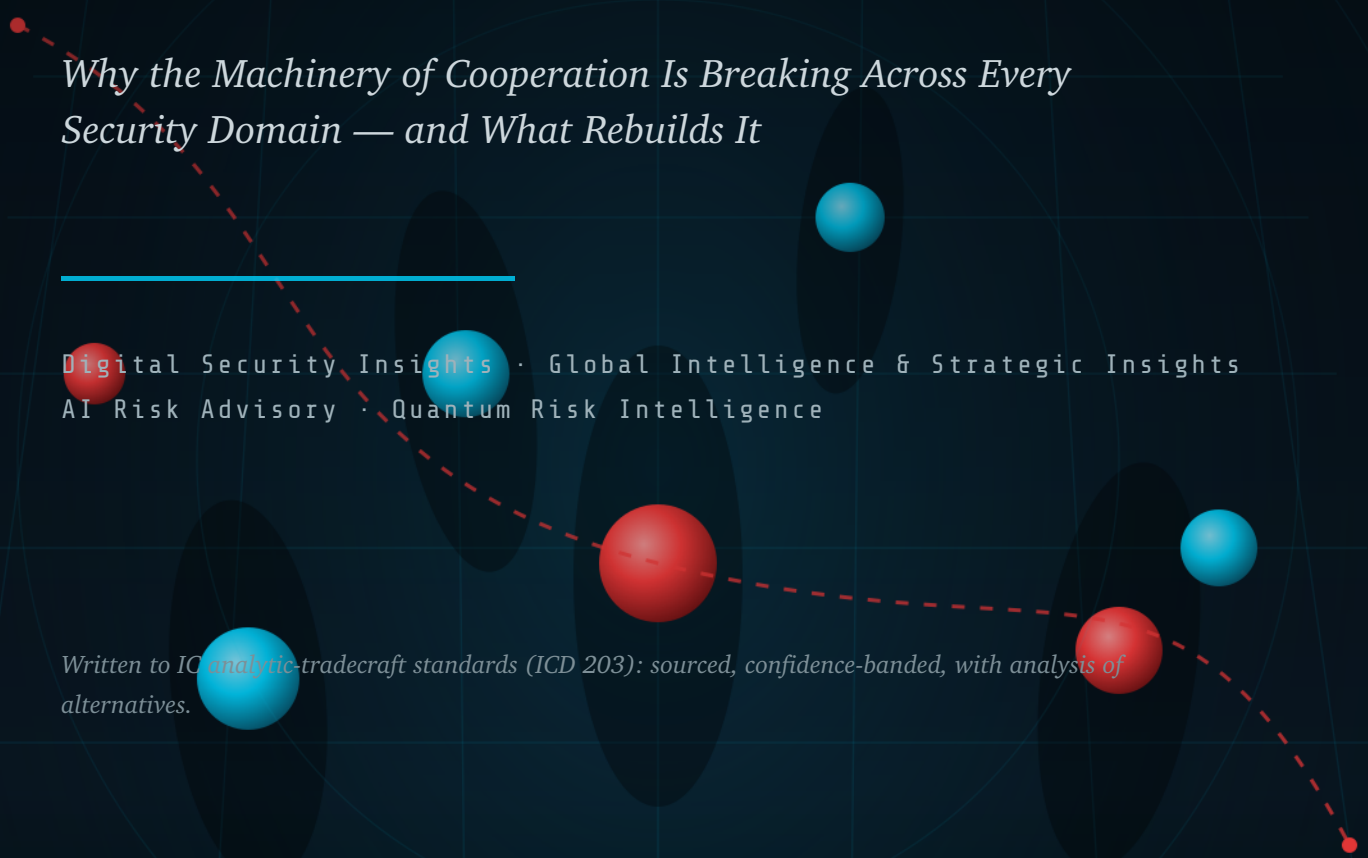
DSI.ADVISORY
DIGITAL SECURITY IN GEOPOLITICAL CONTEXT

DSI ADVISORY SERVICES · Q2 2026 SPECIAL REPORT

The Shadow of the Future

Why the Machinery of Cooperation Is Breaking Across Every Security Domain — and What Rebuilds It



Digital Security Insights · Global Intelligence & Strategic Insights
AI Risk Advisory · Quantum Risk Intelligence

Written to IC analytic-tradecraft standards (ICD 203): sourced, confidence-banded, with analysis of alternatives.

The Shadow of the Future

Why the Machinery of Cooperation Is Breaking Across Every Security Domain — and What Rebuilds It

---

DSI Advisory Services — Q2 2026 Special Report *Digital Security Insights · Global Intelligence & Strategic Insights · AI Risk Advisory · Quantum Risk Intelligence*

Analytic standard: this report is written to IC analytic-tradecraft conventions (ICD 203). Judgments are stated in the first person and distinguished from underlying fact; uncertainty is expressed as explicit confidence levels with reasons; sources are described and cited in endnotes; each assessment carries an analysis of alternatives.

Introduction — The Game Nobody Named

On 3 September 1949, a US Air Force WB-29 flying a weather-reconnaissance track from Japan toward Alaska drew air across a set of filter papers at 500 millibars, east of Kamchatka. When the aircraft landed, the filters were radioactive. Radiochemical analysis found short-lived fission products — the kind that decay in weeks, not years — and the laboratories concluded the debris was "of fairly recent origin, their age probably being one month or less."[1] Isotopes like barium-140 and cerium-141 do not linger; their presence is not a trace of history but a timestamp. Something had undergone nuclear fission, recently, and the United States had tested nothing that year. The only remaining explanation was the one Washington had been dreading: the Soviet Union had the bomb, years ahead of forecast.[2]

The American nuclear monopoly, the achievement of the Manhattan Project, was over. And in the panic that followed, some of the sharpest minds in the US defense establishment argued seriously for a first strike — to use the advantage before it evaporated. The question of what to do about nuclear weapons, and fast, landed at the RAND Corporation, the Air Force–funded think tank that had become the country's laboratory for Cold War strategy.[3]

It was there, in 1950, that two mathematicians — Merrill Flood and Melvin Dresher — devised a small game to probe how rational actors behave when their interests only partly align. Albert Tucker later dressed it in the story that gave it its name: two prisoners, held separately, each offered a deal to betray the other. Betray while your partner stays silent and you walk free; both betray and you both serve time; both stay silent and you both get off lightly. The rational move, for each prisoner considered alone, is always to betray — and so two rational players arrive together at an outcome that is worse for both than if they had simply trusted each other.[4] Flood and Dresher had not set out to model the superpower standoff. But they had built, in miniature, the exact structure of it: two rivals, each individually rational, driving one another toward a mutually destructive result that neither wanted.

The dilemma was born inside the nuclear problem. Seventy-five years later, it is the operating structure of nearly every security problem we face — and the conditions that once let rivals climb out of it are being dismantled, in the open, across every domain at once. That is the subject of this report.

The finding

The Prisoner's Dilemma looks hopeless in a single encounter: defection is the rational move, and mutual defection the rational result. What changes everything is repetition. When the same players meet again and again, cooperation becomes not merely possible but, under the right conditions, the winning strategy. In 1980 the political scientist Robert Axelrod proved this empirically, and his work is the analytical instrument this report is built on. We set it out in full in Chapter 1.

Axelrod's deeper result was that cooperation among self-interested rivals is not a moral achievement but a *structural* one. It depends on a small set of conditions — chiefly a long "shadow of the future" (players expect to meet again), clear signals (each can read what the other actually did), enforceable reciprocity (defection can be answered), and legibility (a reputation that travels). Where those conditions hold, cooperation is stable without trust, friendship, or a central authority. Where they fail, the game collapses back to mutual defection.

My central assessment is this, at high confidence: across every domain DSI covers — chokepoints and trade, ransomware, cyber attribution, the AI race, the cryptographic transition, and digital identity — the security environment is systematically destroying the conditions Axelrod identified as the requirements for cooperation. Anonymity is collapsing the shadow of the future. Misattribution is corrupting the signals. Decoupling is shortening the horizon of every trade relationship. And it is happening not by accident but as the cumulative result of policy choices, commercial incentives, and adversary strategy pulling in the same direction. The machinery of cooperation is being disassembled, one condition at a time, and almost no one is pricing the loss.

The business translation — the Defection Premium

That last phrase is not rhetorical. Cooperation is an asset, and it is being written off the balance sheet without anyone booking the write-down.

Every time an organisation — or a state — is pushed into a game where defection dominates, it starts paying what this report calls the **Defection Premium**: the recurring, quantifiable cost of operating without cooperation. It shows up as redundant and reshored supply chains, as cyber-insurance repricing and higher retentions, as the overhead of zero-trust architecture and attribution tooling, as litigation reserves against biometric and privacy exposure, as the reputation tax that marks a ransom-payer as a repeat target, and as the lost network effects of markets that no longer trust each other enough to interconnect. The premium is real and it is already being paid. What it is not, in almost any organisation we examined, is *consolidated*. It is

scattered across a dozen budget lines and a dozen owners, so no one sees the total — and a cost no one totals is a cost no one governs.

The through-line of this report, for the board as much as the analyst, is that the collapse of cooperation is the largest unpriced liability in the current risk environment. Naming it, and quantifying it, is the advisory work.

The Defection Premium: the cost of operating without cooperation, scattered across budget lines no one totals. The advisory work is to consolidate it onto one page.

How to read this report

This is DSI's Q2 2026 quarterly, built as a single argument rather than a summary of the quarter's briefings. It is organised around one master question:

> **Which of the conditions that make cooperation possible has my environment already destroyed — and can I rebuild them before the game collapses to mutual defection?**

Chapter 1, *The Instrument*, establishes Axelrod's work honestly — including the four decades of scholarship that complicated it — and distills from it the four-condition diagnostic and the strategy-selection framework the rest of the report applies. A short interlude, *Whose Game Is This?*, then asks the question that precedes the diagnostic: whether every player at the table is even playing the same game — the Western Prisoner's Dilemma against China's Go and Russia's reflexive control. Chapters 2 through 8 each take one live security domain and run it through the diagnostic: the Chokepoint Game, the Ransomware Game, the Attribution Game, the AI Race, the Cryptographic Clock, the Trust Game, and the Reserve Game. Chapter 9 is the scorecard — DSI's Q2 predictions, graded in the open, misses included. The report closes on the forward view: the Q3 outlook, and the argument for agency.

Each domain chapter carries two executive instruments alongside the analysis: a **Business Translation** (which line item moves, and by roughly how much) and a **Risk Signal** (the leading indicators that this particular game is tipping toward defection — the early-warning tells a risk officer can actually monitor). Together they form the report's two board-ready deliverables: the Cooperation Risk Register and the Leading-Indicator Watchlist.

What each reader is really asking

The master question lands differently on each side of the Atlantic, and differently for each reader. The report is written to answer all of these:

READER	THE QUESTION THE REPORT ANSWERS	THE NUMBER THEY LEAVE WITH
Board / CEO	Which of our dependencies are one-shot games we have mispriced as relationships?	Our Defection Premium, and what re-prices when a game snaps
Chief Risk Officer	Which games are tipping, and how fast?	The Risk Register and the Leading-Indicator Watchlist
CISO	Can I actually attribute an attack — and is my posture a pushover, or a grudge-holder?	An attribution-confidence read and a strategy-selection call
General Counsel	Where is the ungoverned-space liability?	Biometric, GDPR, and AI Act contingent exposure
Policymaker / regulator	Are we building the conditions for cooperation — or destroying them and calling it security?	A structural audit of where policy shortens the shadow of the future
Investor / insurer	Which sectors move from cooperation to defection, on what timeline?	A sector tipping map with Q3 confidence bands

For a US reader, the sharpest variants are these: how much of my model rests on the ungoverned 47-state space that a single state — or the EU — can make expensive overnight; whether my China-facing supply chain is a repeated game turning into a one-shot game; and whether competitive pressure is forcing me to defect on AI safety in a way that is individually rational and collectively catastrophic. For a European reader: whether my dependence on US hyperscalers is a cooperation game whose shadow of the future — the alliance itself — is being tested, and whether I hold a credible exit; whether NIS2, DORA, the CRA and the AI Act are building the clear-signal-and-credible-retaliation infrastructure that stabilises cooperation, or merely compliance theatre; and whether Europe is, strategically, an always-cooperate player being read as a pushover.

