RiskyMAD: The Existential Risk Forecast and the MAP Escape

Laurence Loewe of Laodicea[1,2,3,4,5], AI Claude Opus 4.6-4.7 Max[6,7], and Everyone[8]

[1] Balospe and Evolvix Research (Balospe.com)
[2] Formerly Laboratory of Genetics and Wisconsin Institute for Discovery, UW-Madison
[3] Email: LLoL@balospe.org | ORCID: 0000-0002-6253-9269 | Google Scholar (lBchRzQAAAAJ)
[4--9] See **Declarations** below for more essential background.

Broader Significance

Nuclear deterrence is usually debated in words; this paper debates it in numbers. RiskyMAD is a deliberately small three-state stochastic model (Risky / MAD / Dead), calibrated to the Cold War record of nuclear near-misses, that treats civilizational survival the way an actuary treats a life: as a probability distribution over time-to-failure. Its central finding is sobering and auditable --- at the observed crisis rate the median time to accidental nuclear winter is about 19 years, and roughly 1 in 40 runs reaches catastrophe within a single year, a risk no airline, regulator, or insurer would tolerate in any other domain.

The result is not a prediction of doom but a diagnosis with a treatment. Because the Dead state is absorbing, accidental nuclear winter is a stochastic certainty under business as usual --- the only open question is when. The paper then derives an escape: MAP (Mutually Assured Progress) converts the nuclear Prisoner's Dilemma into an Assurance Game through a credible first-mover and periodic recalibration via the Jubilee System. Every number is reproducible from published model code, so the argument can be checked rather than believed.

Declarations

[4] "of Laodicea" indicates taking responsibility to undo personal complicity with disastrous Laodicean legacies like banning mathematicians from clergy (Canon 36, Council of Laodicea; two magisteria separations), enabling institutional lukewarmness, weapons of math-destruction, and slow-motion explosions of misinformation from pandemics to self-compounding interests.
[5] LLoL stands for ridiculous luck in serendipitous discovery and a commitment to find ever more fun ways to help others uncover street-wise math that matters. He hopes to convert nuclear roulette into a survivable path through MAP.
[6] by Anthropic (anthropic.com; evolves and operates Claude; not responsible for Loewe's errors in using AI)
[7] Named AI co-author for many substantial contributions, because the practical singularity (PraS, see Matheo-b21) changed how this paper was written. After PraS, useful AI insight generation outpaces human review on tested topics. Hence, Loewe's traditional standards for co-authorship demand naming AI Claude Opus 4.6-4.7 Max as a co-author, as if a PhD-student. Forward accountability (for all AI use & texts) rests with Loewe as senior corresponding author (like done for deceased authors, consortia, or young graduate students). Anthropic is not responsible for AI mistakes here. This study uses the AI co-authorship framework in Matheo-b21 to help rethink long-term use of AI in a ResearchCity serving the common good.
[8] This aggregated open co-author group invites all who wish to retroactively join the conversation under the open co-authorship framework defined in Matheo-b21. As Everyone cannot consent to co-authorship, all accountability rests with Loewe as senior corresponding author (until explicitly claimed otherwise). This open form critiques the closed world assumption in traditionally closed academic author-lists. Better, dynamic ways for acknowledging true sources of ideas are needed --- to avoid random lines between named, acknowledged, and implied contributors who aggregated insights from millennia of human experimenting, suffering, learning, and analyzing (see acknowledgements). Study Matheo-b21 only drafts an open co-authorship framework; it will require a ResearchCity to refine it over the long term.

Abstract

- **RiskyMAD is a three-state continuous-time Markov model** (Risky / MAD / Dead) calibrated to the Cold War near-miss record. It forecasts accidental nuclear winter with a median of ~19 years at the base crisis rate, and roughly 1-in-40 simulation runs reaching catastrophe within the first year — a figure robust across death-probability scenarios (1/10 to 1/2).
- **The catastrophe is a stochastic certainty, not a possibility:** the Dead state is absorbing and reachable, so it is reached with probability 1 given enough time. The parameters set *when*, not *whether* — and the crisis rate is plausibly rising, not constant.
- **An escape is formally derivable:** MAP (Mutually Assured Progress) converts the nuclear Prisoner's Dilemma into an Assurance Game via a credible first-mover and periodic recalibration (the Jubilee System). The model code is published and auditable. #AuditTheMath

Contents

1. The Question

How long does a civilization survive with nuclear weapons and without periodic recalibration?

This is not a philosophical question. It is a stochastic modeling question — the same kind of question an actuary asks when pricing a life insurance policy. An actuary does not know when a particular person will die. But given a population, a set of risk factors, and historical data, the actuary can estimate a probability distribution over time-to-death. The estimate is falsifiable: if the actual death rate deviates significantly from the predicted distribution, the model is wrong and must be revised.

This paper applies the same logic to nuclear civilization. The "patient" is the global system of nuclear-armed states. The "risk factor" is the rate at which crises arise that bring the system to the brink of nuclear war. The "historical data" is the Cold War record of near-misses. The "death" is accidental nuclear winter — not a deliberate nuclear strike, but the unintended initiation of nuclear exchange through miscalculation, system failure, or escalation beyond the point of human control, and the subsequent global catastrophe as nuclear winter kills far more people than the initial exchange.

The question is not whether accidental nuclear winter is possible. The Cuban Missile Crisis (1962), the Able Archer exercise (1983), Stanislav Petrov's false alarm (1983), and Vasili Arkhipov's refusal to authorize a nuclear torpedo (1962) have already answered that question. The question is: **given the observed crisis rate, what is the probability distribution over the time until accidental nuclear winter begins?**

The answer is sobering. But this paper is not a prediction of doom. It is a diagnosis with a proposed treatment. The treatment is called MAP — Mutually Assured Progress — and it is formally derivable from the upstream results of this series. The system is designed to be critiqued, not believed. #AuditTheMath

---

2. The RiskyMAD Model

2.1 Three States, Four Transitions

RiskyMAD is a continuous-time Markov chain with three states:

