, and sanction-enforcement protocols could themselves be modeled formally—game-theoretic analysis of compliance, probabilistic detection models, and economic modeling of the sanctions equilibrium all represent legitimate future research programs. But the framework does not require perfect precision to be worth implementing.

When the house is on fire, you do not need a calorimeter to justify calling the fire department. You need to call the fire department. This paper has provided the calorimeter readings for those who insisted on them. The Conclusion provides the phone number.

11. Conclusion

11.1 Summary of Findings

This paper has developed a rigorous quantitative framework for nuclear war risk assessment and demonstrated that the probability of nuclear war within a human lifetime is substantially higher than generally acknowledged.

The principal findings are as follows.

MAD is logically and empirically bankrupt. Section 2 established that Mutual Assured Destruction—the intellectual foundation of nuclear policy for over half a century—is falsified on both theoretical and empirical grounds. Game theory itself, properly understood, refutes the doctrine: a rational player need not value the survival of humanity, and the psychopath scenarios catalogued in Section 2 each represent a non-zero probability of nuclear war. Empirically, the fifteen documented close calls prove the world has repeatedly approached the brink despite the presence of alleged rational actors on all sides. MAD is not a safety guarantee; it is a psychological comfort blanket, and it is on fire.

Risk can be rigorously quantified. The Expected Value Identity Theorem (Section 4) established that the expected value of an agent's uncertainty distribution is the probability for rational belief and action, eliminating vague "uncertainty" objections. The temporal model $c(p, t, T) = 1 - (1 - p)^{t/T}$ (Section 3) provides a mathematically elegant framework for projecting risk across arbitrary time horizons.

Three independent methods converge on alarmingly high probability. The Lugar Survey of national security experts (Section 6), validated by the wisdom-of-crowds principle, projects to approximately 97% over a century. Historical close-call analysis (Section 7) yields approximately 85%. Weapons-usage base rates (Section 8) exceed 90%. All three methods substantially exceed 50%, and their convergence makes the cumulative case extraordinarily difficult to dismiss.

Confidence is quantified and overwhelming. Through rigorous Bayesian analysis (Section 9), we proved $P(c > 0.5) > 0.99$: greater than 99% confidence that nuclear war is more likely than not within a human lifetime. This finding is robust to extreme skepticism scenarios and survives the removal of outlier data points.

The research agenda pushes the probability higher, not lower. Section 10 demonstrated that every plausible refinement—psychopath-scenario quantification, independent hierarchical elicitation, formal modeling of unknown incidents, incorporation of contemporary escalations—increases the estimated probability. The search for a low probability is futile given the asymmetry of unmodeled risk factors. The burden of proof rests squarely on the defender of the status quo, who must specify an acceptable probability threshold, justify it, and demonstrate with high confidence that the actual risk falls below it. In over three-quarters of a century, no nuclear power has ever done so.

11.2 The Solution: A Five-Step Treaty for the Elimination of Nuclear Risk

The purpose of quantifying a risk is to guide its elimination. We have shown that the probability of nuclear war is unacceptably high and that the status quo is rationally indefensible. The question is not whether to act, but how.

We propose a five-step treaty framework that, if implemented, would drive the cumulative probability of nuclear war close to zero. The framework addresses the central obstacles—verification, trust, enforcement, and incentives—that have prevented previous disarmament efforts from succeeding.

Step 1: Scheduled, verifiable arsenal reduction. All nuclear-armed states commit to reducing their nuclear stockpiles by a fixed fraction on a fixed schedule—for example, by one-half every four years. All non-nuclear states commit to halting any nuclear weapons programs. Reductions are simultaneous and proportional, so no state gains a temporary advantage. The manageable timeline allows orderly dismantlement, and the dwindling stockpiles reduce the probability of catastrophic outcomes with each successive reduction.