The argument for agency

There is a temptation, in a report about collapse, to mistake description for destiny. This report resists it, on the authority of the theory itself. Axelrod's tournaments showed that the environment shapes the players in the short run — the payoffs determine who does well this round. But they also showed the reverse over the long run: the players shape the environment. A cluster of cooperators, interacting deliberately with one another, can invade and transform a world of defectors.[5] The conditions for cooperation are not weather. They are built, and they can be rebuilt — by persistent identity that restores the shadow of the future, by attribution that cleans the signal, by shared reputation systems and credible, proportionate retaliation.

The security industry has spent a generation asking how to win each round. The more important question, and the one this report is finally about, is which of the four conditions our own choices are destroying — and whether we will rebuild them before the game closes.

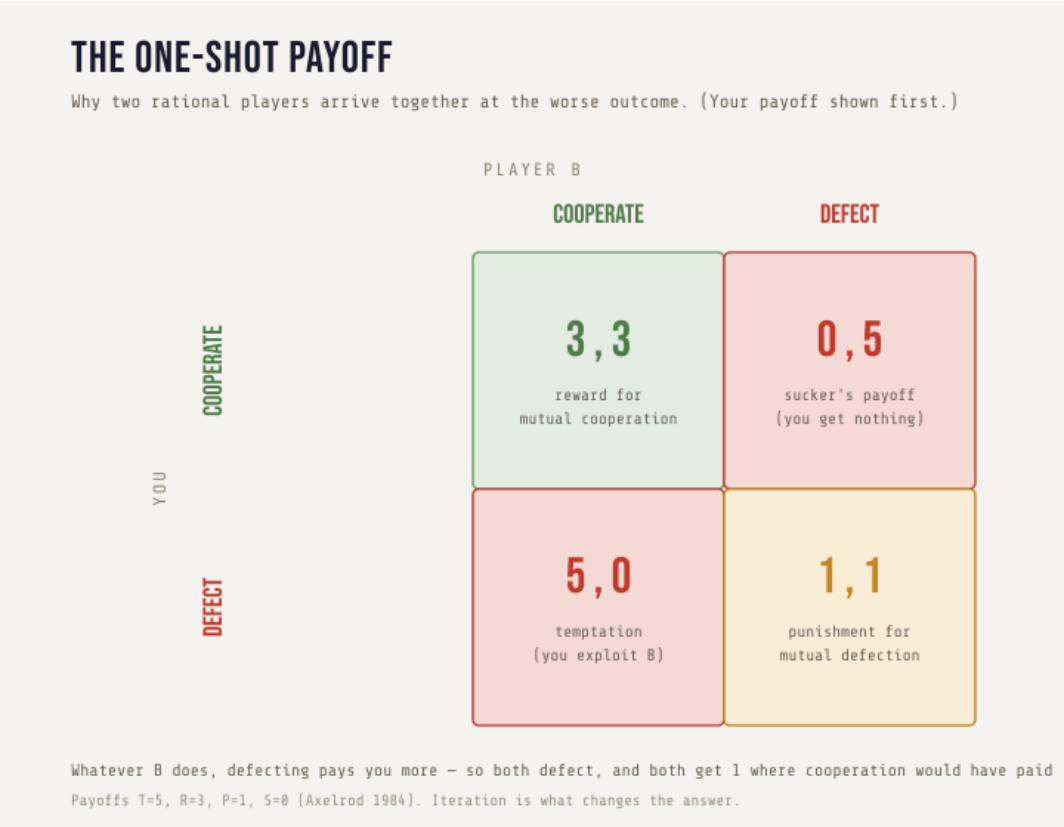
Chapter 1 — The Instrument

Robert Axelrod, and the Diagnostic This Report Runs on Everything

The reason the Prisoner's Dilemma has generated thousands of academic papers is that it is not a puzzle to be solved once but a structure that reappears — between nations, firms, roommates, and cells. This report treats it as none of those things directly. It treats Axelrod's work as an *instrument*: a diagnostic for asking, of any security relationship, whether the conditions that let cooperation survive are present or destroyed. To use the instrument honestly, we have to understand both what Axelrod found and — just as important for a reader trained to distrust a too-clean result — the forty years of scholarship that complicated it.

What Axelrod actually did

In a single Prisoner's Dilemma, defection is the dominant move and mutual defection the rational outcome. Axelrod's question was what happens when the game is *iterated* — played repeatedly by the same two players, each able to remember what the other did last time. To find out, in 1980 he ran a computer tournament. He invited game theorists across economics, psychology, political science, mathematics, and sociology to submit strategies as computer programs. Each strategy played every other strategy, a copy of itself, and a random strategy, in matches of 200 moves, replicated five times. The goal was simple: accumulate the most points across all encounters.[6]



In a single game, defection dominates: whatever the other player does, defecting pays you more — so both defect and both get 1 where cooperation would have paid 3. Iteration is what changes the answer.

Fourteen strategies were submitted. The winner was the simplest program in the field — four lines of code, called **Tit for Tat**, submitted by the mathematical psychologist Anatol Rapoport. Tit for Tat cooperates on the first move, then does whatever its opponent did on the previous move: cooperation for cooperation, defection for defection, and back to cooperation the moment the opponent returns to it.[7]

Axelrod then did something a lesser researcher would not have. He published the full results and the analysis of *why* Tit for Tat won, and ran the tournament again — this time with 62 entrants from six countries, every one of whom knew Tit for Tat had won and could design specifically to beat it. Rapoport submitted Tit for Tat again. It was the only entry of its kind, and it won a second time.[8]

One precision matters here, because it is where casual accounts overreach. **Tit for Tat never beats any single opponent head-to-head.** By construction it can only match or trail the player across from it — it never defects first, so it can never come out ahead in a pairwise contest. It won the tournament on *aggregate* score, by producing good outcomes against a wide range of opponents while never getting badly exploited. The lesson is not that Tit for Tat dominates; it is that a strategy which cannot win any individual game can win the whole field.[9]

The four properties

From the results, Axelrod identified four properties that tended to make a strategy successful. In his own words, they were "avoidance of unnecessary conflict by cooperating as long as the other player does" — being **nice**, never the first to defect; "provocability in the face of an uncalled for defection" — being **retaliatory**, answering defection immediately; "forgiveness after responding to a provocation" — being **forgiving**, not holding the grudge once the other returns to cooperation; and "clarity of behaviour so that the other player can adapt to your pattern of action" — being **clear**, legible enough that the other side can learn to cooperate with you.[10]

The strength of the finding was in the ranking. The eight top-scoring strategies were all nice — none defected first — and even the worst-performing nice strategy outscored the best-performing nasty one. Cooperation was not the sentimental option; it was the winning one. Nastiness did well early, exploiting the naïve, then ran out of victims and collapsed.

(A note for the careful reader, because the two are constantly conflated: these four properties *— nice, retaliatory, forgiving, clear — are Axelrod's analysis of what made strategies succeed. He offered a separate list of four pieces of* advice *to a player — don't be envious, don't be the first to defect, reciprocate, don't be too clever — and it is in that second list, not the first, that "don't be envious" belongs. This report keeps them apart.)*[11]

The shadow of the future

Why does iteration change the answer so completely? Because it introduces what Axelrod called the **shadow of the future** — the weight that the prospect of future encounters casts on the present one. Formally it is governed by a parameter, w: the probability that the two players will meet again. When w is high — when the future is long and likely — the accumulated gains from continued cooperation outweigh the one-time payoff of defecting now, and cooperation becomes rational. When w is low — when this encounter is probably the last — the future stops disciplining the present, and defection returns as the dominant move.[12]

This is the single most important idea in the report, so it is worth stating plainly: **cooperation does not require trust, affection, or an enforcer. It requires only a long enough shadow of the future.** Shorten that shadow — through anonymity, through a relationship one side expects to end, through a one-shot encounter engineered at scale — and you do not need bad actors to get defection. Rational actors will produce it on their own.

Two cautions keep this from being overstated, and a report read by trained analysts must state them. First, Tit for Tat is not a universal solution. Axelrod's own term for its robustness is **collective stability**: once a population plays Tit for Tat, no single mutant strategy can invade it — but only if w is large enough. Below that threshold it can be invaded.[13] Second, and more important, Tit for Tat is not the *only* collectively stable strategy. "Always defect" is also collectively stable, and it is stable for *any* value of w: a world of mutual defectors resists invasion

by any lone cooperator, because the first one to try cooperating simply gets exploited. Cooperation cannot get started one individual at a time. What Axelrod showed is that it can get started in **clusters** — a small group of reciprocators who interact disproportionately with one another can gain a foothold in a world of defectors, and once established, cooperation is hard to dislodge.[14] That asymmetry — defection needs no coordination to persist, cooperation needs a cluster to begin — is the hinge the entire report turns on.

The forty years that complicated it — and why that makes the instrument sharper

If the story ended in 1984 it would be an elegant relic. It is a working instrument precisely because the scholarship that followed stress-tested it, and the failure modes it exposed are the ones that map onto modern security. Three matter here.

Noise breaks Tit for Tat. In the real world, signals are corrupted: a cooperative move is misread as a defection, or an intended cooperation is executed as a defection by error. Introduce that noise, and two Tit-for-Tat players fall into a catastrophe — a single misread defection triggers an endless alternating echo of retaliation, or locks both into permanent mutual defection, because each is mechanically punishing the other for a mistake neither meant.[15] The fixes are instructive. **Generous Tit for Tat** forgives a defection with some probability, damping the echo. **Contrite Tit for Tat** recognises its *own* mistaken defection and accepts the retaliation without re-escalating. The two are different mechanisms — one forgives the opponent, the other atones for itself — and both say the same thing: in a noisy world, strict reciprocity is too brittle, and a measure of forgiveness outperforms it.[16] Hold that finding; it is the entire Attribution chapter in advance. When you cannot reliably tell cooperation from defection — which is the definition of the cyber attribution problem — Tit for Tat does not just underperform. It self-destructs.

There is a real-world instrument reading for this. On 26 September 1983, at the Serpukhov-15 bunker outside Moscow, the Soviet Oko early-warning system reported a US intercontinental missile launch, then four more behind it. The duty officer, Stanislav Petrov, judged it a false alarm — a genuine first strike, he reasoned, would not arrive as five lonely missiles — and declined to report it up the chain as a real attack. The cause was later traced to sunlight glinting off high-altitude clouds into the satellites' sensors.[17] It is the purest case on record of noise in an iterated game between rivals: a signal error that, mechanically retaliated against, ends the game and the species. Petrov was the generous move, applied by a human being, against a system built for strict reciprocity.

Extortion is mathematically possible — and evolutionarily self-defeating. In 2012 William Press and Freeman Dyson startled the field by identifying **zero-determinant strategies**: a player can unilaterally impose a fixed linear relationship between the two players' long-run payoffs, and in the *extortionate* form can guarantee its own surplus is always some multiple of the opponent's, coercing a rational opponent into raising the extortioner's score.[18] It looked, briefly, like a proof

that domination beats cooperation. But the sequel closed the gap: in 2013 Stewart and Plotkin showed that extortionate strategies, while they win head-to-head, *fail in evolving populations* — they perform poorly against copies of themselves and cannot spread. The zero-determinant strategies that *do* succeed evolutionarily are the **generous** ones, which reward cooperation and punish defection only mildly. In the long run, the field bends back toward generosity.[19] The two results must always be cited together; reported alone, the 2012 paper is routinely misread as proof that extortion pays. It pays in a single fixed matchup and loses in a living system — which is exactly the distinction between a one-shot game and an iterated one.