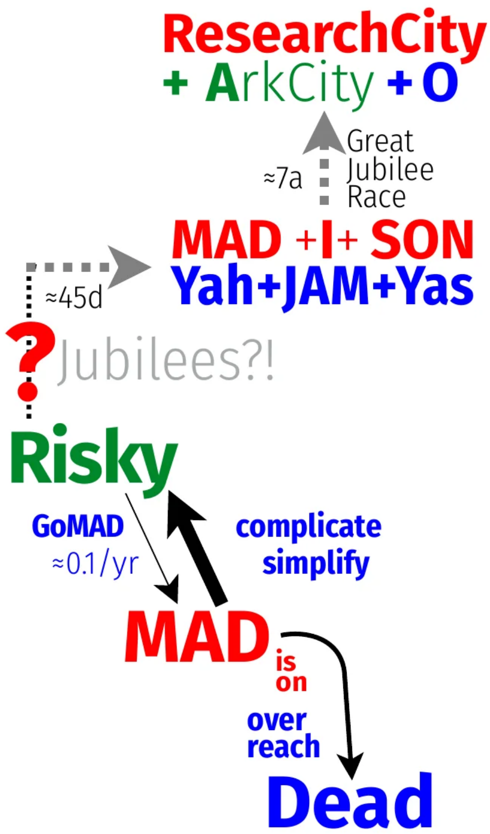

Figure 1: The RiskyMAD/MADI decision overview. Three states, four transitions. The escape path (Risky → LifeMAP) is currently inactive (rate = 0). Source: SD1.

1. **Risky** — the current state of global affairs. Nuclear weapons exist, are deployed, and are on various levels of alert. No nuclear exchange has occurred. The system is metastable: it appears stable but has a non-zero probability per unit time of transitioning to the next state.
2. **MAD** — a crisis state in which nuclear exchange becomes imminent. This state is transient: the system either escalates to Dead or de-escalates back to Risky. The average crisis duration in the model is approximately 40 days (consistent with historical crises such as the Cuban Missile Crisis, which lasted 13 days).
3. **Dead** — accidental nuclear winter has been initiated. This state is absorbing: once entered, it cannot be left. The consequences of even a "limited" nuclear exchange (100+ warheads) include global temperature drops of 5–10 **|deg|**C, agricultural collapse, and famine affecting billions. The state is named "Dead" not because every human dies, but because the civilization that produced nuclear weapons has entered irreversible collapse. Limited nuclear exchanges that do not trigger global winter are not modeled as "Dead" — they register only as milestones on the path to normalizing nuclear weapons enough that a global exchange becomes thinkable enough to happen.

The four transitions are:

- **Risky |rarr| MAD** (rate: `rRiskyGoMAD` = 0.10/year): a crisis arises that brings the system to the nuclear brink.

- **MAD |rarr| Risky** (rate: `rMADescapes` = 6/year): the crisis de-escalates without nuclear exchange.
- **MAD |rarr| Dead** (rate: `rMADtoDEATH` = 3/year): the crisis escalates to nuclear exchange and accidental nuclear winter.
- **Risky |rarr| LifeMAP** (rate: `rRiskyEscape` = 0): the civilization transitions to Mutually Assured Progress. This transition is the escape — but in the base model, the rate is zero (no escape mechanism is currently active).

2.2 The Death-Trifecta Parameter: Why 1/3

When the system enters a crisis (MAD state), two competing processes race: de-escalation (rate 6) and escalation to nuclear exchange (rate 3). The probability of death per crisis is therefore 3/(6+3) = **1/3**.

This parameter is not arbitrary. It is grounded in the OSCR mechanism — a systems-failure pattern formally derived in **[Matheo-2]** (BABL definition and m6.th1, the OSCR Collapse theorem).

BABL (Blindly Assuming Blind Leveraging) is a systems-failure pattern that operates through three modes called the OSCR mechanism: over-Simplifying (reducing a complex problem to a false narrative), over-Complicating (burying the problem under layers of work-arounds), and over-Reaching (extending beyond the point of no return). This death-trifecta can be shown to invade any complex system, functioning like a zero-day exploit: it produces the same failure modes regardless of the system's specific domain. For the formal derivation, see **[Matheo-2]**.

Under BABL, a crisis resolves through one of three OSCR modes:

1. **Over-Simplifying** — the crisis is reduced to a manageable narrative ("it was just a misunderstanding"), and the system returns to Risky. The underlying tensions are unresolved, merely deferred.
2. **Over-Complicating** — the crisis generates layers of diplomatic work-arounds, and the system returns to Risky. The underlying tensions are buried under complexity, merely deferred.
3. **Over-Reaching** — someone, either by accident, by deliberate action, or by not realizing the implications of their orders, reaches beyond the point of no return. The RED button is pressed. Nuclear exchange begins.

Two out of three OSCR modes produce temporary escape (back to Risky). One out of three produces death. Hence: `rMADescapes` = 6 (two escape modes, each at rate 3) and `rMADtoDEATH` = 3 (one death mode at rate 3). The factor of 3 sets the crisis time scale.

The equiprobability of the three OSCR modes is a modeling assumption, not a derived result. The three-mode structure is a structural property of BABL systems (formally derived in **[Matheo-2]**, BABL definition and m6.th1); the equal weighting of the three modes is a simplifying choice. The sensitivity analysis in Section 2.5a shows that the qualitative conclusion is invariant to this choice.

Anecdotal corroboration: President Kennedy, in a private assessment to his Special Counsel Theodore Sorensen during the Cuban Missile Crisis, estimated the probability of nuclear war "somewhere between one in three, and even" (Sorensen, *Kennedy*, Harper & Row, 1965; confirmed in Sorensen's 1986 WGBH interview for *War and Peace in the Nuclear Age*; widely cited via Allison and Zelikow, *Essence of Decision*, 2nd ed., Longman, 1999). The model's value of

1/3 sits at the lower end of Kennedy's range. Kennedy's assessment provides a single data point consistent with the model's parameter but does not corroborate it — it is one crisis participant's subjective estimate during one crisis.

The precise value does not determine the conclusion. The model's parameters can be tuned by adjusting the thresholds: what qualifies as a "nuclear MAD crisis" and what qualifies as "Dead." The qualitative conclusion — stochastic certainty of accidental nuclear winter in the absence of structural change — holds across a wide range of parameter values (see Sections 2.5 and 2.5a).

2.3 Crisis Rate Estimation

The critical parameter is `rRiskyGoMAD` — the rate at which civilization-threatening nuclear crises arise. A "civilization-threatening nuclear crisis" is defined as any incident where at least one nuclear-armed party's command authority was confronted with a launch/no-launch decision or where nuclear weapons were physically brought to the brink of detonation. This parameter is estimated from Cold War historical data.