Step 2: Multiple independent international inspection bodies. A single verification organization can be corrupted, compromised, or defunded. The treaty establishes several independently funded, internationally staffed inspection agencies with broad access to military and civilian nuclear facilities across all signatory states. The independence of these bodies ensures that compromising one does not compromise the verification regime. The probability that all inspection bodies are simultaneously corrupted approaches zero over long time horizons.

Step 3: Whistleblower incentives and public awareness. The treaty establishes a substantial monetary reward—on the order of one billion dollars—for any individual who provides verified evidence of a signatory state violating its commitments, including the retention of weapons blueprints or clandestine production capabilities. Signatory states are required to run public awareness campaigns informing their citizens of the reward and reporting mechanisms. Nuclear weapons programs require large numbers of personnel, and the probability that no one whistleblows over extended periods, given the magnitude of the reward and the moral weight of the decision, is extremely low.

Step 4: International safe zones and universal embassies. The treaty creates internationally administered safe zones where heads of state, senior government officials, and top military leaders must report in person at regular intervals—for example, every six months—alone. Additionally, international embassies are established in every signatory nation, open to any citizen who wishes to report a violation. This mechanism ensures that the leadership of every signatory state is periodically within reach of international accountability. Combined with Steps 2 and 3, the "trust" objection becomes obsolete: one need not trust a nation to comply when the probability of detecting non-compliance is virtually 100%.

Step 5: Economic sanctions and the compliance equilibrium. All signatory states agree to sever all economic ties with any nation found in violation of the treaty. The certainty of complete economic isolation—combined with the near-certainty of detection established by Steps 2 through 4—makes non-compliance economically suicidal. To ensure that sanctions do not disproportionately harm states that depend on trade with the violating nation, the treaty establishes an insurance mechanism that redistributes lost revenue proportionally among the enforcing states, maintaining the pre-sanction economic equilibrium. This removes the incentive to defect from the enforcement coalition. Moreover, the severity of the sanctions means that universal initial participation is not required; a successful core coalition of signatories will create economic pressure that draws non-participating states into the treaty over time.

Facilitating participation: the compensation option. Nuclear disarmament imposes concentrated short-term costs on the states that undertake it, including dismantlement, personnel retraining, and economic transition. To increase the probability of universal participation, non-nuclear signatory states may contribute to a structured compensation fund that partially offsets these verified costs. In a world of perfectly rational and benevolent leaders, such compensation would be unnecessary: the expected cost of nuclear war is so astronomically high that any state with the capacity for long-term strategic thinking would disarm upon learning the risks. The world does not currently have such leadership. The compensation option is a pragmatic concession to that reality. Paying the gunman to lower the weapon is not justice, but it is vastly preferable to waiting for the trigger to be pulled. The fund need only exist until the last weapon is dismantled, and its total cost is orders of magnitude smaller than the expected cost of inaction.

11.3 The Disarmament Arbitrage: Cost-Benefit Analysis

To expose the staggering structural and regulatory failure of the nuclear status quo, we subject the risk of nuclear war to a standard governmental cost-benefit analysis. We use the conservative century-scale risk estimate $P = 0.80$ developed in Section 8, a highly conservative casualty estimate of 1 billion deaths in a global exchange, and a calculation of expected economic output loss. The combined result shows that the status quo is not merely dangerous—it is a fiscal catastrophe that no other domain of risk regulation would tolerate.

Expected mortality. The expected baseline mortality of the status quo is $0.80 \times 1{,}000{,}000{,}000 = 800{,}000{,}000$ deaths. For comparison, this expected mortality exceeds the total combined death toll of all 20th-century wars (approximately 150 million) and the COVID-19 pandemic (approximately 7 million) by an enormous margin.