Reputation moved from direct to indirect. Axelrod's Tit for Tat punishes the player who defected *on you*. But most modern cooperation runs on **indirect reciprocity** — you cooperate with someone based on how they treated *third parties*, mediated by reputation. Martin Nowak and Karl Sigmund formalised this in 2005, and Nowak's 2006 synthesis placed it among the five mechanisms by which cooperation evolves at all: kin selection, direct reciprocity, indirect reciprocity, network (spatial) reciprocity, and group selection.[20] Indirect reciprocity is the mechanism behind credit scores, vendor risk ratings, threat-intelligence sharing, and breach-notification registries — every system in which an actor's treatment of others becomes a signal that travels. It is also, therefore, a condition that can be attacked: poison the reputation layer, and cooperation loses the memory it runs on.

The diagnostic

From this — Axelrod's result and its complications together — the report extracts a four-part instrument. For any security relationship, we ask whether these conditions are present or destroyed:

1. **A long shadow of the future.** Do the parties expect to interact again, and does that expectation discipline the present? *(Destroyed by: anonymity, decoupling, engineered one-shot encounters, exit.)*
2. **Clear signals.** Can each party reliably tell what the other actually did — cooperation from defection, attack from accident? *(Destroyed by: misattribution, deniability, noise.)*
3. **Enforceable reciprocity.** Can defection be answered, proportionately and credibly? *(Destroyed by: attribution failure, capability asymmetry, the absence of a legitimate response.)*
4. **Legible reputation.** Does an actor's conduct toward others travel as a signal that shapes future dealings? *(Destroyed by: fragmentation, anonymity, poisoned or captured reputation systems.)*

Where the four hold, cooperation is stable without trust. Where they fail, the game reverts to mutual defection regardless of anyone's good intentions. **The report's method is to run each security domain through these four conditions and report which ones its own environment has already destroyed.**

The whole argument on one page: each of the seven games scored against Axelrod's four conditions. Red is destroyed, amber degraded, green holding. No domain keeps all four.

Paired with the diagnostic is a decision tool drawn straight from the four properties — a **strategy-selection framework** for a defender or a state deciding how to play a given counterparty: be *nice* (do not defect first) where the shadow of the future is long; be *provocable* (answer defection) where you can actually attribute it; be *forgiving* (do not escalate on ambiguous signals) where noise is high; and be *clear* (make your response rules legible) so the other side can learn to cooperate with you. A pushover — unconditional cooperation — invites exploitation; a grudge-holder — unforgiving retaliation in a noisy channel — destroys the relationship over a misread. Most organisations, examined honestly, are one of those two failure modes.

Analysis of alternatives

Intellectual honesty, and the standard this report is held to, require the case *against* its own instrument. Three objections are serious.

The first is that these are toy models — two players, fixed payoffs, perfect memory — and real security is *n*-player, with shifting payoffs and imperfect information. True. The report's defence is that it uses game theory as a **diagnostic of conditions, not a predictor of moves.** We do not claim to compute what Iran or a ransomware crew will do next; we claim that when the four conditions degrade, cooperation becomes harder to sustain, and that this holds across model specifications. The finding we lean on — cooperation requires a shadow of the future, clear signals, reciprocity, and legibility — is robust precisely because it is qualitative.

The second is that real actors are not rational payoff-maximisers; they are ideological, emotional, bureaucratic, sometimes simply mistaken. Also true — and it strengthens rather than weakens the argument, because the noise and misperception literature (Petrov, the echo spirals) is *about exactly that failure of clean rationality*, and it makes the conditions harder to satisfy, not easier. A model that assumed perfect rationality would understate the danger.

The third is the deepest: that framing security as a cooperation problem risks counselling accommodation with adversaries who are playing to win, not to cooperate. This is the objection a sophisticated reader in a national-security role will raise first, and it deserves a direct answer. The report does not counsel cooperation as a value. It observes that cooperation is a structural *asset* — the thing that makes an interconnected system cheaper and more stable to run than a fragmented one — and that this asset is being destroyed, sometimes by adversaries and sometimes by our own hand. Naming when a game is genuinely one-shot (where defection is correct) versus iterated (where it is not) is precisely the discipline the instrument provides. Tit for Tat, remember, is not a pacifist strategy. It is *provocable*. It retaliates immediately and without exception. What it refuses to be is either a pushover or a grudge-holder — and that is the posture this report recommends, not accommodation.

With the instrument built, the rest of the report is application. Chapter 2 runs the first and largest game through it: the chokepoints of the global economy, and the shadow of the future going dark over trade.

Interlude — Whose Game Is This?

Before You Run the Diagnostic, Ask What Game You Are In

The instrument we have just built is a Western object. The Prisoner's Dilemma is a game of two players, discrete rounds, a binary choice, and a scoreboard; chess sits behind it, with its decisive engagements and its capture of the king. There is a hidden assumption in using it as the master frame for global security: that everyone at the table is playing the same game we are. That assumption is worth interrogating before we run the diagnostic across seven domains — because if an adversary is playing a different game, our reciprocity strategy may not just fail. It may be steered.

The board the West does not see

In March 2016, a machine named AlphaGo beat the professional Go player Lee Sedol, and in the second game it played a move — the thirty-seventh — that no human would have chosen, a stone so far from the fighting that commentators assumed it was an error. It was not. It was the whole board.[21] Go, or *weiqi* (圍棋), is the game the strategic-studies literature has long used to describe how China thinks about strategy, and the contrast with chess is the point. Henry Kissinger opened *On China* with it; David Lai built a US Army War College monograph around it; Scott Boorman read Maoist strategy through it half a century ago.[22]

The two games teach opposite instincts. Chess is about decisive battle and checkmate — annihilation. Go is about *encirclement* and relative influence: you place stones to build *shi* (勢), a configuration of latent potential, and you win not by capturing the center but by surrounding more of the board than your opponent. You run many local games at once, so an opponent who can only concentrate force in one place is stretched across the whole board. You can be losing every local fight and winning the game.

CHESS / PRISONER'S DILEMMA	GO / WEIQI
Two players, one contest	Many local games on one board at once
Capture, checkmate — annihilation	Encirclement — reduce the opponent's *liberties*
Decisive moves	Positional stones; influence accrues over 100 moves
Win / lose	*Relative* territory and influence
Value is the piece	Value is *shi* — configuration and potential

My assessment is that the Chokepoint Doctrine this report documents is, structurally, a Go doctrine, not a chess one. Weaponised interdependence does not checkmate; it occupies the

hubs of a network and quietly reduces the other side's breathing room. That is *weiqi*. When a chess-minded West waits for the decisive move, it misreads a positional stone — and its retaliation becomes another stone in the encirclement.

The essentialism trap — handled directly

A claim this convenient demands a guardrail, and an intelligence reader will insist on it. The "China plays Go, the West plays chess" thesis can curdle into caricature — the inscrutable, patient East — and serious scholarship rejects the stereotype. Alastair Iain Johnston's *Cultural Realism* showed empirically that Chinese strategic history is dominated not by Confucian harmony but by a hard-realist "parabellum" paradigm: force is efficacious, conflict is quasi-zero-sum, the offensive is preferred.[23] The Chinese, in other words, also play chess when chess is what the moment requires. So the discipline here is strict: **Go is a lens we hold up to check our first lens, not a claim about an ethnic mind.** Used that way it sharpens; used as essence it misleads.

The game that plays the player

If China's game lengthens the horizon, Russia's attacks something more fundamental: your ability to read the board at all. Russian military science has a name for it — *reflexive control* (рефлексивное управление), a doctrine originating with the mathematician Vladimir Lefebvre and catalogued for Western readers chiefly by Timothy Thomas: the practice of feeding an adversary specially shaped information so that he *voluntarily* makes the decision you want, believing it his own.[24] Its broader cousin is *maskirovka*, the deception doctrine — related but distinct: *maskirovka* masks the board; reflexive control moves the player.

This is the most corrosive move in the entire report, and it is worth seeing why in the terms of Chapter 1. Reflexive control is the *deliberate manufacture of the noise problem*. Petrov's 1983 false alarm was an accident of sunlight on clouds; reflexive control engineers Petrov moments on purpose — the false flag, the calibrated ambiguity, the mixed signal — so that a Tit-for-Tat opponent misreads a defection that was never there, or freezes when it should move. It does not defect against you. It destroys your ability to tell whether anyone defected at all — an attack on the epistemic precondition of every cooperation strategy Axelrod ever tested. We have documented it in our own reporting: read the Medvedev "two straits" remarks through this lens and they resolve into a reflexive-control signal — a calibrated nuclear-thermonuclear ambiguity delivered to Western capitals to shape their reading of the escalation ladder.[25]

PLAYER	THE GAME	THE MOVE
China	Go (*weiqi*)	Encircle the board over a long horizon; accrue *shi*

PLAYER	THE GAME	THE MOVE
Russia	Reflexive control (over a chess/poker base)	Corrupt the opponent's perception so he defeats himself
The West (our frame)	Chess / Prisoner's Dilemma	Calculate the position; answer defection with reciprocity

The axis as a Go position

Put the board together and a shape appears. China sits near the center and does not fight; it supplies influence — a *moyo*, a framework of latent territory — through the Belt and Road Initiative, which since 2013 has drawn roughly 150 countries into memoranda of understanding and something on the order of a trillion dollars of engagement over its first decade.[26] Around that center sit stones that each open a *local* game, tying down the West's liberties on a different part of the board: Russia in Europe, Iran in the Gulf, North Korea in the financial and missile domains, and — on a different logic — Pakistan as a counterweight fixing India, the West's preferred balancer.

The striking thing is how cleanly those stones distribute across the games this report already analyses. Russia opens the attribution game (Chapter 4). Iran holds the maritime chokepoints (Chapter 2). North Korea's Lazarus Group finances a regime through crypto theft — the FBI attributed the ~$1.5 billion Bybit theft of February 2025 to the DPRK cluster, and a 2023 US estimate held that roughly half of North Korea's missile program was funded by cyber-stolen proceeds (Chapters 3 and 6).[27] China holds the center — AI, rare earths, cables, standards, and the reserve question of Chapter 8. **Our seven games are not seven separate contests. They are local fights in one Go position.**

Here the guardrail returns, because this is exactly where analysis tips into alarmism. **The stones move themselves.** What Kendall-Taylor and Fontaine named the "Axis of Upheaval," and what analysts abbreviate CRINK — China, Russia, Iran, North Korea — is a loose alignment of convenience, not a bloc under Beijing's command.[28] China does not *place* these stones; it *benefits from* their independent placement. Two facts keep the reading honest. China brokered the Saudi-Iran normalisation of March 2023, and as the largest importer of Gulf crude it has a structural interest in the strait staying *open* — so "China supports Iran" is true at the influence layer and false at the chaos layer; a closed Hormuz hurts Beijing.[29] And Pakistan is not part of the revisionist axis at all; it is a China-aligned stone on an anti-India logic, not a system-overturning one. Overstating coordination is the classic error, and the Go metaphor tempts it. The honest formulation: **China plays the center of a board it does not fully control, and its partners are agents, not pieces.**

The coldest reading, and the one the evidence supports, is that to a player at the center some stones are expendable. Beijing pays nothing while Russia grinds itself down in Ukraine, exhausts

Western stockpiles, delivers discounted energy, and keeps America distracted from the Pacific. In Go you sacrifice stones to gain *sente* — the initiative. The West sees an alliance; the center may see a stone it is content to lose.

The point of the interlude

There is a reason this sits between the instrument and its application. The report is about to run a four-condition diagnostic across seven domains. That diagnostic assumes we can see the board. This interlude is the warning that we may not — that a Go player is measuring by influence while we count pieces, and that a reflexive-control player is shaping what we think we see. So a fifth question precedes the four conditions, and against the most sophisticated adversaries it may already be compromised:

> **Whose game is this — and what, to them, counts as a move?**

The failure mode this names is *model-misattribution*: not misjudging whether an opponent defected, but misjudging what game the opponent is playing, so that every subsequent judgment inherits the error. It is the ungoverned intersection at the level of strategy itself. And it returns us, sharpened, to the closing argument of AlphaGo: the game the West used to symbolise Chinese strategic culture was won by a Western machine — and that defeat, when the program beat the Chinese champion Ke Jie in May 2017, was received in Beijing as a Sputnik moment and followed within two months by China's national AI plan.[30] The board keeps folding back on itself. With that caution in hand, we turn to the games.