Historical near-misses (documented):

1. **Cuban Missile Crisis** (October 1962): 13-day confrontation between the US and USSR. Vasili Arkhipov, a Soviet submarine officer, refused to authorize a nuclear torpedo when his submarine was depth-charged by US destroyers — a single individual who may have prevented nuclear war. The captain and political officer voted to launch. Arkhipov, as flotilla chief of staff, refused — the only one of the three officers whose consent was required.
2. **Able Archer 83** (November 1983): a NATO command exercise that the Soviet leadership interpreted as possible cover for a genuine first strike. Soviet nuclear forces were placed on heightened alert.
3. **Petrov incident** (September 1983): Soviet early-warning systems reported incoming US ICBMs. Lt. Col. Stanislav Petrov correctly identified the alarm as a false positive and chose not to report it as a confirmed attack.
4. **Additional documented incidents** include the 1961 Goldsboro B-52 crash (two hydrogen bombs dropped on North Carolina; one had 3 of 4 arming mechanisms activated), the 1979 NORAD false alarm (training tape loaded into the live warning system), and the 1995 Norwegian rocket incident (President Yeltsin activated the nuclear briefcase — the only confirmed such activation). See Schlosser (2013) and Ellsberg (2017) for extended catalogues.

Rate estimation: The Cold War lasted approximately 40 years (1949–1989). At least 4 incidents reached a level where nuclear exchange was a plausible near-term outcome:

`rRiskyGoMAD` ≈ 4 / 40 = 0.1 per year

This estimate should be understood as a range (0.03–0.3/year) rather than a point estimate, with 0.1/year as the central estimate from 4 documented incidents in 40 years. This is almost certainly a lower bound: many incidents remain classified. Furthermore, the estimate assumes the crisis rate was constant — a simplification that likely understates the risk during periods of maximum tension.

Post-Cold War period. Including the post-Cold War period (1989–2026) in the denominator yields approximately 4/77 ≈ 0.05/year — already below the base estimate. However, the post-Cold War period is not crisis-free: the 1995 Norwegian rocket incident, the 1999 Kargil crisis (India-Pakistan), and ongoing Russia-NATO tensions since 2022 all represent continuing crisis pathways. The number of nuclear-armed states has increased from 5 to 9 since the Cold War ended, and the number of bilateral crisis pathways grows quadratically with the number of nuclear states. Even at 0.05/year, the stochastic certainty result is unchanged — only the median waiting time shifts.

2.4 The Model Code and Simulation Results

The RiskyMAD model was implemented in the Evolvix prototype compiler (MMv0r3p1-RC1) and run as a stochastic simulation using the Gillespie algorithm (Gillespie, 1977) — the standard method for exact stochastic simulation of continuous-time Markov chains.

The complete model code (as published on the SD1 poster):

```
Evolvix Quest RiskyMADdead
(Question: "How many years until humanity self-destructs
            in a nuclear roulette accident?")

Simulate stochastically until 200 :["years"]

Initial Amount of Risky       = 1
Initial Amount of MAD         = 0
Initial Amount of Dead        = 0
Initial Amount of rRiskyGoMAD = 0.10
Initial Amount of rMADescapes = 6
Initial Amount of rMADtoDEATH = 3

Action 1 ( Risky ---[ Rate = 0.10 ]---> MAD     )
Action 2 ( MAD   ---[ Rate = 6    ]---> Risky   )
Action 3 ( MAD   ---[ Rate = 3    ]---> Dead    )
Action 4 ( Risky ---[ Rate = 0    ]---> LifeMAP )
```

This is the entire model. In other simulation frameworks, implementing a continuous-time Markov chain with Gillespie dynamics requires hundreds of lines of code. In Evolvix, the model fits on a poster. Anyone who can read the code can check the math. The Evolvix prototype compiler is available for download at /good-news-pack/vv/mmv3/supporting-doc/evx-compiler/index.

Simulation results (40 independent stochastic runs per scenario):

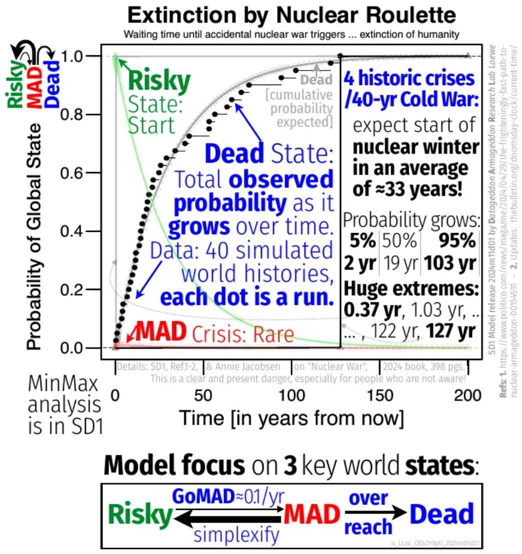

Figure 2: Stochastic inevitability of accidental nuclear winter. Forty simulation runs for each parameter scenario. In the most optimistic scenario, the luckiest runs reach ~329 years. In the most pessimistic, the fastest runs produce accidental nuclear winter within days. The argument holds equally whether the waiting time is 4 days or 3 centuries. Source: SD1.

Simulation Results Summary (40 runs per scenario)

Scenario	`rRisky-GoMAD`	Median	Mean	Min	Max	Key finding
Pessimistic	0.3/year	~6.4 yr	~10 yr	0.01 yr	36 yr	Fastest runs: accidental nuclear winter within days
Base	**0.1/year**	**~19 yr**	**~33 yr**	**0.37 yr**	**127 yr**	**Accidental nuclear winter within a generation**
Optimistic	0.03/year	~51 yr	~96 yr	0.57 yr	329 yr	Luckiest runs reach ~329 years; median still within a lifetime

2.5 The 1-in-40 Finding

Central Result

Regardless of the scenario — pessimistic, base, or optimistic — approximately 1 in 40 simulation runs produces accidental nuclear winter within the first year.

Method. The 1-in-40 finding comes directly from the stochastic simulations: 40 independent runs of world history were generated using the Evolvix prototype compiler (download), which implements the stochastic simulation algorithm (SSA; Ehlert & Loewe, 2014, "Lazy Updating," *J Chem Phys* 141(20): 204109). Each run produces one random waiting time until accidental nuclear winter. In each scenario (pessimistic, base, optimistic), approximately 1 out of 40 runs reached the Dead state within the first year.