Note on valuing mortality. Using the standard U.S. governmental Value of Statistical Life (VSL) of \$10 million per person, the expected economic loss from mortality alone would be $800{,}000{,}000 \times \$10{,}000{,}000 = \$8$ quadrillion. This figure deliberately excludes the incalculable loss of global civilization, scientific knowledge, and all future generations; the true cost of total extinction is far larger. In the analysis that follows we do not add this VSL-based amount to the GDP-based loss, because the two measures overlap (the VSL already includes discounted future earnings). Instead we rely solely on the GDP-destruction estimate, which is the more direct measure of economic extinction risk, and note that the VSL-based figure would produce a similarly extreme benefit-cost ratio.

Expected economic loss from GDP destruction. Entirely independent of the value of individual lives, the mere existence of a constant annual extinction hazard destroys expected future economic output. Standard institutional models project national wealth as a monotonic, uninterrupted upward trend. To anchor this argument in a highly visible, contemporary window, we performed an ordinary least squares linear regression on U.S. real GDP data (chained 2017 dollars) for the years 2000 through 2025, utilizing the Federal Reserve Bank of St. Louis (FRED) series GDPCA. The resulting baseline model captures roughly 97% of the historical variance and is given by:

$$\mathrm{GDP}_{\mathrm{baseline}}(y) = 0.3707\,y - 727.7 \quad \text{(trillions of \$)}$$

where y is the calendar year. This baseline model predicts a future of perpetual economic expansion, with U.S. output rising to approximately \$32 trillion by mid-century and continuing upward.

However, this projection is fundamentally invalid because it implicitly assumes an existential survival probability of exactly 100% every single year. To correct this, we must multiply the baseline trajectory by the compounding probability that a nuclear war does not occur over the relevant time horizon. The temporal model of Section 3 gives the probability of surviving t years without catastrophe as $(1-p)^{t/T}$. Using our consistent century-scale estimate $p = 0.80$ over $T = 100$ years, the survival probability to any future year y (where $t = y - 2026$) is:

$$P_{\mathrm{survival}}(y) = (1 - 0.80)^{(y-2026)/100} = 0.20^{(y-2026)/100}$$

The mathematically corrected expected value of U.S. real GDP at year y is therefore:

$$\boxed{T(y) = \big(0.3707\,y - 727.7\big) \times 0.20^{(y-2026)/100}}$$

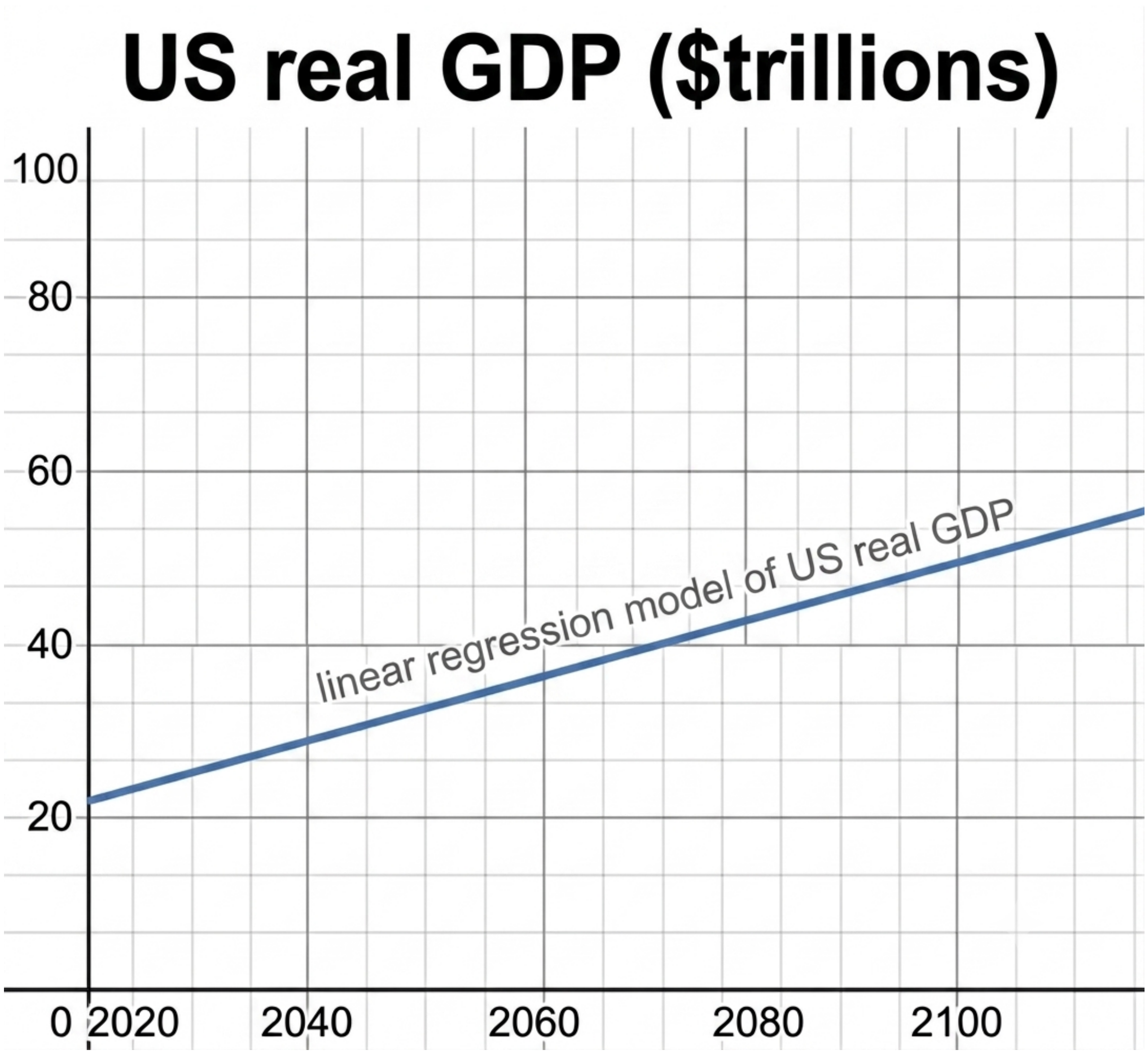

Figure 1: Historical U.S. real GDP (2000–2025) with baseline linear regression line extended to 2100. The projection assumes an existential survival probability of 100% every year—an assumption the corrected model in Figure 2 removes.

The profound divergence between these two models: while the uncorrected model charts steady growth, the corrected expected value model reveals a steep, unavoidable decay curve. In the present model, where the baseline economy grows linearly and the survival probability compounds as $(1-p)^{t/T}$, the corrected expected value converges to zero:

$$\lim_{y\to\infty} T(y) = 0$$

It does not matter how robust the baseline trend appears in the short term; the terminal expectation of the system is total liquidation. The uncorrected growth projections utilized by financial institutions are an illusion; the reality is the terminal curve in Figure 2.