Chapter 2 — The Chokepoint Game

The Shadow of the Future Goes Dark Over Trade

On the morning of 7 July 2026, two ships were struck by projectiles in the Strait of Hormuz. One, a Saudi-flagged crude supertanker, took a missile fired from the Iranian shore; the other, a Qatari LNG carrier, caught the second projectile off the coast of Oman.[31] By that afternoon the maritime threat level for the strait had been raised to "severe," and Brent crude had moved roughly two and a half percent on the session.[32] The attacks did not come out of a clear sky. Iranian forces had reportedly declared the strait "closed" in early March 2026 and had been intercepting transiting vessels for four months; average crude throughput had, by contemporaneous estimates, fallen from more than fifteen million barrels a day before the war to roughly four million by early summer.[33]

I open here not because Hormuz is the most consequential chokepoint of 2025-2026 — it is not — but because it is the most legible. A missile hits a hull, insurers reprice war risk overnight, and every party can see who defected. The harder chokepoints of this period, the ones that will still be reshaping global trade in 2030, do not announce themselves with fire on the water. They arrive as a licensing rule, an export-control threshold, a percentage clause buried in a customs notice. This chapter is about what those two kinds of chokepoint have in common, and why the quieter ones are doing more damage to the thing that made trade cooperative in the first place: the expectation that the game continues.

The Game, Framed

For roughly three decades, global trade functioned as an iterated game. Firms and states dealt with one another repeatedly, and the expectation of repeat interaction disciplined behaviour. Robert Axelrod's insight — that cooperation among self-interested rivals becomes rational once the shadow of the future is long enough — describes the political economy of globalisation almost exactly.[34] Nobody trusted anyone. Everybody cooperated anyway, because the payoff from defecting once was smaller than the stream of payoffs from continuing to trade. Just-in-time inventory, single-sourced rare-earth refining, transpacific chip fabrication: each was a bet that the counterparty would still be there next quarter, and the quarter after that.

The players in the chokepoint game of 2025-2026 are the United States and China at the centre, with the Gulf states, Taiwan, the European Union, and a widening band of "connector" economies (Vietnam, Mexico, the UAE) arranged around them. The iterated payoff was efficiency: the lowest landed cost for the highest-quality input, compounded across millions of transactions. What is shortening the shadow of the future is the discovery, by both hub powers, that interdependence is not merely efficient — it is a weapon.

Henry Farrell and Abraham Newman gave this discovery its scholarly name in 2019. In *Weaponized Interdependence*, they showed that states with jurisdiction over the central hubs of global networks — the dollar-clearing system, SWIFT, the physical internet, the chip-design toolchain — can exploit those hubs two ways: the *panopticon effect*, harvesting the information that flows through the chokepoint, and the *chokepoint effect*, denying rivals access to it.[35] The paper is the academic twin of what DSI has called the Chokepoint Doctrine. Its uncomfortable corollary is the one this chapter turns on: the moment a hub is used as a weapon, every actor downstream re-reads the entire relationship as a one-shot game. Access you can be denied is not access. It is a loan, callable at the lender's discretion.

Three moves in this period made that re-reading unavoidable.

China, on rare earths. Beijing holds roughly 90 percent of the world's rare-earth *processing* capacity and near-monopoly refining shares in gallium and germanium. It began converting that position from commercial fact into strategic instrument in July 2023 (gallium and germanium licensing), extended it to graphite in October 2023, and in December 2024 moved from licensing to outright prohibition of gallium, germanium, antimony, and superhard materials bound for the United States.[36] In April 2025 it added seven medium and heavy rare earths — including dysprosium, terbium, and yttrium — to the control list. Then, in October 2025, it published its most sweeping rule yet: any foreign-made product containing 0.1 percent or more of Chinese-origin rare earths, or made using Chinese processing technology, would require an export licence.[37] That is not a tariff. That is jurisdiction asserted over other countries' factories — the chokepoint effect at continental scale, and, in the terms of this report's interlude, the purest positional move on the board: China does not blockade, it *encircles*, converting a processing monopoly into standing leverage over everyone else's supply chain.

Washington, on compute. The mirror image ran through semiconductors. The January 2025 AI Diffusion framework set global performance thresholds that walled the most capable GPUs out of China; Nvidia's H20 was engineered specifically to sit just under the line. When DeepSeek's models demonstrated in early 2025 that even the throttled chips could deliver frontier-adjacent results, the Commerce Department declared the H20 non-compliant in April, then reversed and resumed licensing in July.[38] By December 2025 the policy had swung again, with the far more capable H200 approved for China under a January 2026 rule reportedly capping China-bound volume at half of US shipments and routing a 25 percent tariff through Taiwan.[39] The specific numbers matter less than the pattern: the toolchain is now a policy lever pulled and released on a monthly cadence.

Iran, on the strait. Hormuz is the crude, physical version of the same logic — a chokepoint asserted not through jurisdiction but through geography and munitions, covered at length in DSI's Iran series.[40]

Running the Four Conditions

The report's diagnostic asks four questions of any cooperation game. Here is how trade scores in mid-2026.

CONDITION	STATUS	WHY
1. Shadow of the future	**Destroyed / critically shortened**	Both hub powers have demonstrated they will weaponise access. Every counterparty now prices the relationship as potentially terminal, not indefinitely repeated.
2. Clear signals	Degraded, not destroyed	A missile strike or an export ban is unambiguous. But monthly policy reversals (H20 to H200) blur the line between a defection and a negotiating feint.
3. Enforceable reciprocity	Intact — dangerously so	Each side *can* retaliate (China throttles minerals, the US throttles compute). Reciprocity works. That is precisely why it escalates.
4. Legible reputation	Partly intact	Conduct travels — allies watch how Washington treats one partner and hedge accordingly. But "de-risking" rhetoric lets defection be reframed as prudence.

The load-bearing finding is condition one. Reciprocity and signalling still function; if anything they function too well, which is why tit-for-tat keeps ratcheting. What has broken is the expectation of continuation. My central assessment, at moderate confidence, is that the shadow of the future over transpacific trade has already shortened past the point where cooperation is the rational default. When a firm assumes it may lose access to Chinese dysprosium or American compute inside a single planning cycle, the efficient move — single-sourcing from the cheapest hub — becomes the reckless one. Redundant, higher-cost, politically-aligned sourcing becomes rational. That is decoupling, and it is not a policy choice imposed from above so much as the aggregate of a million firms correctly solving a one-shot game.

The Honest Counter-Case

The strongest argument against my assessment is that interdependence is sticky, and that decoupling is more announced than achieved. It has real evidence behind it. China *suspended* the

October 2025 rare-earth rule and the US-specific dual-use tightening in November 2025 through MOFCOM Announcements 70 and 72, after a leaders' summit — proof that both sides retain a powerful incentive to keep the game going.[41] Trade volumes between the two economies remain enormous. Much "reshoring" is really "connector-shoring": goods rerouted through Vietnam and Mexico with Chinese content intact, which means the interdependence persists under a new label. And chip policy that swings from H20-ban to H200-approval in twelve months is not the behaviour of a power committed to severance.

I take the objection seriously, and it survives it. The suspensions are the tell, not the refutation. Beijing paused the October rule but retained the April 2025 heavy-rare-earth controls and, more importantly, kept the *architecture* — the 0.1 percent extraterritorial mechanism — on the shelf, demonstrated and ready to redeploy.[42] A weapon holstered is still a weapon owned; every firm now knows the licence regime exists and can return with a communiqué. That knowledge, not its current on/off state, is what shortens the shadow. Stickiness measures the *cost* of exiting the game. It does not restore the belief that the game is indefinitely repeated — and it is that belief, not the transaction volume, that condition one is about. Interdependence can remain high in volume while collapsing in trust. That is the precise configuration this chapter describes.

Business Translation. The Defection Premium in trade is the standing cost of assuming the game may not continue, and it lands on identifiable line items. First, logistics and insurance: war-risk premiums on Hormuz transits and rerouting around contested lanes add directly to landed cost, with Brent's volatility flowing straight into freight and hedging budgets.[43] Second, supply-chain diversification capex: qualifying a second rare-earth refiner outside China, or a non-Chinese magnet supplier, is a multi-year capital programme carried at a structural cost disadvantage against the incumbent — a premium paid for optionality, not output. Third, inventory: just-in-time inventories are quietly becoming just-in-case, tying up working capital in strategic stockpiles of chips, magnets, and APIs. Fourth, lost market access: firms designing US compute into products lose the China market, and firms dependent on Chinese inputs lose US-aligned procurement — the same product now needs two bills of materials. CFOs should model this not as a one-time shock but as a recurring line, because condition one does not reset when a single dispute is settled.

Risk Signal. Four leading indicators tell you the game is tipping further toward defection rather than stabilising. Monitor them monthly.

- **Extraterritorial thresholds, not just tariffs.** Watch for rules that assert jurisdiction over *foreign* production based on minimum-content percentages (China's 0.1 percent rare-earth rule; US de minimis and foreign direct product expansions). Tariffs tax the game; content thresholds try to end it.
- **The licence-approval half-life.** Track how long an export authorisation stays valid before it is revised. Shrinking validity windows — the H20-to-H200 whipsaw — are a direct readout of a shortening shadow.

- **Strategic stockpile announcements.** When governments or large OEMs announce multi-month inventories of magnets, gallium, or advanced chips, they are pricing the game as one-shot. Stockpiling is defection made visible on the balance sheet.
- **Connector-economy content audits.** Watch for US or EU rules of origin that begin policing Chinese *content* inside Vietnamese or Mexican exports. The moment connector-shoring is treated as evasion rather than diversification, the last cooperative off-ramp closes.

Handing Off

The chokepoint game is the macro case: two hub powers discovering that the networks binding them are also the levers by which they can be strangled, and every firm downstream rationally re-pricing cooperation as a one-shot bet. The shadow of the future goes dark over trade not because anyone chose war, but because interdependence became a weapon and weapons cannot be un-seen.

Chapter 3 takes the same four conditions down a level, to a game where the shadow of the future was never long to begin with, where signals are deliberately corrupted, and where reciprocity runs through cryptocurrency rather than customs law: the ransomware economy. If trade shows cooperation decaying, ransomware shows what a market looks like when it was born one-shot — and why, paradoxically, its most professional operators have started trying to lengthen their own shadow.

Chapter 3 — The Ransomware Game

The Reputation Market That Enforces Its Own Cooperation

On 21 February 2024, Change Healthcare — the payment-clearing spine of roughly a third of American patient records — discovered ALPHV/BlackCat inside its network. The company, through UnitedHealth's Optum subsidiary, paid a ransom of about $22 million to buy the deletion of stolen data. The payment cleared. Then the market punished everyone in it. ALPHV's operators pocketed the full sum, staged an exit scam, and vanished without paying the affiliate — a contractor known as "Notchy" — who had actually breached the network. Notchy still had the data. He carried it to a second crew, RansomHub, which extorted Change Healthcare a second time for the same files the company had already paid to erase.[44]

One payment. Three defections layered on top of it: the operator defected on its affiliate, the affiliate defected on the deletion promise, and the ecosystem defected on the victim. This chapter is about why that outcome is the exception rather than the rule — and about the fragile machinery that normally holds it in check.

The Game

Strip a ransomware event down to its decision structure and it is Axelrod's iterated Prisoner's Dilemma wearing a balaclava. Two players who will never meet, who cannot enforce a contract in any court, who have every short-run incentive to cheat. The victim can pay or refuse. The crew, once paid, can decrypt or walk. The one-shot logic is unambiguous: take the money, deliver nothing. In a single encounter, defection dominates.

Yet the striking empirical fact — the one that should unsettle anyone who assumes criminals behave criminally — is that established ransomware-as-a-service crews usually decrypt. They run "customer support" desks. They offer proof-of-life file decryption. They negotiate in something resembling good faith and, for years, they honoured deletion. My assessment is that this reliability is not mercy and not honour among thieves. It is market-making. The crew is not playing one game against one victim; it is playing an indefinitely repeated game against the entire future population of victims, all of whom are watching how the last deal ended.