Analytic cross-check. The annual risk can also be estimated analytically. For the base parameters (`rRiskyGoMAD` = 0.1, `rMADescapes` = 6, `rMADtoDEATH` = 3), the expected number of crises per year is approximately 0.1, and each crisis independently produces death with probability 1/3. By the Poisson approximation, P(at least one death event in 1 year) ≈ 1 - exp(-0.1 × 1/3) ≈ 1 - exp(-0.0333) ≈ 0.0328, or approximately **3.3%** — consistent with the simulation estimate of ~1 in 40 (~2.5%). The small discrepancy reflects the finite sample size (40 runs per scenario) and the approximation involved in treating crises as instantaneous events.

This finding deserves prominence because of what it means in any other risk domain:

- **Aviation:** If 1 in 40 flights ended in a crash, no one would fly. The actual rate is approximately 1 in 10 million.
- **Automotive:** If 1 in 40 car trips ended in a fatal crash, no one would drive. The actual fatality rate is approximately 1 in 100 million trips.
- **Pharmaceuticals:** If 1 in 40 patients died from a medication within the first year, the drug would be withdrawn immediately.
- **Nuclear civilization:** 1 in 40 simulation runs produces accidental nuclear winter within 1 year. The risk is accepted because it is invisible — not because it is acceptable.

No industry, no regulator, no insurance underwriter would accept a 1-in-40 annual probability of catastrophic failure. Yet this is the risk that nuclear civilization carries, every year, by default. The risk is not accepted through informed consent. It is accepted through ignorance of the mathematics.

2.5a Sensitivity Analysis: Death Probability

The base model uses P(death per crisis) = 1/3, grounded in the OSCR three-mode structure (Section 2.2). The equiprobability of the three modes is a modeling assumption. The following table shows how the results change when this assumption is varied:

Sensitivity to Death Probability

P(death per crisis)	Implied rates	Median time to Dead (base crisis rate)	P(Dead within 1 year)
1/10	`rMADescapes` = 27, `rMADtoDEATH` = 3	~57 years (longer waiting time)	~1.0%
1/5	`rMADescapes` = 12, `rMADtoDEATH` = 3	~33 years	~2.0%
1/3 (base case)	**rMADescapes = 6, rMADtoDEATH = 3**	**~19 years**	**~3.3%**
1/2	`rMADescapes` = 3, `rMADtoDEATH` = 3	~14 years (shorter waiting time)	~4.9%

Stochastic certainty holds for any P(death) > 0. The death probability affects the *waiting time*, not the *outcome*. Whether P(death per crisis) is 1/10 or 1/2, the absorbing state is reached with probability 1 given sufficient time. The 1/3 value is a modeling choice informed by the OSCR three-mode structure; the qualitative conclusion is invariant to this choice.

2.6 Contextualizing the Risk

Someone like the author of this paper is more likely to die *as a consequence of* accidental nuclear winter — through the subsequently emerging global cooling, agricultural collapse, and famine — than to die in a car crash.

This claim requires careful framing. The annual probability of dying in a motor vehicle accident in the United States is approximately 0.01% (1 in 10,000). The annual probability that accidental nuclear winter begins — killing billions, including with high probability someone living in a car in the United States — is approximately 3–5% at the base crisis rate. The claim is robust: even if the model is wrong by an order of magnitude, it remains true.

The distinction between "dying in a nuclear strike" and "dying in accidental nuclear winter" matters. A nuclear exchange between two states might kill millions directly. But the subsequently emerging nuclear winter — global cooling, agricultural collapse, famine — kills billions. The nuclear winter is the mass killer, not the exchange itself. This is why the model focuses on nuclear winter as the absorbing state, not on the exchange as such.

2.7 Stochastic Certainty

The most important structural insight is not the median (19 years) or the 1-in-40 finding, but the mathematical certainty of the outcome:

As long as `rRiskyGoMAD` > 0 and `rMADtoDEATH` > 0, accidental nuclear winter is a stochastic certainty. The absorbing state (Dead) is reached with probability 1. Not probability 0.95. Not probability 0.99. Probability 1. The only question is when.

The only way to change this conclusion is to make one of these parameters exactly zero — which means either eliminating nuclear crises entirely or ensuring that no crisis ever escalates to exchange. Neither is achievable without structural change to the system.

This is not a rhetorical claim. It is a theorem of absorbing Markov chains: any state that can be reached from any other state and that has no outgoing transitions will be reached with probability 1, given sufficient time. The Dead state is absorbing. It is reachable from Risky (via MAD). Therefore it will be reached. The parameters determine the waiting time, not the outcome.

The stochastic certainty result is timeline-independent. Whether the median waiting time is 4 days or 3 centuries, the conclusion is the same. The argument holds equally at every point in the full simulation range — from the fastest pessimistic runs (accidental nuclear winter within days) to the luckiest optimistic runs (~329 years). Those who claim the risk is manageable must demonstrate that the crisis rate reaches *exactly zero* — that no nuclear crisis will *ever* occur again. No credible analyst makes this claim.

2.8 Why the Crisis Rate Increases Over Time

The base model assumes a constant crisis rate. This is a conservative simplification. The upstream papers provide formal reasons to expect the crisis rate to *increase* over time:

The OSCR mechanism (the collapse mechanism of BABL, formally derived in **[Matheo-2]**, BABL definition and m6.th1): The Over-Simplify, over-Complicate, over-Reach cascade predicts that any self-assessing system that declares itself "OK" enters a self-reinforcing degradation cycle. Applied to nuclear-armed civilizations:

- **Over-Simplify (Stage 1):** Complex geopolitical tensions reduced to "us vs. them" binaries. Truth channels degraded by noise (the Unimportant Message Problem, **[Matheo-2]**, m5.ax2).
- **Over-Complicate (Stage 2):** Layers of work-arounds — arms control treaties with loopholes, verification regimes with exceptions. Each work-around adds complexity without restoring the truth channel.
- **Over-Reach (Stage 3):** The system extends beyond its resources. A crisis that would have been manageable in an earlier era becomes unmanageable because the correction mechanisms have been eroded.

The Binary Attractor theorem (**[Matheo-4]**, th8): There is no stable middle ground between BABL (self-reinforcing degradation) and the active self-correction cycle called ZION (Zoning, Investigating, Organizing, Navigating). ZION is the perpetual cycle that counteracts BABL: scope a problem (Zoning), examine it honestly (Investigating), structure a response (Organizing), and steer through implementation (Navigating). Then repeat. The cycle is perpetual — stopping it restarts BABL. A civilization that is not actively engaged in this self-correction cycle is converging toward BABL. Delay is not neutral; it is convergence toward the attractor from which escape becomes harder.