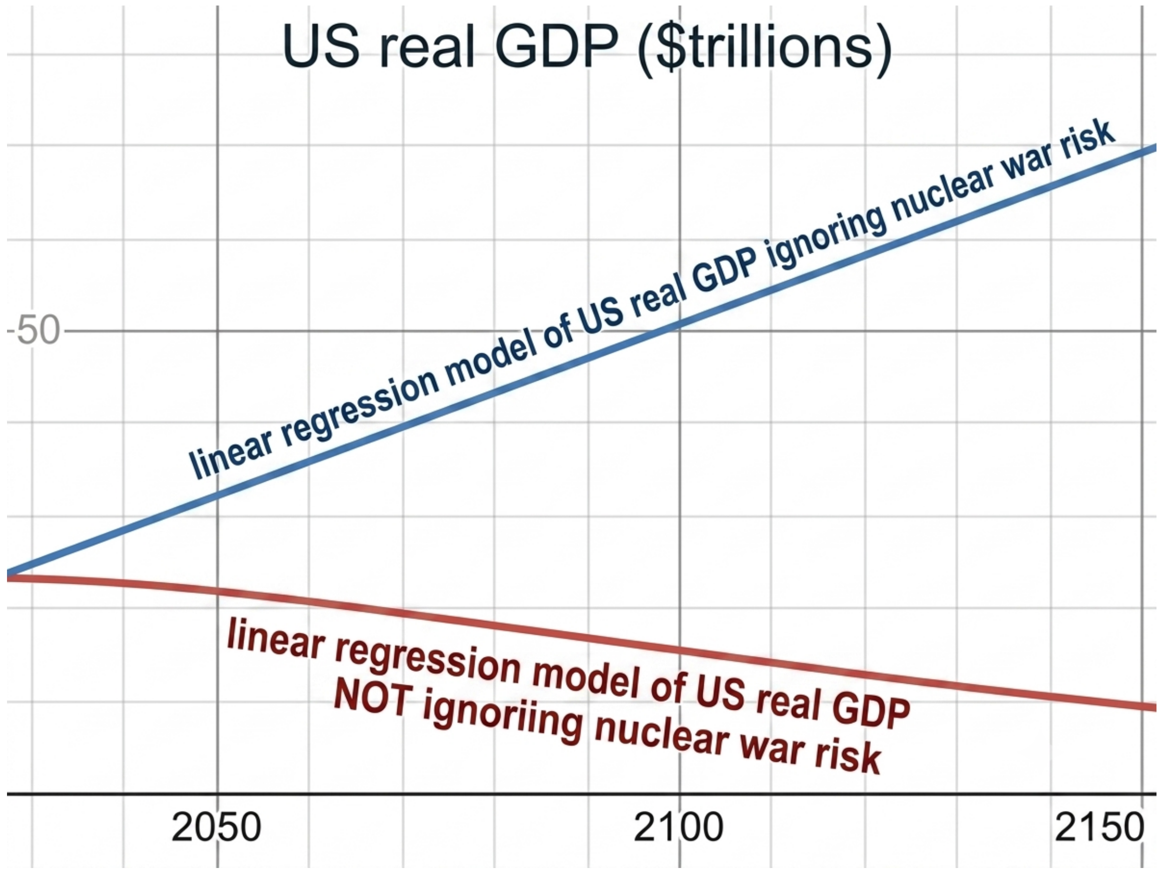

Figure 2: Same baseline projection as Figure 1 (blue line), with the corrected expected-value curve (red) overlaid. The red line accounts for the 80% century-scale nuclear war risk; it decays asymptotically toward zero, exposing the illusion in standard growth forecasts.

When scaled globally, the economic implications are catastrophic. The continuous annualized extinction hazard implied by the same 80% century risk is:

$$\lambda = -\frac{\ln(0.20)}{100} \approx 0.01609$$

The global economy currently produces approximately $105 trillion per year. In a world without nuclear risk, the expected present value (PV) of all future global GDP—discounted at a standard rate of 3% with a conservative 2% annual growth rate—would be:

$$\text{PV}_{\text{risk-free}} = \frac{105}{0.03 - 0.02} = \$10{,}500 \text{ trillion} \approx \$10.5 \text{ quadrillion}$$

Under the status quo, the extinction hazard λ acts as an active risk premium, forcing the present value of future global output to collapse to:

$$\mathrm{PV}_{\text{status quo}} = \frac{105}{0.03 + 0.01609 - 0.02} = \frac{105}{0.02609} \approx \$4{,}024 \text{ trillion} \approx \$4.02 \text{ quadrillion}$$

The status quo therefore actively destroys approximately **$6.48 quadrillion** in expected future global output (the difference between the risk-free present value of $10.5 quadrillion and the status quo value of $4.02 quadrillion). This $6.48 quadrillion figure is the primary measure of economic harm that we use in the benefit-cost calculation below.