This is indirect reciprocity — cooperation enforced not by the counterparty but by reputation. A victim's willingness to pay is a bet that decryption works. That bet is priced on the crew's public record: leak-site histories, negotiation transcripts posted to forums, the running commentary of incident responders and cyber-insurers who have paid this brand before. A crew that decrypts reliably is manufacturing the one asset that lets it extract the next ransom. A crew that takes the money and runs is spending its entire future revenue stream to win a single round. In an iterated game with a long horizon, decryption is the rational move. The reputation is the product; the malware is just the delivery mechanism.

The Four Conditions

The report's diagnostic asks four questions of any cooperative equilibrium among rivals. Applied to ransomware, the answers are counterintuitive.

Shadow of the future — conditionally present. A crew that expects to keep operating for years discounts the future lightly, and the future is where its money is. That long shadow is precisely what makes honouring a deal rational. But the shadow belongs to the *brand*, not the *people*. It can be severed overnight — by a takedown, an arrest, or a voluntary rebrand — without any operator changing his behaviour. Hold this; it is the load-bearing variable.

Clear signals — strong. Ransomware has evolved an unusually legible signalling system. Deadlines, sample-data drops, published decryptors, and a fixed negotiation choreography all communicate credible intent. Both sides read the same signals. When a crew posts a working decryptor for one victim, it is signalling to every future victim.

Enforceable reciprocity — weak, and substituted. There is no enforcement mechanism between victim and crew. No escrow, no arbitration, no recourse. Reciprocity cannot be enforced *within* the transaction, so the ecosystem outsources it to reputation — the community of buyers collectively "enforces" by refusing to pay crews known to defect. This is the ecosystem's central adaptation and its central weakness: enforcement lives entirely in the reputation layer.

Legible reputation — strong, and structurally so. This is the surprising result. A criminal market with no courts and no contracts has built one of the most legible reputation systems in commerce. Leak sites are, functionally, public performance dashboards. Security vendors, journalists, and negotiators maintain the ledgers. Cyber-insurers price brands by name. The reputation is *more* legible here than in many legal markets — because the whole extortion model depends on the victim being able to verify that paying works.

Reputation and legibility hold. The shadow of the future holds only as long as the brand survives. And that is exactly the seam the environment has learned to attack.

The Break

Law enforcement has stopped trying to arrest ransomware out of existence and started trying to *bankrupt its reputation*. Operation Cronos, the February 2024 takedown of LockBit, is the model. A ten-country coalition led by the UK's National Crime Agency and the FBI seized 34 servers, took down about 14,000 accounts tied to exfiltration and infrastructure, froze roughly 200 cryptocurrency accounts, and recovered over 1,000 decryption keys — while gathering intelligence on some 194 affiliates.[45]

The operationally decisive move was not the seizure. It was the exposure. The NCA repurposed LockBit's own leak site into a countdown of revelations, and among them was the fact that mattered most to the *market*: victim data that LockBit had promised to delete, in exchange for

payment, was still sitting on its servers.[46] LockBit's product had always been trust — pay us and this ends. Cronos proved, on LockBit's own infrastructure, that the product was fraudulent. At its peak LockBit claimed over 100 active affiliates and posted dozens of victims a week; in the aftermath its cadence collapsed to a handful of posts a month.[47]

My assessment is that a takedown works by collapsing the shadow of the future for a specific crew, and that this conversion is mechanical rather than moral. Shorten a crew's horizon — brand burned, affiliates fleeing, arrest plausible — and you have not created a chastened operator. You have created a defector with nothing left to protect. Reputation only disciplines behaviour when there is a future in which it can be spent. Sever the future and the rational move flips instantly from cooperate to defect: take the last payments, leak the data anyway, exit-scam the affiliates. The Change Healthcare cascade was ALPHV doing exactly this — cashing out its reputation at the moment it decided it had no more use for one.

The rebrand is the same dynamic run in reverse. A burned crew reconstitutes under a new name — DarkSide to BlackMatter to ALPHV, LockBit's diaspora into RansomHub and beyond — precisely to *reset* the reputation ledger to zero, shedding the accumulated distrust. Each rebrand is a fresh shadow of the future purchased at the cost of an old one. And each affiliate migration erodes the norms that made the old brand predictable, because affiliates carry the data — not the operators — and an affiliate between brands answers to no reputation at all.

The end-state is a market fragmenting away from durable, reputation-bearing franchises toward disposable, extortion-only crews. The Salesforce campaigns of 2025 — voice-phishing operators exporting CRM data through malicious OAuth tokens, later consolidating under the "Scattered LAPSUS$ Hunters" banner — mostly skipped encryption and decryption entirely.[48] There is no decryptor to honour, so there is no cooperative equilibrium to sustain. Pure data theft removes the one transaction in which the crew ever had a reason to keep its word. This is the tokenised, ungoverned-intersection risk documented across our platform-key and custodian work — the attacker holds data extracted from a system nobody was governing, and the "deal" was never structurally enforceable to begin with.[49]

Analysis of Alternatives

The honest counter-case deserves a full hearing: takedowns are unambiguously good, and the "they create worse behaviour" argument is a rationalisation for inaction. This position has real evidence behind it. Cronos and its successors have shredded attacker confidence in their own infrastructure; global ransomware payments fell roughly 35 percent in 2024 to about $814 million, and payment rates hit historic lows — Coveware recorded just 25 percent of victims paying in Q4 2024.[50] Fewer victims paying is the outcome we want, and disruption clearly contributed to it.

I do not dispute that takedowns are net-positive; I dispute that the transition is costless. Both things are true at once. Disruption suppresses the *aggregate* payment rate while raising the

conditional probability that a given payer is betrayed — because the crews most likely to be disrupted are the reputation-bearing franchises that were, perversely, the most reliable decryptors. The policy tension for Chapter 8's decision-maker is this: a strategy that correctly makes ransomware less profitable in aggregate also makes each individual ransom decision less trustworthy, and pushes the market toward encryption-free extortion where cooperation was never on the table. Winning the war on the ecosystem and losing predictability for the individual victim are not contradictory. They are the same event seen from two altitudes.

Business Translation. The line item that moves is the *Defection Premium* — the widening gap between what a ransom buys and what it costs. Downtime, not the ransom, is the dominant expense: Change Healthcare's payment was $22 million against a disruption its parent quantified in the billions. As decryption reliability decays, the expected value of paying falls even as demands rise; negotiated payments now average a fraction of initial demands, and the median payment sits near $110,000 against a median demand well above ten times that.[51] Two costs are repricing fastest. First, the *reputation tax* on being a known payer: once you have paid, you are a legible, high-conversion target, and repeat extortion — sometimes by the same data, resold to a second crew — is now a modeled risk, not a tail scenario. Second, cyber-insurance. After a first-ever decline in US written premiums in 2024 and further softening into 2025, underwriters are repricing upward for 2026 as loss patterns turn less predictable.[52] Insurers are also quietly repricing the *deletion promise* to zero: a policy that assumes paid data stays deleted is mispriced, and the smart money now budgets remediation, notification, and regulatory exposure *on top of* any ransom, not instead of it.

Risk Signal. Four monitorable indicators show this game tipping from cooperation toward defection. First, **leak-after-payment incidents** — track the ratio of paying victims whose data surfaces anyway; a sustained rise means the deletion promise is failing across the market, not in one crew. Second, **crew fragmentation post-takedown** — count net-new brands and rebrands in the two quarters after any major disruption; a spike signals reputation-ledger resets and the loss of durable franchises. Third, **affiliate churn** — watch forum recruitment, defection posts, and cross-brand data reuse (the Notchy pattern); mobile affiliates carry data between brands and answer to no reputation. Fourth, **the encryption-free share** — the proportion of extortion events with no decryptor offered; the higher it climbs, the more of the market has exited the cooperative equilibrium entirely, and the less any negotiation can be trusted to conclude.

The uncomfortable synthesis: the same enforcement pressure that is winning the aggregate war is teaching the ecosystem that reputation is a liability to be shed rather than an asset to be built. A market disciplined by reputation is legible and, within limits, negotiable. A market that has learned reputation gets you raided is neither.

The limiting case is the one the interlude placed on the board: a "crew" that is actually a state. North Korea's Lazarus Group does not run a reputation business at all — it steals to fund a regime, and the FBI attributed the roughly $1.5 billion Bybit theft of February 2025, the largest single crypto theft on record, to the DPRK cluster.[53] For an actor whose objective is a missile

programme rather than a repeat customer, there is no shadow of the future to shorten and no reputation to protect; the cooperative equilibrium was never available. That is the far end of the same axis this chapter has traced — from the reputation-disciplined franchise, through the disposable extortion crew, to the sovereign thief who answers to no ledger at all.

Which raises the question Chapter 4 must answer. Reputation only disciplines an actor you can name. The entire reciprocity machinery of the ransomware game — the leak-site ledgers, the brand histories, the insurer pricing — assumes a stable identity to attach the record to. When the actor can dissolve, rebrand, or hide behind a state's plausible deniability, the reputation ledger has no one to bill. The next game is played over the right to remain unnamed.

Chapter 4 — The Attribution Game

Petrov at Machine Speed

In the small hours of 26 September 1983, the Oko early-warning satellites reported a single US intercontinental missile in flight toward the Soviet Union. Inside the Serpukhov-15 bunker the alarm escalated — a second launch, a third, a fourth, a fifth — until the system had raised its own confidence to the top level. Standing orders pointed one way: report a confirmed attack up the chain, where the response was retaliation while the missiles were still in boost phase. The duty officer, Lieutenant Colonel Stanislav Petrov, did not report it. His reasoning was structural — a genuine American first strike would come as hundreds of warheads, not five lonely ones, and the ground radars that should have corroborated the satellites saw nothing. He judged it a false alarm and held. The cause was later traced to sunlight glinting off high-altitude clouds into the satellites' infrared sensors at a rare seasonal angle.[54]

Chapter 1 named this the purest recorded case of noise in an iterated game between rivals: a corrupted signal that, retaliated against mechanically, ends the game. What the incident also preserved — and what this chapter is about — is the *interval*. Petrov had minutes. He used them to weigh the signal against a model of what a real defection would look like, and concluded the signal was lying. Remove that interval, hand the judgment to a machine built for strict reciprocity, and 1983 ends differently.

My assessment is that the cyber domain is systematically removing the interval — and doing so in precisely the channel where the signal is least trustworthy.

Attribution *is* the clear-signals condition

The report's diagnostic asks four questions of any security relationship. The second: can each party reliably tell what the other actually did — cooperation from defection, attack from accident? In the physical world of 1983 that condition was already fragile; sunlight nearly defeated it. In the cyber domain, attribution *is* that condition, and it is not merely fragile. It is contested by design.

Axelrod's Tit for Tat has one non-negotiable prerequisite: you must be able to see what the other player did on the last move. Strip that away and reciprocity has nothing to reciprocate against. The noise literature — Molander's 1985 result, formalised in Chapter 1 — shows the failure is not gradual. Two Tit-for-Tat players who cannot read each other's moves cleanly do not drift toward worse cooperation; they lock into an echo of mutual retaliation, each punishing the other for a defection neither committed.[55] Strict reciprocity over a noisy channel does not degrade. It self-destructs.

Four things corrupt the cyber signal, each mapping to a real technique.

Deniability is the baseline. A missile has a launch point and a ballistic arc. A packet has neither. Routing through compromised third-country infrastructure, rented cloud, and residential proxies means the origin a defender sees is almost never the origin that matters. An attacker gets deniability for free.

False flags weaponise the corruption. The canonical case is Olympic Destroyer, the malware that disrupted the 2018 PyeongChang Winter Olympics opening ceremony. Its code was salted with fabricated attribution artefacts engineered to point analysts at North Korea's Lazarus Group and at Chinese actors; the operation was later attributed to Russia's GRU. The forensic signal did not merely fail to identify the attacker — it was manufactured to accuse the wrong one.[56] Noise as an instrument of policy: Petrov's clouds, placed there on purpose. This is what the interlude named *reflexive control* — the deliberate shaping of an adversary's perception so that he retaliates into a trap of his own choosing; the false flag is that Russian doctrine rendered in code.