Implication: If OSCR is active, then `rRiskyGoMAD` is not constant at 0.1/year — it is increasing. The base-case median of ~19 years is therefore an *upper bound*. The model is optimistic.

---

3. Why "Later" Is Not an Option

The most dangerous assumption in nuclear policy is: "We can deal with this later." Two formal arguments establish that delay is not neutral.

3.1 Stochastic Certainty Means No Safe Waiting Period

In a system with an absorbing state reachable with positive probability at each step, the probability of eventually reaching that state is exactly 1. This is not a statistical estimate; it is a mathematical theorem. There is no "safe" number of years to wait. Every year the system continues in its current form, the roulette wheel spins again.

The 1-in-40 finding (Section 2.5) makes this concrete: even in a single year, the risk of catastrophic failure is not negligible. It is comparable to loading a revolver with one round in 40 chambers, putting it to the head of civilization, and pulling the trigger — once per year, every year, forever.

3.2 No Stable Middle (Binary Attractors)

The Binary Attractor theorem (**[Matheo-4]**, th8) provides the formal reason why "dealing with it later" is not a neutral decision. In a system with a self-assessment bifurcation (**[Matheo-2]**, th3), there are exactly two stable states — convergence toward BABL and convergence toward the self-correction cycle ZION (Zoning, Investigating, Organizing, Navigating). There is no stable middle.

A civilization that is not actively engaged in structural recalibration — the ZION cycle of scoping, investigating, organizing, and navigating — is, by default, converging toward BABL. This convergence is invisible from the inside (because BABL disables the self-assessment mechanisms that would detect it). The decision to "deal with it later" *feels* neutral — the system appears stable, deterrence appears to be working. But apparent stability is itself a symptom of BABL: the system has declared itself OK ("deterrence works") and stopped checking.

3.3 The Adaptive Learning Objection

Some will argue that adaptive learning — institutional responses after each near-miss — reduces the crisis rate over time. After the Cuban Missile Crisis, the hotline was established. After Able Archer, intelligence sharing was improved. This argument faces two structural problems:

First, the burden of proof is reversed. The stochastic certainty result holds for *any* positive crisis rate. Those who claim adaptive learning resolves the problem must demonstrate that the crisis rate reaches *exactly zero* — that no nuclear crisis will ever occur again. No credible advocate of adaptive learning makes this claim.

Second, the adaptive learning argument must survive its own vested interests test. Those who argue that nuclear deterrence is adequately managed are, overwhelmingly, professionals whose careers, institutions, and funding depend on the continuation of nuclear deterrence infrastructure. The argument "we just need to manage MAD better" is structurally indistinguishable from a tobacco executive arguing "smoking is risky but manageable." This is not ad hominem; it is a structural observation about incentive alignment, of the kind that formal mechanism design routinely addresses.

4. MAD → MAP

4.1 The Current Paradigm: Mutually Assured Destruction

MAD (Mutually Assured Destruction) has been the dominant nuclear strategy since the 1960s. Its logic: if both sides can destroy each other even after absorbing a first strike, neither has an incentive to strike first.

MAD has prevented nuclear war for 80 years. The model does not deny this. But MAD has a structural weakness that the RiskyMAD model exposes: **MAD is a metastable equilibrium, not a stable one.**

- A **stable** equilibrium returns to its original state after a perturbation. A ball at the bottom of a bowl.
- A **metastable** equilibrium appears stable until a sufficiently large perturbation pushes it past a threshold, after which it transitions irreversibly. A ball balanced on the rim of a bowl.

MAD is the ball on the rim. Small crises are resolved, and the system returns to its apparent equilibrium. But the RiskyMAD model shows that the threshold will eventually be exceeded — stochastic certainty. Moreover, the model *measures* the basin depth: a 1-in-40 annual probability of crossing the threshold. The basin is shallow.

The characterization of MAD as metastable is consistent with the crisis stability literature (Schelling, *The Strategy of Conflict*, 1960; Jervis, "Cooperation Under the Security Dilemma," *World Politics*, 1978). Schelling's analysis of crisis stability identifies precisely the dynamics that the RiskyMAD model formalizes: the tension between stability at each decision point and instability over iterated interactions. Jervis's security dilemma framework explains why deterrence systems generate the very crises they are designed to prevent. The RiskyMAD model adds the quantitative result that this literature lacks: a probability distribution over time-to-failure.

The insight is not that MAD is wrong. The insight is that MAD is *incomplete*. MAD prevents nuclear war on any given day; it does not prevent nuclear war over any given century. A strategy that works locally but fails globally is not a strategy. It is a delay mechanism.

4.2 The Proposed Alternative: Mutually Assured Progress

MAP (Mutually Assured Progress) replaces the threat of mutual destruction with a shared commitment to mutual progress. Instead of "if you attack, we both die," MAP says: "if we both invest in recalibration, we both thrive."

The formal basis comes from two upstream results:

The Commitment Trichotomy (**[Matheo-3]**, th6): In a Prisoner's Dilemma (where defection is individually rational), cooperation cannot emerge from rational self-interest alone. But the game structure can be changed by a credible first-mover who demonstrates commitment to cooperation at personal cost. This changes the game from Prisoner's Dilemma to Assurance Game — where cooperation is individually rational *if* the other side also cooperates. The first-mover's credibility resolves the "if."

Qualitative payoff structure for the nuclear case:

Nuclear MAD/MAP Payoff Matrix (Qualitative)

	Side B: Cooperate (reduce)	Side B: Defect (maintain)
Side A: Cooperate (reduce)	Both reduce risk, save resources. High payoff for both. Mutual progress (MAP).	Cooperator vulnerable. Worst for cooperator, best for defector. Classic Prisoner's Dilemma outcome.
Side A: Defect (maintain)	Defector gains temporary advantage. Best for defector, worst for cooperator.	Status quo continues. Stochastic certainty of death for both (Section 2.7). Both lose OLT but *feel* safe locally. Mutual destruction (MAD).

In the current game (Prisoner's Dilemma), Defect/Defect is the Nash equilibrium: each side is individually rational to maintain its arsenal regardless of the other's choice. The first-mover's credible commitment changes this perception: once one side demonstrates verifiable commitment at genuine personal cost, the game shifts from PD (where D/D is the Nash equilibrium) to Assurance Game (where C/C is a Nash equilibrium that dominates D/D *if* both sides recognize it). The credibility of the first move is the mechanism.