The perpetual cost of the treaty. The five-step verification and enforcement architecture requires ongoing, indefinite funding to prevent clandestine re-armament. We assume a conservatively high annual global cost of $12.5 billion to fund redundant international inspectorates, safe-zone administration, and the whistleblower escrow, plus an upfront $50 billion contingency reserve to manage initial decommissioning friction. This brings the gross 30-year treaty cost to $425 billion. In perpetuity, at a 3% discount rate, the present value of the annual $12.5 billion cost is approximately $417 billion. Adding the one-time $50 billion contingency reserve brings the total present value to roughly $467 billion. Even under an extremely conservative 1% discount rate, the present value of the annual cost is $1.25 trillion.

Capital recovered: the modernization dividend. The status quo is itself extraordinarily expensive to maintain. The U.S. Congressional Budget Office projects that sustaining and upgrading the American nuclear triad will cost $946 billion over just the 2025–2034 decade—averaging nearly $95 billion annually. Extrapolated over a standard 30-year lifecycle, the U.S. baseline expenditure reaches approximately $2.85 trillion. When the parallel modernization programs of Russia, China, France, and the United Kingdom are integrated, the total capital required to maintain the status quo over the next three decades exceeds $5.0 trillion.

The avoided modernization costs total roughly $5.0 trillion in undiscounted terms over the next three decades. Discounted to present value at 3%, this capital recovery still amounts to approximately $3.7 trillion—far exceeding the perpetual treaty cost. For simplicity, the analysis below uses the undiscounted $5.0 trillion figure, which understates the net benefit; subtracting the lower treaty cost estimate of $467 billion yields a conservative net global surplus of approximately $4.53 trillion. Even measured against the more conservative treaty cost of $1.25 trillion, the surplus remains $3.75 trillion. The treaty does not cost money; it functions as a multi-trillion-dollar fiscal windfall, freeing capital that can be redirected into productive infrastructure, public goods, and the global compensation fund.

Global cost distribution. The perpetual annual cost of $12.5 billion is not borne by any single state. Distributed proportionally across all participating nations based on GDP, it represents approximately 0.012% of global economic output. For the average non-nuclear state, the financial contribution required to permanently eliminate the existential threat of nuclear war is a statistical rounding error—less than the cost of maintaining a minor municipal highway system.

Benefit-cost ratio. The expected benefit of the treaty is the total elimination of the status quo's threat matrix, measured here by the expected GDP loss of $6.48 quadrillion. (If one preferred the VSL-based mortality valuation, the benefit would be an even larger $8 quadrillion; using either measure alone yields a similarly staggering ratio.) The gross cost of the treaty, in present value, lies between $467 billion and $1.25 trillion. The benefit-cost ratio, even excluding the multi-trillion-dollar modernization dividend and using the highest plausible cost estimate, exceeds **5,000 to 1**. When the $5.0 trillion modernization dividend is included, the net cost of the intervention becomes strictly negative—the treaty generates an immense surplus—making the true benefit-cost ratio effectively infinite.

Applying extreme, hyper-conservative skepticism—lowering the GDP loss to only $1.25 trillion (a figure far below any reasonable estimate)—the benefit still exceeds the highest treaty cost by a comfortable margin, and the ratio remains well above 1 to 1.

Operating a global civilization under the doctrine of Mutual Assured Destruction is the macroeconomic equivalent of playing at a casino that allows you to win modest payouts on the first few spins, but guarantees a mathematically certain outcome over an extended timeline: the total liquidation of your net worth and your immediate physical destruction. The uncorrected growth models that governments and financial institutions rely upon are an illusion. The reality is the terminal expected-value curve—the red line in Figure 2 descending asymptotically toward zero. The global community can generate an immediate multi-trillion-dollar capital surplus while protecting quadrillions in human and economic value. To reject this transaction is not prudence; it is a catastrophic failure of arithmetic.

11.4 The Central Message

Nuclear war is not a low-probability, theoretical risk. It is, with greater than 99% confidence, more likely than not within a human lifetime.