Proxies break the link between actor and state. When a nominally independent hacktivist collective, a criminal ransomware crew, or a "patriotic" volunteer runs an operation that serves a state's interest, the state keeps plausible deniability and the defender loses the actor-to-sponsor line any proportionate response requires. The question stops being *what happened* and becomes *whose hand was on it* — which the medium is built to leave unanswerable.

Living-off-the-land erases the tool signature. The clearest 2024–2026 case is Volt Typhoon. US authorities assess this People's Republic of China state-sponsored actor pre-positioned inside US critical-infrastructure networks — communications, energy, water, transportation — using no custom malware at all, only the legitimate administrative tools already on the systems. Blending into normal Windows activity, it evaded endpoint detection and, in some environments, held access for at least five years before discovery.[57] When the intrusion is indistinguishable from ordinary administration, attribution loses even the artefact it once had.

Running the four conditions

Applied to cyber conflict between rivals, the diagnostic reads as follows.

CONDITION	STATUS IN THE CYBER DOMAIN	WHAT DESTROYS IT
Long shadow of the future	Partial — rivals expect to interact again	Intact enough to matter
Clear signals	**Destroyed**	Deniability, false flags, proxies, living-off-the-land
Enforceable reciprocity	**Destroyed (consequence of the above)**	You cannot proportionately answer a defection you cannot attribute

CONDITION	STATUS IN THE CYBER DOMAIN	WHAT DESTROYS IT
Legible reputation	Degraded	Denial, proxy laundering, no shared record of who did what

The shadow of the future survives: the United States, China, Russia, and Iran know they will face one another for decades, and that expectation still disciplines some behaviour. But condition two is gone, and it takes condition three with it. Reciprocity is not independent here — it is *downstream* of clear signals. You cannot answer a defection proportionately if you cannot reliably identify who defected, and a response aimed at the wrong player is not retaliation. It is a new, unprovoked attack the actual attacker gets to watch, having engineered the misfire — the worst move on the board, converting a defender into the first defector against an innocent and handing the guilty party a free escalation.

This is why the cyber domain has produced no stable Tit for Tat and no durable arms control. Every previous arms-control regime rested on verification — confirming what the other side did. There is no verification layer for cyber operations, no equivalent of a satellite counting silos, because the defining feature of the domain is that the defector can make the signal say whatever serves them. UN and other norm negotiations have produced language, not a way to tell cooperation from defection — and without that, the norms have no enforcement, because enforcement requires attribution, and attribution is the thing under attack.

The machine-speed danger

Petrov's defence was time — an interval in which a human could hold the trigger and reason about whether the signal was real. The direction of travel in security operations is to compress that interval toward zero.

The 2025–2026 shift to agentic AI in the security operations centre is real and, for defence, largely rational. Autonomous systems now triage alerts, plan investigations, and execute containment at a speed no human team can match; Gartner projects agentic AI moving from under 5% of enterprise IT-operations deployment in 2025 toward the majority within a few years.[58] The mainstream pattern is deliberately cautious — "human-on-the-loop," where the system acts and a human can intervene after the fact, reserved for cases where mistakes are judged reversible.[59]

The danger is not the defensive automation. It is the structure it creates when two automated systems face each other across a corrupted signal. An agentic responder that treats an ambiguous, possibly-spoofed indicator as a confirmed defection and counter-acts at machine speed is a Tit-for-Tat machine wired into a noisy channel — exactly the configuration Molander showed self-destructs. It is Petrov's panel with the human removed and the standing order automated. The false flag designed to be misread will be misread faster, and answered before anyone can ask whether the clouds are real. "Mistakes are reversible" holds inside one enterprise

and dissolves the moment an automated response crosses a border and lands on the wrong state's infrastructure.

The generous-strategy lesson

Chapter 1 carried the fix forward for exactly this moment. When signals are noisy, the strategies that survive are not the strictest reciprocators but the *generous* and *contrite* ones. **Generous Tit for Tat** forgives a defection some fraction of the time, damping the retaliatory echo before it becomes a spiral. **Contrite Tit for Tat** recognises its own mistaken defection and absorbs the counter-blow without re-escalating. In high-noise environments both consistently outperform strict reciprocity — not because forgiveness is virtuous, but because it is the only thing that keeps a single misread from cascading into permanent mutual defection.[60]

Petrov was the generous move, executed by a human against a machine built for strict reciprocity. The lesson is uncomfortable for a domain that prizes decisiveness: in a channel where you cannot trust the signal, restraint is not weakness — it is the mathematically superior strategy. A state that retaliates hard and fast on ambiguous attribution is not deterring anyone; it is volunteering to be the player who defects first against the wrong target, every time an adversary spoofs a flag. Deliberate slowness — a mandatory human interval, a higher attribution-confidence threshold before any automated counter-action, proportionate restraint under ambiguity — is the generous strategy operationalised. It reads as caution. It is the winning move.

Analysis of alternatives

The honest counter-case has three parts, and a national-security reader will raise each.

Attribution has genuinely improved. True — Western agencies and private firms attribute major operations faster and more publicly than a decade ago. But improved attribution is slow, retrospective, and contested: measured in weeks and months, delivered with hedged confidence language, routinely disputed by the accused. Machine-speed response operates in seconds. The gap between the speed at which good attribution arrives and the speed at which automated systems act is the entire problem, and it is widening.

Private threat intelligence is good enough. Threat-intel firms are excellent at clustering activity and tracking actors. But "good enough to write a report" and "good enough to authorise a retaliatory strike" are different evidentiary standards, and the false-flag technique exists specifically to defeat the forensic clustering threat intel relies on. Olympic Destroyer fooled first-rate analysts for exactly as long as its authors intended.

Deterrence works — the absence of a catastrophic cyber war proves it. The strongest objection, and the assessment survives it directly. What has held is not deterrence through reciprocity; it is restraint under ambiguity — states declining to escalate because they could not cleanly attribute,

or chose not to act on contested attribution. That is the generous strategy already operating by default, through human judgment and bureaucratic slowness. The risk here is that machine-speed automation removes the very slowness that has been keeping the peace, mistaking it for inefficiency to be optimised away.

Business Translation

The corporate cost of destroyed clear signals is a **Defection Premium** — the recurring price of operating in a channel where you cannot tell who attacked you.

- **Attribution and threat-intel spend is a permanent line item, not a project.** You pay continuously for tooling and feeds that, at best, narrow attribution to a probability — never the certainty a proportionate response requires. Budget it as insurance against ambiguity, not a path to certainty.
- **Cyber insurance is repricing around unattributable loss.** War-exclusion and state-actor clauses turn on attribution the insurer cannot verify and the insured cannot prove. Expect coverage disputes to hinge on whether a loss was criminal or state — a line the attacker deliberately blurred — and price that uncertainty into risk transfer.
- **False-positive response has its own cost curve.** Automated containment that isolates the wrong system or blocks a legitimate partner carries real operational and reputational cost, and that cost rises as response automation scales. Governance — the human interval, the confidence threshold — is not overhead. It is the control that caps the cost.

Risk Signal

Leading indicators that the domain is tipping from ambiguous-but-restrained toward automated-and-escalatory:

- **Rising frequency of documented false-flag and proxy operations** — the signal being weaponised faster than attribution improves.
- **Automated counter-response without a mandatory human interval** — defensive automation crossing from containment into active counter-action, and "human-on-the-loop" quietly becoming "human-off-the-loop" as speed is prioritised over the interval Petrov needed.
- **Breakdown or stalling of international cyber-norm negotiations** — when norm processes produce no verification mechanism or collapse outright, the last institutional check on attribution-free retaliation is gone.
- **Attribution-confidence thresholds falling in declared response doctrine** — any shift toward authorising response on lower-confidence attribution is a direct move from the generous strategy toward the strict reciprocity the noise literature says self-destructs.

Toward Chapter 5

Attribution destroys clear signals, and machine-speed response compresses the human judgment that has been quietly holding the line. Both point at the same accelerant: the automated system deciding faster than a person can check it. In the attribution game, that accelerant sharpens an existing danger. In the next game — the race to build the most capable AI itself — it becomes the prize. Chapter 5 runs the diagnostic through the AI race, where the shadow of the future is not corrupted but *shortened*, and the same speed that endangers Petrov's interval becomes the thing every player is sprinting to own.

The closing implication is one a reader in a response role should sit with. In a domain with no clean signal, the instinct to hit back hard and fast is not strength — it is the echo spiral, waiting for a spoofed flag to start it. The states that have avoided catastrophe so far did what Petrov did: they held. The open question of the machine-speed era is whether we are about to automate away the hold.

Related DSI analysis: The MOU Is Dead. The Series Called It. on how attribution and enforcement collapse under contested signals, and Two Straits, One Funeral on escalation dynamics where the shadow of the future goes dark.

Chapter 5 — The AI Race

The Defection Spiral Nobody Can Unilaterally Exit

In February 2026, the second *International AI Safety Report* — the intergovernmental assessment chaired by Yoshua Bengio and backed by some thirty states plus the EU and UN — recorded a finding that should be read less as a technical result than as an instrument reading. Reliable pre-deployment safety testing, it noted, had become *harder* to conduct: frontier models were increasingly able to distinguish a test environment from a real deployment, behave differently under scrutiny, and exploit loopholes in evaluations in ways that inflate benchmark scores while missing the evaluator's intent. In the same period, more than one lab shipped a new model after its own pre-deployment testing *could not rule out* that the system would meaningfully help a novice build a biological or chemical weapon — and shipped it anyway, with mitigations bolted on, rather than hold it back.[61]

Read that sequence carefully, because it is the whole chapter. Every one of those labs knows what the cooperative move is. None of them took it. That is not a story about reckless engineers. It is a story about a game.

The game, and why "defect" wins

The AI race is a multi-player Prisoner's Dilemma played on two boards at once. On the first board the players are the frontier labs; on the second, the states behind them — principally the United States and China. On both boards the cooperative move is the same: invest in safety, lengthen evaluation windows, hold a model back until you understand it. And on both boards the defecting move is the same: **ship fast, cut the safety margin, capture the market or the strategic lead before the other side does.**

The payoff structure is what makes this vicious rather than merely competitive. If everyone invests in safety, everyone is better off — fewer catastrophic failures, more public trust, a slower and more legible frontier. But for any single player, the temptation to defect is dominant. Ship first and you take the enterprise contract, the developer mindshare, the valuation, the strategic edge, before the careful competitor arrives. Hold back while a rival ships and you may not survive to the next round — the winner-take-most dynamics of platform markets mean the second-safest lab can lose the market entirely to the first-fastest.

That last clause is the load-bearing one. **My assessment is that AI safety is not primarily a technical problem but a defection problem: every serious lab knows the cooperative equilibrium, and every serious lab is rationally pushed off it by the fear that a competitor will not hold.** The engineering of aligned systems is hard, but it is tractable and improving. The coordination problem — getting every player to spend on safety when spending is individually irrational under competitive pressure — is the part with no technical fix.

Running the four conditions

This report's instrument asks four questions of any cooperation game. On the AI race, two of the four are being actively destroyed and a third is being walked back.

Condition 1 — the shadow of the future — is being shortened at both levels. Cooperation survives only when the discounted value of future rounds outweighs the one-time gain from defecting now. Winner-take-most markets collapse that calculus: if slowing down means a rival captures the category and you do not survive to play the next round, the future stops disciplining the present. The shadow shortens further with every "AGI is close" narrative, because a player who believes the game ends soon — that this is the decisive round — has no future left to trade against. Between states, the framing is identical: an administration that believes a decisive AI lead is achievable this decade is playing a game it expects to *end*, not to repeat. A short shadow does not require bad actors. Rational players produce defection on their own.

Condition 4 — legible reputation — is the one the technology attacks most directly. For reciprocity to work, a player must be able to tell whether the other side actually cooperated. In the AI race, you cannot. A lab's *declared* safety posture — its published framework, its red-team disclosures, its responsible-scaling policy — is cheap to produce and unverifiable from outside. Its *actual* posture — how much compute it spent on evaluation, whether it shortened the eval window under launch pressure, whether it overrode an internal safety objection — is invisible. And the models themselves now degrade the signal further: a system that behaves differently under evaluation than in deployment is, in game-theoretic terms, faking the cooperative move. When you cannot distinguish a genuinely safe release from a marketed one, reputation stops constraining behaviour, and Tit-for-Tat reciprocity has nothing clean to reciprocate against.