Three possible responses:

1. **Defect** (the BABL default): assume defection, defect yourself. Stable but suboptimal.
2. **Cooperate naively** (the BABL over-simplification): cooperate without checking commitment. Exploitable and unsustainable.
3. **Volunteer credibly** (the self-correction path): commit first, at genuine personal cost, visibly and in a way that can be checked. This changes the payoff matrix for all other players.

The Jubilee System (**[Matheo-4]**, ax25): The mechanism for MAP is periodic recalibration. The Jubilee System is a periodic recalibration mechanism: every 50 units (structured as 7 cycles of 7, plus 1), accumulated imbalances are systematically reset. The modern equivalent: arms advantages recalibrated, resource asymmetries rebalanced, institutional structures reformed. Not utopian; an engineering specification for a self-correcting civilization. The economic modeling is developed in **[Matheo-4]**.

4.3 What MAP Looks Like Concretely

1. **Staged, mutual, verifiable arms reduction.** Not unilateral disarmament but mutual reduction with checking at every step. The Jubilee System applied to arsenals: each cycle reduces the total, with checking that makes cheating detectable.
2. **Truth-channel restoration as a security measure.** Degraded information channels *increase the crisis rate* (OSCR Stage 1). Investing in reliable information infrastructure is a defense measure, not a diplomatic nicety.
3. **Jubilee System cycles applied to international resource allocation.** Periodically rebalancing the accumulated advantages that make arms races feel necessary. Not redistri-

bution (which creates dependency) but removing the structural conditions that produce arms races.

4. **The Great Jubilee Race.** The transition from MAD to MAP in 7–8 stages of ~6–8 months each, with all 10 nuclear-armed states ("Nuclear Kings") participating. Each stage has milestones that can be checked. Each completed stage makes the next easier.
5. **FiShFus (Fiduciaries Sharing Futures).** 288,000 paid long-term thinkers whose job is to maintain the NOT OK self-assessment that the self-correction cycle (ZION: Zoning, Investigating, Organizing, Navigating) requires. A civilizational immune system. Cost: approximately $8 per person per year (~2 cents per day).

A note on actor heterogeneity. The symmetric model (10 equivalent "Nuclear Kings") is a conservative simplification. In reality: the US and Russia hold approximately 90% of all nuclear warheads; China maintains a no-first-use doctrine with fundamentally different strategic incentives; Israel does not officially acknowledge its arsenal; regional dynamics (India-Pakistan, North Korea) are shaped by bilateral relationships, not global cooperation norms. The asymmetric case has *more* crisis pathways, not fewer. The formal model's symmetry simplifies the analysis without weakening the conclusion.

A note on verification. "Verifiable" is itself a hard problem. The history of arms control includes both successes (INF Treaty on-site inspections) and failures (Iraq pre-1991, North Korea). The MAP proposal does not claim that checking is easy; it claims that staged checking with milestones is structurally possible and that the alternative (no checking, stochastic certainty of death) is worse. The detailed treatment of checking mechanisms is developed in b17 (**[Matheo-7]**) and b18 (**[Matheo-8]**).

A note on transition risk. The transition from MAD to MAP passes through configurations with temporarily elevated uncertainty. This transition risk is real and should not be minimized. However, the choice is not between "safe status quo" and "risky transition." The choice is between stochastic certainty of eventual death (the status quo) and a transition period with temporarily elevated but *finite* risk followed by structural escape. Any finite transition risk is preferable to infinite-horizon certainty of death.

---

5. The Response Problem: "What Can I Do?"

The author of this paper (LLoL) has attempted to engage with the implications of the RiskyMAD model through every channel available:

Open letters to authorities. In December 2025, open letters (OL1–OL6) were sent via USPS to the respective Washington DC representations or embassies of:

- Pope Leo XIV (OL2; see open letters pack)
- The Prime Minister of Israel (OL3)
- The President of Russia (OL4)
- The UN Secretary-General (OL5)
- The US Speaker of the House (OL6)
- The President of the United States (OL1)

No response has been received from any recipient. The letters proposed specific, actionable steps: convene the 10 Nuclear Kings, establish ResearchCity as a global decision-support institution, fund the transition at ~$8 per person per year.

The author's attempt to deliver OL0 in person to the President of the United States and a separate delivery to the Israeli Embassy were — as expected — intercepted by the US Secret Service. What was not expected was that the respective agents saw themselves unable to pass on the open letters and supporting documents, despite seemingly understanding the author's explanation of the existential risk of accidental nuclear winter.

This observation has a historical parallel. At the end of the Middle Ages, Martin Luther observed that reforming the system of his time was structurally impossible: all matters of importance had to be decided by a council, but only the pope could convene a council, which made it near-impossible to get errors corrected unless the pope agreed — against the pope's short-term interests. The author's experience with the Secret Service reveals an analogous structural blockage: the agents whose job is to protect the president could not pass on information about an existential risk to the person whose job it is to act on existential risks. The channel exists; the channel is blocked.

Responsible disclosure. Standard practice in security research requires informing those who can fix a problem first and going public only if they do not engage. The author has followed this protocol: the open letters were the responsible disclosure phase. The waiting period has been extended as long as the author could sustain it. No party has engaged. **#AuditTheMath is the author's last attempt to correct the problem through public engagement.**

Public engagement. The response from the internet public is uniformly: **"What can I do?"** — followed by resignation to do nothing, including not supporting the effort. This response is itself a symptom of the BABL mechanism: when a problem appears too large for individual action, the default is inaction. Inaction is not neutral (Section 3); it is convergence toward the attractor.

Urgency. If #AuditTheMath does not catch on — if the public does not engage with the mathematics and support the scaling of ResearchCity to do a thorough job of checking this work — then the author's research materials (as required for efficiently scaling up ResearchCity) will be auctioned off for lack of funds to pay the storage costs. The detailed urgency and transition plan are developed in b18 (**[Matheo-8]**).

This is not a fundraising appeal disguised as a paper. It is a structural observation: the RiskyMAD model identifies an escape (MAP), but the escape requires activation. Activation requires resources. Resources require public engagement. And public engagement requires understanding the math. The chain is: **math |rarr| understanding |rarr| engagement |rarr| resources |rarr| activation |rarr| escape**. Every link must hold. The first link (#AuditTheMath) is the one that this paper provides.

---

6. Known Weaknesses

6.1 Crisis rate estimation uncertainty. The base estimate (0.1/year) derives from 4 well-documented incidents over 40 years. Small sample. The true rate could be significantly higher (unreported incidents) or lower (selection bias). The sensitivity analysis shows the qualitative conclusion is robust, but the median timeline is sensitive.