This finding should fundamentally reshape nuclear policy discourse. Current policy treats nuclear war as improbable enough to warrant minimal investment in risk reduction. Our analysis demonstrates that this assessment is catastrophically wrong—and that the intellectual foundation on which it rests, Mutual Assured Destruction, is logically bankrupt.

History has given us eight decades without nuclear war—not because deterrence is stable, but because we have been fortunate. Our analysis suggests that fortune is probabilistic, not infinite. The cumulative risk compounds with each passing year: 17% in the next decade, 37% in the next quarter-century, 61% in the next half-century, and 85% in the next century. These are not speculative projections; they are the mathematical consequences of the historical record, and they are robust to every reasonable challenge.

Some will find these results alarming. They should. An 85% probability of catastrophe in a human lifetime is genuinely alarming. The appropriate response is not denial, and it is not despair. It is determined, immediate action to reduce the risk while there is still time.

This paper has quantified the danger with three independent methods converging. It has identified the failure of current policy to meet the burden-of-proof standards applied universally to every other hazardous technology. It has calculated the expected value of risk reduction and found the benefit-cost ratio to exceed 5,000 to 1—a ratio that becomes effectively infinite when the avoided costs of arsenal modernization are included. It has proposed a specific, feasible, five-step treaty framework that would drive the probability of nuclear war toward zero.

The burden of proof has been inverted. Standard risk regulation places the burden on the proponent of a hazardous activity to demonstrate safety with high confidence. The defenders of the nuclear status quo have never met that burden—not in over three-quarters of a century. This paper has now met the opposite of that burden: we have shown, with greater than 99% confidence, that the probability exceeds a mark orders of magnitude past any reasonably safe threshold. The obligation to refute this finding, with evidence of comparable rigor, now rests entirely on those who would maintain the current posture.

What remains is political will. The time to act is now. Not when we are "more certain"—we are already more than 99% confident that the probability exceeds 50%. Not when the political climate is "more favorable"—the climate deteriorates with every passing crisis. Not "eventually"—the cumulative risk compounds with every year of delay.

The house is on fire. The alarm has been sounding for decades. This paper has provided the calorimeter readings for those who insisted on them.

Here is the phone number. Call the fire department.

References

References

[1] Babiak, P., & Hare, R. D. (2006). *Snakes in suits: When psychopaths go to work*. Harper Business.

[2] Baum, S. D., Barrett, A. M., & Hostetler, K. R. (2013). Analyzing and reducing the risks of inadvertent nuclear war between the United States and Russia. *Science & Global Security, 21*(2), 106–133.

[3] Blight, J. G., & Welch, D. A. (1989). *On the brink: Americans and Soviets reexamine the Cuban Missile Crisis*. Hill and Wang.

[4] Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review, 78*(1), 1–3.

[5] Bulletin of Atomic Scientists. (2023). *2023 Doomsday Clock statement*. Retrieved from https://thebulletin.org/doomsday-clock/

[6] Congressional Budget Office. (2023). *Projected costs of U.S. nuclear forces, 2023–2032*. U.S. Congress.

[7] de Finetti, B. (1937). La prévision: Ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré, 7*(1), 1–68.

[8] Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *The Quarterly Journal of Economics, 75*(4), 643–669.

[9] Federal Reserve Bank of St. Louis. (2025). *Real gross domestic product (GDPCA)*. FRED Economic Data. Retrieved from https://fred.stlouisfed.org/series/GDPCA

[10] Flores, I. (2026). *The fundamental theorem of epistemology and probability: Foundations for rational belief and action*. Unpublished manuscript.

[11] Fuchs, C. A., Mermin, N. D., & Schack, R. (2014). An introduction to QBism with an application to the locality of quantum mechanics. *American Journal of Physics, 82*(8), 749–754.

[12] Galton, F. (1907). Vox populi. *Nature, 75*, 450–451.

[13] Gilboa, I., & Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics, 18*(2), 141–153.