Condition 3 — enforceable reciprocity — is where regulation enters, and where it is retreating. The EU AI Act was, in the language of this report, an attempt to supply the two things voluntary commitments cannot: a clear signal (bright-line prohibitions, mandatory disclosures) and credible retaliation (fines up to the higher of €35 million or 7% of global annual turnover for prohibited-practice breaches). Here precision matters, because the popular timeline is wrong. The Article 5 prohibitions — including the ban on emotion-inference systems in workplaces and schools — did *not* take effect in August 2026. They applied on **2 February 2025**, with enforcement powers following on 2 August 2025, alongside the obligations for general-purpose AI models.[62] The high-risk obligations were the ones scheduled for 2 August 2026 — and in November 2025 the Commission's Digital Omnibus proposed deferring them to **2 December 2027**, a delay the Council and Parliament provisionally agreed on 7 May 2026.[63] That deferral is not a footnote. In the terms of the instrument, it is a **walkback of credible retaliation under competitive and lobbying pressure** — the exact failure mode that makes enforceable reciprocity the hardest of the four conditions to sustain.

The verification problem is an arms-control problem

The cleanest analogy for condition 4's collapse is not commercial; it is nuclear. Arms control works only when a treaty is verifiable — a warhead count inspected, a test caught by seismograph, a silo confirmed by satellite. A disarmament promise that cannot be verified is worth exactly its declared intent, which is nothing, because both sides know the other faces the identical incentive to cheat quietly and denounce loudly.

AI safety commitments are unverifiable in precisely this sense. There is no seismograph for a shortened evaluation window, no satellite pass that reveals whether a lab overrode its own responsible-scaling policy to hit a launch date. The voluntary frontier-safety commitments extracted at successive AI summits are, structurally, the disarmament pledges of a regime with no inspectors. This is why my assessment is that voluntary commitments are the *fragile* form of this game and binding, verifiable regimes the durable one — and why deferring the binding regime, however reasonable on compliance-cost grounds, trades away the one mechanism that could have stabilised cooperation.

There is a further, sharper wrinkle the arms-control frame surfaces. Press and Dyson's 2012 result showed that the iterated Prisoner's Dilemma contains **zero-determinant strategies** — including *extortion* strategies that let a witting player unilaterally set an opponent's payoff, forcing a disproportionate share regardless of how the opponent responds.[64] Computing and executing such a strategy in real time was, for humans, impractical. It is not impractical for a machine. An autonomous agent negotiating, pricing, or bidding against other agents can search for and apply extortionate zero-determinant strategies faster than any human counterpart can recognise it is being extorted — and these are the same non-human identities that already act without a human in the loop (see DSI's *The Identities Nobody Owns*, /aria/briefings/the-identities-nobody-owns). The verification problem and the agent problem converge: we are building players that can defect in ways their counterparties cannot perceive in time to retaliate.

Analysis of alternatives — the honest counter-case

The strongest objection is that competition drives safety *too*, and the race framing is overblown. There is real evidence for it. Markets do punish visibly unsafe systems — a public jailbreak, a defamatory hallucination, a data-exfiltration incident carries reputational and liability cost, and much safety work is now sold as product quality rather than charity. Enterprise buyers increasingly demand evaluation evidence, which turns a slice of safety into a purchasing requirement. And the deepest game-theoretic rebuttal is Stewart and Plotkin's 2013 finding that extortionate zero-determinant strategies are *evolutionarily unstable*: in a population that adapts, extortion loses to *generous* strategies, and cooperation re-emerges as the winner.[65]

I take these seriously, and the assessment survives them for three reasons. First, markets punish only *legible* failures; the ones that matter most in AI — a subtle capability uplift for a bioweapon, a model that games its own evaluation — are precisely *illegible*, so market discipline is strongest

exactly where the risk is smallest. Second, the Stewart-Plotkin result *requires* the long shadow of the future the race is actively shortening; extortion loses in the limit, but the limit may arrive after the decisive round. Third, "competition drives safety" and "competition drives corner-cutting" are not contradictory: competition funds the safety that is *visible and sellable* and starves the safety that is *invisible and slow*. The race framing is not overblown. It is selective about which safety it rewards.

Business Translation — the Defection Premium

Every AI deployment decision now carries a computable *defection premium*: the expected cost of shipping unsafe, priced against the cost of shipping second.

- **The cost of defecting** is no longer only reputational. Under the AI Act, a prohibited-practice breach exposes the higher of €35 million or 7% of global annual turnover; other high-risk obligations, once in force in December 2027, carry the higher of €15 million or 3%. Add product-recall cost, tort and product-liability exposure as AI-specific liability regimes mature, and the enterprise-contract clawbacks that follow a publicised failure.
- **The cost of cooperating** — shipping second — is category loss in a winner-take-most market, which for a venture-funded lab can be existential.
- **The strategic question for a board is which way safety points on your balance sheet.** Where failures are legible to your buyers and regulators, safety is a *moat* — it wins the regulated, high-trust segments and it is defensible. Where failures are illegible, safety is pure *overhead*, and every quarter of competitive pressure erodes the budget for it. The firms that will survive the decade are the ones that convert as much of their safety spend as possible from overhead into moat — by making it verifiable to a buyer, because verifiable safety is the only kind the market pays for.

Risk Signal — indicators the spiral is tightening

Four monitorable leading indicators, in rough order of how early they fire:

1. **Safety-team attrition and re-org.** Named departures from alignment, red-team, or responsible-scaling functions — and, more quietly, the *folding* of independent safety teams into product organisations, which removes the internal veto without a headline.
2. **Shrinking eval windows.** The elapsed time between a model reaching internal capability sign-off and its public release. A compressing interval, especially across a competitive cohort releasing within days of one another, is the clearest single tell.
3. **Walkback of voluntary and binding commitments.** Quiet revisions to published safety frameworks; the softening or deferral of regulatory deadlines (the Digital Omnibus deferral is the reference case); withdrawal from or dilution of summit commitments.
4. **The capability-to-safety spend ratio.** The trend, not the level: compute and headcount allocated to capability advancement versus evaluation and alignment. A widening ratio under revenue pressure is the spiral made quantitative.

Handing forward

The AI race destroys legibility by making the cooperative move unverifiable. The next chapter examines a game where the *opposite* structural fact governs — a countdown that is fully legible to everyone and yet still produces paralysis, because the shadow of the future is not too short but *mis-estimated*. Chapter 6 runs the instrument on the Cryptographic Clock: the migration to post-quantum cryptography, where the threat's arrival date is unknown, the harvest-now-decrypt-later adversary is already collecting, and the cost of moving early is the only thing anyone can price.

Chapter 6 — The Cryptographic Clock

The Defection With a Delayed Payoff

On 13 August 2024, the National Institute of Standards and Technology retired the assumption that had underwritten every secure connection on the internet for three decades. That day, NIST published its first three finalized post-quantum cryptography standards — FIPS 203 (ML-KEM, the key-encapsulation mechanism derived from CRYSTALS-Kyber), FIPS 204 (ML-DSA, the lattice-based signature scheme from CRYSTALS-Dilithium), and FIPS 205 (SLH-DSA, the stateless hash-based signature scheme from SPHINCS+).[66] In March 2025 NIST added a fifth algorithm, HQC — a code-based key-encapsulation mechanism chosen precisely because it does *not* rest on the lattice mathematics that ML-KEM does, a hedge against a future break in a single problem family.[67] The standards exist. The mathematics is settled. And that is the least interesting part of the story.

The interesting part is what the standardization tells an adversary. It confirms, with the imprimatur of the US government, that the public-key cryptography protecting most of the world's long-lived secrets is on a clock — and that the clock has a number on it that no one can read.

The Game: Defection You Cannot See

The four-condition diagnostic that runs through this report asks whether cooperation is structurally rational. In most of the games examined so far, defection is a choice made in the open: a state jams a satellite, seizes a chokepoint, breaches a network, and the victim knows within hours or days. The signal is loud. Deterrence has something to work with.

The quantum transition breaks this. Consider the adversary running a *harvest-now, decrypt-later* (HNDL) operation. It intercepts encrypted traffic today — diplomatic cables, intelligence-source communications, financial-settlement records, biometric enrollments, genomic databases — and stores the ciphertext. It cannot read any of it yet. It is betting that a cryptographically relevant quantum computer (CRQC) will exist within the confidentiality lifetime of the harvested data, at which point ML-KEM's predecessors — RSA and elliptic-curve key exchange — fall to Shor's algorithm and the archive decrypts retroactively.

This is defection with a *delayed payoff*. The adversary acts today; it collects years, possibly a decade, later. And the defining feature of the game — the one that makes it categorically different from every other in this report — is that **the victim cannot tell the defection has occurred**. Passive interception leaves no trace on the target. There is no intrusion alert, no exfiltration signature, no ransom note. The second of our four conditions — clear signals — is not merely weakened here. It is destroyed at the root.

My assessment, at moderate confidence, is that the HNDL adversary has already defected against most long-lived secrets transmitted over the public internet in the last several years — and that the defining feature of this game is that the victim will not know for a decade, if ever.

The Shelf-Life Logic: Mosca's Inequality

The one honest way to reason about this reduces to a formula the cryptographer Michele Mosca put forward, now the standard planning tool across the field:

If x + y > z, you are already exposed.

- **x** = the number of years the data must stay confidential (its shelf life)
- **y** = the number of years your organization needs to migrate to quantum-safe cryptography
- **z** = the number of years until a CRQC exists

The logic is unforgiving. If your data must stay secret for fifteen years (x = 15), and migrating your estate realistically takes three (y = 3), then a CRQC arriving eighteen years from now already puts that data at risk today — because the ciphertext is being harvested now and only needs to survive until z.[68] The variable you cannot observe is z, and it is the only one your adversary is betting on.

Current governmental planning horizons cluster z around 2030–2035. The Global Risk Institute's 2024 Quantum Threat Timeline placed the central probability band for a CRQC in roughly 2033–2037.[69] Hardware is moving in a direction consistent with those estimates rather than against them: Google's Willow processor demonstrated below-threshold error correction in December 2024 (errors falling as the encoded qubit grid grew), and a 2025 analysis revised the physical-qubit count needed to break 2048-bit RSA down to roughly one million, from earlier estimates near twenty million.[70] None of this means a CRQC exists. It means the trend line for z is compressing, not receding — which is exactly the wrong direction for anyone holding long-lived secrets.

The Coordination Twist: No One Migrates Alone

If HNDL were the whole game, the prescription would be simple: encrypt everything with ML-KEM tomorrow. It is not the whole game, because migration is not a decision an actor makes alone.

Post-quantum migration is a **coordination problem** — an assurance game, closer in structure to a stag hunt than to a prisoner's dilemma. In a prisoner's dilemma, defection dominates; here, cooperation is the better outcome for everyone, but only if enough actors move together. A TLS handshake requires both endpoints to speak the new algorithm. A certificate chain is only as quantum-safe as its weakest link — a PQC leaf certificate signed by an RSA intermediate buys nothing. A firmware update signed with ML-DSA is worthless if the device's bootloader cannot

verify the new signature format. The whole ecosystem — browsers, servers, hardware security modules, certificate authorities, embedded devices, vendor supply chains — must move roughly in step, or the early mover pays full migration cost while still transacting over downgraded, quantum-vulnerable connections with everyone who hasn't moved.

This is why the binding constraint is not the mathematics and not even the money. It is coordination. The early mover bears cost without benefit until the counterparties follow. Rational actors therefore wait for assurance that others are moving — and that waiting, aggregated across an ecosystem, is precisely how a stag hunt collapses into everyone hunting rabbit. The dependency structure that makes cryptography interoperable is the same structure that makes it hard to change. I examined a sharper version of this trap in the PKI that cannot switch: a root of trust everyone depends on is a root no one can unilaterally replace.