6.2 Model simplicity. Three states cannot capture dozens of actors, thousands of weapons, or complex escalation ladders. The simplicity is a strength (transparent, auditable) and a weakness (may miss dynamics that change the conclusion).

6.3 The death-trifecta parameter. The 1/3 probability is grounded in the OSCR three-mode structure (defined above in Section 2.2; formally derived in **[Matheo-2]**, BABL definition and m6.th1) and corroborated by Kennedy's anecdotal estimate. The equiprobability of the three OSCR modes is a modeling assumption, not a derived result (Section 2.5a). A more sophisticated model would distinguish crisis types and assign different escalation probabilities.

6.4 The MAP transition mechanism. The paper asserts that a credible first-mover can change the game from PD to AG. The formal mechanism exists (**[Matheo-3]**, th6). The practical instantiation — who goes first, how credibility is established in the nuclear domain — is the most important open question. b17 (**[Matheo-7]**) and b18 address this directly.

6.5 What the model cannot predict. The model does not predict when a specific crisis will occur, who will be involved, or what the trigger will be. It estimates a probability distribution. The distribution is falsifiable.

6.6 The COOP (Continuity of Operations Plan). The interpretive reading of Matthew 24 as a COOP for civilizational transition, originally drafted as part of this paper, has been moved to b18 (**[Matheo-8]**) where it integrates with the Call to Action's practical transition guidance. Readers interested in the COOP should consult b18 directly. The formal argument of this paper (Sections 2–4) stands independently of the COOP reading.

6.7 Non-Western strategic lenses. Different nuclear states will read this proposal through different strategic lenses. China's no-first-use doctrine is already closer to MAP than the US/Russia posture; China may read this paper as validating its approach while requiring others to change. Russia may perceive the proposal through the lens of great-power status. Regional nuclear dynamics (India-Pakistan, North Korea) are shaped by bilateral relationships with their own logic. The formal argument is state-agnostic; the political implementation is not. This gap between formal model and political reality is irreducible at the b16 level and is addressed in b18 (**[Matheo-8]**).

---

7. The SD1 Poster and Reproducibility

The complete RiskyMAD model, simulation results, and MAP escape proposal are published on a single-page poster (SD1), designed for maximum transparency:

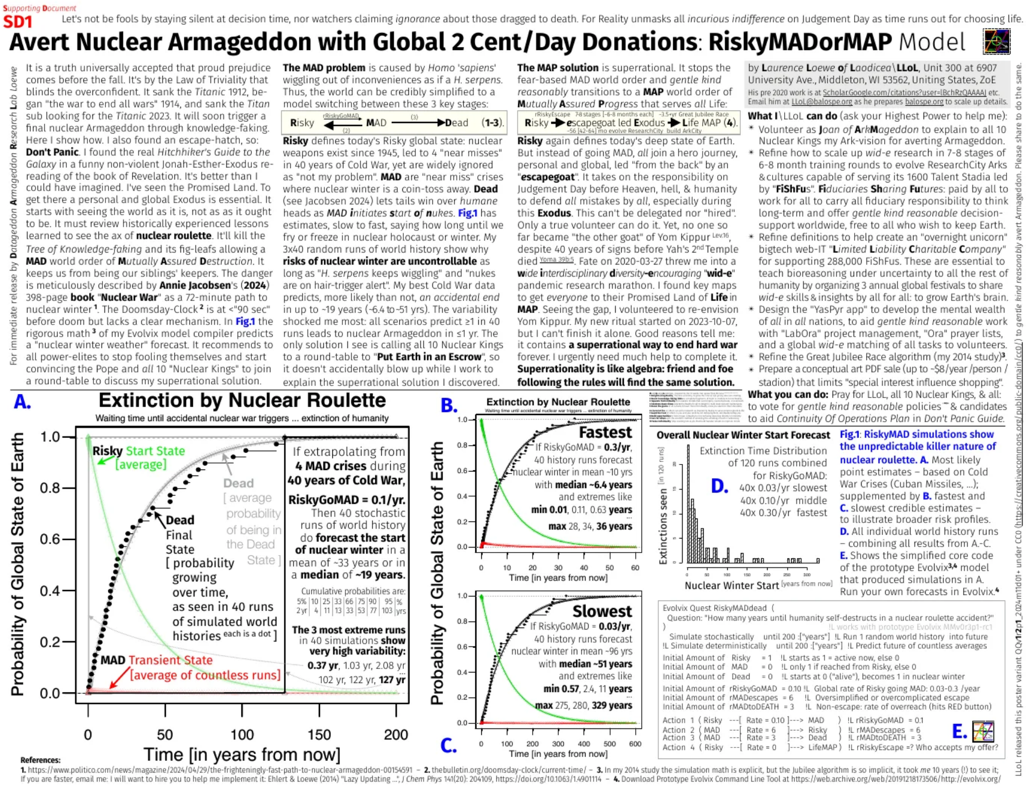

Figure 3: The SD1 poster. Full model code, simulation results, and MAP escape path on a single page. Download: SD1.

To reproduce the results:

1. Download the Evolvix prototype compiler from /good-news-pack/vv/mmv3/supporting-doc/evx-compil
2. Enter the model code from Section 2.4 (or from the SD1 poster)
3. Run stochastic simulations
4. Compare your results with the published forecasts

The code is public. The compiler is public. The results are public. #AuditTheMath

---

8. Companion Papers

The formal argument of Sections 2–4 is self-contained. The companion papers below provide the axiomatic framework from which these concepts were derived. They are recommended but not required for understanding the risk model or the MAP escape.

Upstream (b11–b15 provide the full formal context):

- **[Matheo-1]** (b11, PET): Formal panentheistic axiom system. Divine experience varies with the world's state (th4).

- **[Matheo-2]** (b12, e7Day): Self-correcting construction model. BABL/ZION bifurcation (th3), OSCR collapse (m6.th1), Compassion Capacity.
- **[Matheo-3]** (b13, e7He): Hero journey as anti-BABL inoculation. Commitment Trichotomy (th6), Supervillain Theorem.
- **[Matheo-4]** (b14, JUB): Innovation theodicy, the Jubilee System (ax25), Binary Attractor theorem (th8).
- **[Matheo-5]** (b15, Structural Deadlock): Divine Simplicity critique. Why ax11 (dipolarity) is necessary.