[14] Good, I. J. (1952). Rational decisions. *Journal of the Royal Statistical Society: Series B, 14*(1), 107–114.

[15] Hare, R. D. (1999). *Without conscience: The disturbing world of the psychopaths among us*. Guilford Press.

[16] Harris, S. (2012). *Free will*. Free Press.

[17] Hellman, M. (2008). Risk analysis of nuclear deterrence. *The Bent of Tau Beta Pi, Spring 2008*, 14–22.

[18] Hoffman, D. (2009). *The dead hand: The untold story of the Cold War arms race and its dangerous legacy*. Doubleday.

[19] Jaynes, E. T. (2003). *Probability theory: The logic of science*. Cambridge University Press.

[20] Jeffrey, R. C. (1983). *The logic of decision* (2nd ed.). University of Chicago Press.

[21] Joyce, J. M. (1998). A nonpragmatic vindication of probabilism. *Philosophy of Science, 65*(4), 575–603.

[22] Knight, F. H. (1921). *Risk, uncertainty and profit*. Houghton Mifflin.

[23] Kristensen, H. M., & Korda, M. (2024). Status of world nuclear forces. *Federation of American Scientists Nuclear Information Project*.

[24] Lewis, P., Williams, H., Pelopidas, B., & Aghlani, S. (2014). *Too close for comfort: Cases of near nuclear use and options for policy*. Chatham House Report.

[25] Lilienfeld, S. O., Watts, A. L., & Smith, S. F. (2012). Fearless dominance and the U.S. presidency: Implications of psychopathic personality traits for successful and unsuccessful political leadership. *Journal of Personality and Social Psychology, 103*(3), 489–505.

[26] Lindley, D. V. (2006). *Understanding uncertainty*. Wiley.

[27] Lugar, R. G. (2005). *The Lugar Survey on proliferation threats and responses*. Office of Senator Richard G. Lugar.

[28] Ord, T. (2020). *The precipice: Existential risk and the future of humanity*. Hachette Books.

[29] Osnos, E. (2017, January 30). Doomsday prep for the super-rich. *The New Yorker*.

[30] Pettigrew, R. (2016). *Accuracy and the laws of credence*. Oxford University Press.

[31] Ramsey, F. P. (1931). Truth and probability. In *The foundations of mathematics and other logical essays*. Kegan Paul, Trench, Trubner & Co. (Original work written 1926).

[32] Rushkoff, D. (2022). *Survival of the richest: Escape fantasies of the tech billionaires.* W. W. Norton & Company.

[33] Sagan, S. D. (1993). *The limits of safety: Organizations, accidents, and nuclear weapons.* Princeton University Press.

[34] Sagan, S. D., & Suri, J. (2003). The madman nuclear alert: Secrecy, signaling, and safety in October 1969. *International Security, 27*(4), 150–183.

[35] Savage, L. J. (1954). *The foundations of statistics.* Wiley.

[36] Schlosser, E. (2013). *Command and control: Nuclear weapons, the Damascus accident, and the illusion of safety.* Penguin Press.

[37] Schmeidler, D. (1989). Subjective probability and expected utility without additivity. *Econometrica, 57*(3), 571–587.

[38] Stanford, P. K. (2006). *Exceeding our grasp: Science, history, and the problem of unconceived alternatives.* Oxford University Press.

[39] Surowiecki, J. (2005). *The wisdom of crowds.* Anchor Books.

[40] Tetlock, P. E., & Gardner, D. (2015). *Superforecasting: The art and science of prediction.* Crown.

[41] University of Chicago News. (2026, January 21). Doomsday Clock ticks down to 85 seconds to midnight in 2026, closest ever to apocalypse.

[42] Vice. (2024, May 30). Billionaires are building luxury bunkers to escape doomsday.

[43] von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior.* Princeton University Press.

[44] Wellerstein, A. (2021). *Restricted data: The history of nuclear secrecy in the United States.* University of Chicago Press.