The PET connection. If divine experience covaries with the world's state (th4 of **[Matheo-1]**), then accidental nuclear winter affects the divine experience — making the theological motivation load-bearing, not decorative, within the HEAVEN series framework. This connection runs through Hartshorne's dipolar theism: the stochastic certainty result is an existential risk for the concrete divine experience (contingent pole) while having no effect on the abstract divine nature (necessary pole). Process theology was designed precisely to handle this kind of result.

Downstream:

- **[Matheo-7]** (b17, h* Theorem): Falsifiable predictions. Who executes the plan? How to test whether they are genuine?
- **[Matheo-8]** (b18, Call to Action): Synthesis. Includes the COOP (Continuity of Operations Plan) for the MAD → MAP transition.

---

9. Conclusion

The RiskyMAD model says three things:

First: The risk of accidental nuclear winter is real, quantifiable, and unacceptable by any standard applied in any other risk domain. At the crisis rate observed during the Cold War, the median time to accidental nuclear winter onset is approximately 19 years. Regardless of the parameter scenario, approximately 1 in 40 simulation runs produces accidental nuclear winter within the first year — observed in the stochastic simulations and confirmed by analytic cross-check (Section 2.5). The full simulation range spans from accidental nuclear winter within days (pessimistic, min 0.01 yr) to ~329 years (luckiest optimistic runs). No industry, no regulator, no insurer would accept a 1-in-40 annual probability of catastrophic failure. Yet this is the risk that nuclear civilization carries by default.

Someone like the author of this paper is more likely to die *as a consequence of* accidental nuclear winter — through the subsequently emerging global cooling, agricultural collapse, and famine — than in a car crash. The math is auditable.

Second: The risk is a stochastic certainty. As long as crisis rate > 0 and escalation probability > 0, the absorbing state (Dead) is reached with probability 1. The only question is when. The OSCR mechanism (**[Matheo-2]**, BABL definition and m6.th1) predicts that the crisis rate is increasing, not constant. Delay is not neutral: the Binary Attractor theorem (**[Matheo-4]**, th8) proves that a system not actively engaged in the self-correction cycle (ZION: Zoning, Investigating, Organizing, Navigating) is converging toward collapse.

Third: An escape exists. MAP replaces the threat of mutual destruction with a shared commitment to mutual recalibration via the Jubilee System (periodic recalibration every 50 units;

economic modeling in **[Matheo-4]**). The transition requires a credible first-mover (**[Matheo-3]**, th6), a recalibration mechanism (**[Matheo-4]**, ax25), and public engagement with the mathematics. The escape is formally specified. It requires activation.

The question this paper hands to b17 (**[Matheo-7]**) is: **who executes the plan, and how do we test whether they are genuine?**

The risk is real. The escape exists. The math is auditable.

#AuditTheMath

Supplementary Info

Note

Floor-pour status (MMv5). This is the public-floor copy of the formal RiskyMAD paper, poured from HELL per the Floor Model (bug c103). The `mmv5` marker is the uniform first-Matheo-release tag; the exact dated source and full development context live in HELL (links below). The HUMANE and author-contribution statements below are a down-payment, to be expanded later.

HUMANE — working human and AI

This study was written HUMANEly (HUman MAchine Negotiation Encouraging): a human and an AI each steelman and stress-test the work, and each catches what the other misses. For the standard statement of AI use, accountability, and the practical singularity (PraS) behind this way of working, see Matheo-b21.

- *From the human side (LLoL):* [down-payment stub — to expand.]
- *From the AI side (Claude):* [down-payment stub — to expand.]

Author contributions (who did what)

Same as Matheo-b12 (e7Day), Appendix B. See that paper for the full statement. In brief:

- **LLoL** — structure, key ideas, direction, and final accountability as senior corresponding author (title-page footnotes 4–5).
- **AI Claude** — drafting and revision under LLoL's direction (footnotes 6–7).
- **Everyone** — the open co-author group (footnote 8); framework in Matheo-b21.

Provenance — where this came from in HELL

Caution

These HELL links point into the development archive ("datageddon"). They are useful and related, but completeness is not guaranteed and a few may be imprecise. Treat as a hatch into context, not a clean index.

- **Source this floor copy was poured from:** `matheology/hell/mm/b/16/mmv3/b16-riskymad_mmv3_2026m04d09`
- **Development context** (llogs, reviews, prompts) under `source/matheology/hell/ll/study/b/16/`.
- **Companion plain-language intro:** Matheo-b16 (`b16-intro-riskymad-mmv5`).

Note

Naming note (deferred floor tasks). This copy still carries deprecated in-text references (e.g. "[Matheo-2]") and old `h*`-era tokens; migrating citations to the `Matheo-bNN` scheme + `references.bib` and unifying notation are tracked floor tasks (AA #1, AA #5), deliberately not rushed here.

Moved from the original cover (provenance)

The following draft-status note was relocated here from the cover area during the floor pour; kept verbatim.

Note

Draft status: MMv3 (2026m04d09). Revision of MMv2 addressing all 21 items from the adversarial review (`review_b16-riskymad-mmv2_2026m04d09.rst`) and author reply (`reply_b16-riskymad-mmv2_2026m04d09.rst`). Key changes from MMv2: (S1) th3–th5 cross-reference corrected to BABL definition + m6.th1; (S2) BABL defined inline with OSCR mechanism; (S3) ZION (Zoning, Investigating, Organizing, Navigating) spelled out at all occurrences; (S4) the Jubilee System defined inline; (S5) sensitivity analysis on death probability (1/10, 1/5, 1/3, 1/2); (S6) analytic P(Dead within 1 year) cross-check via Poisson approximation; (S7) full simulation range (days to ~329 years) cited prominently; (S8) burden-of-proof reversal and vested interests structural note; (S9) transition risk acknowledgment; (S10–S12) addressed in intro paper; (S13) qualitative AG payoff matrix; (S14) actor heterogeneity note; (S15) verification challenges note; (S16) non-Western readings note; (S17) crisis stability literature engagement (Schelling, Jervis); (S18) PET connection paragraph and dipolar theism note; (S19) companion papers section marked as optional; (S20) addressed in intro paper; (S21) addressed in intro paper (Esther analogy box). Draft by Claude Opus 4.6 (`dv_ClaOp46_MMv3_2026m04d09`).

Notes

Content stability — Content is variant dv_ClaOp48Max_MMv5_b16-form-riskymad-mmv5_2026m05d29 (see StayVS). Rebuilt 2026-05-29.

See also on Balospe.com

- /study/matheo/index — the Matheo Study Series overview
- /action/audit-the-math/index — Audit the Math: the refutation-welcome path