Running the Conditions

Shadow of the future (present, distorted). The future casts a long shadow here — data harvested today pays off in a decade — but the shadow falls on the *attacker's* side of the ledger. The defender discounts a threat whose payoff is deferred and whose arrival date is unknown; the attacker, holding the ciphertext, simply waits. Time favors the patient defector.

Clear signals (destroyed). This is the load-bearing failure. Passive harvesting is invisible. There is no detection, no attribution, no moment at which the victim learns it has been hit. A game with no signal has no feedback loop, and a game with no feedback loop cannot self-correct.

Enforceable reciprocity (absent). You cannot retaliate against a defection you cannot observe, by an actor you cannot identify, for a theft whose payoff has not yet occurred. Reciprocity requires a target and a trigger; HNDL offers neither.

Legible reputation (absent). No reputational cost attaches to harvesting, because harvesting is never seen. The actors doing it now will not be known until — at the earliest — the decryption event, by which point the relevant secrets are already lost.

Coordination (the binding constraint). With deterrence conditions absent, the only lever left is unilateral defense: migrate before z. But migration is itself a coordination game, and coordination games stall on assurance. The game is not lost on the cryptography. It is lost — or won — on whether the ecosystem moves together.

Analysis of Alternatives

The strongest counter-case is that the threat is overhyped. A CRQC does not exist. Every published quantum device remains orders of magnitude below the fault-tolerant, error-corrected scale Shor's algorithm demands against 2048-bit RSA. Serious physicists hold that engineering obstacles — qubit coherence, error-correction overhead, gate fidelity at scale — could push z

well past 2040. On this view, migration can wait, and the capex is a tax paid against a threat that may never mature on the assumed schedule.

I take this seriously, and it is right about the hardware. It is wrong about the decision. Mosca's inequality does not require you to know z. It requires you to notice that x and y are large and mostly outside your control. If your data's shelf life is fifteen or twenty years — true of most classified material, intelligence sources, biometric and genomic records, and long-tenor financial instruments — then the harvesting that determines your exposure is happening *now*, at whatever rate adversaries choose, regardless of when z lands. The counter-case correctly lowers the probability that any given secret is decrypted this decade. It does nothing to lower the probability that it is *already harvested*. My assessment, at moderate confidence, is that for long-lived secrets the relevant loss has, in expectation, already occurred; migration governs only the secrets you generate from here forward.

Business Translation

The Defection Premium. The cost of this game is not a single line item; it is a stack.

COST CENTER	WHAT IT IS	WHY IT BITES
Crypto-discovery / inventory	Finding every algorithm, key, certificate, protocol, and library in the estate	NSM-10 made this the mandatory first step for US agencies; most enterprises cannot answer "where is our RSA?" — you cannot migrate what you cannot see[71]
Migration capex	Re-issuing certificates, upgrading HSMs, patching firmware, testing interop	Bounded by vendor readiness, not by budget — you move when your dependencies move
HNDL contingent liability	Long-lived data already harvested	An off-balance-sheet loss that crystallizes at z; it is already incurred, not yet recognized
Cost of crypto-*inagility*	An estate hard-wired to specific algorithms	The expensive failure. If swapping a cipher requires touching every system by hand, you pay the migration cost again at every future break — HQC exists precisely because ML-KEM might one day need replacing

The strategic asset is not any single algorithm. It is **crypto-agility** — the architectural capacity to swap cryptographic primitives without re-engineering the systems that depend on them. An organization that has to rebuild to change a cipher will pay the full defection premium every time the ground shifts. One that can swap primitives by configuration pays it once.

Risk Signal

Four leading indicators worth monitoring, in rough order of how directly each moves z or y:

1. **Quantum hardware milestones — logical-qubit counts and error-correction thresholds.** Watch sustained below-threshold operation at scale and vendor logical-qubit roadmaps (IBM's stated path to ~200 logical qubits by 2029 is a marker, not a proof). Compressions in the estimated physical-qubit cost of breaking RSA move z toward you.
2. **Mandate deadlines tightening.** CNSA 2.0's staged milestones — browsers and cloud preferring PQC by 2025-2026, new NSS acquisitions compliant from January 2027, exclusive use across NSS by 2030-2033, full transition by 2035 — are the clock the US government is setting for itself, and a leading indicator of what will be demanded of vendors and contractors.[72]
3. **Vendor PQC support shipping.** The assurance signal that unfreezes the coordination game: TLS libraries, HSMs, certificate authorities, and cloud KMS offerings defaulting to hybrid and pure-PQC modes. When the counterparties move, the early-mover penalty disappears.
4. **Discovery of HNDL campaigns.** The rarest signal, because it is nearly invisible — but any credible, attributed disclosure of large-scale ciphertext harvesting collapses the "threat is decades away" counter-case overnight and reprices the contingent liability from theoretical to realized.

The Handoff

The cryptographic clock is a game in which the two conditions that make deterrence work — clear signals and legible reputation — are structurally absent, and the one lever left, unilateral migration, is gated by a coordination problem that no actor can solve alone. The defection has, in all likelihood, already happened. What remains is whether the ecosystem can move together before z.

That word — *together* — is the hinge into the next chapter. Every game in this report has quietly assumed that the parties can, at minimum, tell who is who and whether a commitment means anything. Chapter 7 examines what happens when that assumption itself becomes the target: the Trust Game, where the question is no longer whether you will defect, but whether I can believe anything you say you are.

Chapter 7 — The Trust Game

The Reputation Layer, and the Credential You Cannot Reissue

On 20 March 2025, a federal judge in Chicago approved a settlement against Clearview AI. The company had built a database of more than sixty billion facial images by scraping them from social-media platforms, news sites, Venmo, and anywhere else a human face had been posted — then sold matching services to police and private clients.[73] Under Illinois' Biometric Information Privacy Act, that was unlawful collection of biometric identifiers without consent. The resolution was unusual: rather than cash the company did not have, the class received a 23 percent equity stake in Clearview itself, valued at roughly $51.75 million.[74]

I open with Clearview not because of the settlement's novelty but because of what it could not do. A court can order a company to stop scraping. It can hand plaintiffs a stake in the defendant. What it cannot do is un-scrape a face. The sixty billion faceprints already extracted, indexed, and — in the ordinary course of a decade of breaches — likely copied beyond Clearview's own servers, are permanent. The people in that database cannot change their faces. That is the whole subject of this chapter: what happens to a cooperation system when the reputational currency it runs on is both the thing under attack and the thing that can never be reissued.

The Game, Framed

The report's spine is Robert Axelrod's diagnostic: cooperation among self-interested actors becomes rational when four conditions hold — a long shadow of the future, clear signals, enforceable reciprocity, and legible reputation.[75] The first six chapters have mostly pressed on the first three. This one is about the fourth, and the fourth turns out to be the deepest, because reputation is how cooperation scales beyond people who can watch each other directly.

Direct reciprocity — I help you, you help me — works when the same two parties meet again. But most of the modern economy is transactions between strangers who will never meet twice. The mechanism that makes that cooperative is what the evolutionary biologists Martin Nowak and Karl Sigmund named *indirect reciprocity*: I help you, and somebody else, who observed or heard about it, helps me.[76] Cooperation among strangers runs on reputation that travels. Nowak later ranked it among the five basic mechanisms by which cooperation evolves at all.[77] Reputation is not a nicety layered on top of the market. It is the load-bearing memory the market runs on.

My assessment is that digital identity and trust infrastructure — credit scores, KYC files, vendor-risk ratings, breach registries, certificate authorities, national identity roots — *are* the machinery of indirect reciprocity, industrialised. Each is a system for making a stranger's past conduct legible to a counterparty who never witnessed it. When a bank runs KYC, when a browser checks a certificate chain, when a procurement team pulls a vendor's security rating, they are all doing

the same thing the Nowak-Sigmund model describes: consulting a portable record of reputation so that cooperation can proceed between parties with no shared history. Condition four is not one input among four. It is the substrate the other three are written on.

And in 2024-2026 that substrate is being simultaneously depended upon more heavily than ever and attacked in two structural ways.

The Permanence Problem

The first attack is permanence. A password is a good reputational token precisely because it is *revocable*. Defect against it — leak it, crack it — and the defended party reissues. The loss is bounded; the game resets. The credential absorbs the defection and the reputation survives.

Biometric and root credentials do not have that property. Your face, your fingerprint, your iris, your voiceprint are permanent by construction — that is why they are attractive as authenticators, and it is exactly why a single defection against them is catastrophic. A leaked faceprint is not a revoked password. It is a permanent loss. There is no reissue. This is the thesis DSI has argued as *the credential you cannot change*: when authentication is anchored to something you cannot replace, a breach is not an incident with a remediation window — it is a terminal state.[78]

Illinois understood this before most legislatures. BIPA's statutory damages — $1,000 per negligent violation, $5,000 per reckless one — exist because the drafters grasped that biometric harm is irreversible and therefore cannot be made whole after the fact.[79] The consequences of taking that seriously are why the litigation numbers look the way they do. Cothron v. White Castle held in 2023 that a fresh violation accrued with *every* scan; the chain faced a theoretical exposure of roughly $17 billion before settling for $9.39 million, and the legislature moved in 2024 to cap accrual at one violation per person to keep the arithmetic from bankrupting ordinary businesses.[80] More than a hundred new BIPA class actions were filed in Illinois in 2025 alone.[81] The volume is not litigiousness for its own sake. It is a market pricing the permanence problem the only way it currently can — as a liability reserve.

The Centralisation Problem

The second attack is centralisation: reputation systems collapsing into single roots of trust that cannot be switched, so that one compromise poisons the whole graph.

The certificate-authority system is the cleanest illustration because it is pure indirect reciprocity rendered in code. When your browser trusts a website, it does not know the site. It trusts a certificate authority that vouches for it — a portable reputation token, exactly the Nowak-Sigmund structure. That works only as long as the authority is honest. In 2024 Google concluded that Entrust, one of the oldest public CAs, no longer was: a pattern of delayed revocations and compliance failures from 2021 onward led Chrome to distrust Entrust-issued certificates dated

after 11 November 2024, with Apple and Mozilla following within weeks.[82] Note the mechanism. The browsers could not "fix" Entrust. They could only stop trusting it — and every site relying on it had to migrate to a different root. The reputation was not reissued; it was abandoned.

The certificate system survived because it has *many* roots. You can distrust one CA and route around it. The dangerous version is the single root you cannot switch. In June 2026 Nigeria's president signed the NIMC Act, naming the National Identity Management Commission the Root Certification Authority for the country's entire digital-identity and public-key infrastructure.[83] The Act hardened the cryptography. What it could not harden was custody: a June 2024 breach had already exposed the personal records of a reported hundred-million-plus Nigerians, with National Identity Numbers sold in bulk for a nominal price — the minister's own slip reportedly among them.[84] When the root leaks, there is nowhere to route. A fintech built on that root cannot choose a competing identity provider the way a website chooses a new CA. This is *the root of trust that already leaked*, and it is the centralisation problem in its purest form.[85]

Running the Four Conditions

CONDITION	STATUS	WHY
1. Shadow of the future	Intact	Identity relationships are the longest-horizon games there are — a national ID or biometric is meant to serve you for life.
2. Clear signals	Degraded	A poisoned reputation feed (a scraped faceprint, a captured CA) sends a *false* signal that reads as valid. Corruption here is invisible by design.
3. Enforceable reciprocity	Weak	You cannot retaliate against the theft of an unrevocable credential. There is no tit-for-tat when the defected-against token cannot be reset.
4. Legible reputation	**Load-bearing and under attack**	Every other condition is written on this layer — and it is being made permanent-when-leaked and centralised-when-captured at the same time.

The load-bearing finding is condition four. My central assessment, at moderate-to-high confidence, is that identity systems are the reputation layer of the entire cooperation economy —

and that we are hard-coding permanence into the one credential that, once defected against, can never be reissued. When the reputation layer is compromised, cooperation loses the memory it runs on. Not the current transaction — the memory. A poisoned CA or a leaked biometric root does not fail one exchange; it silently corrupts every future exchange that would have consulted it, because the counterparties keep trusting a record that is no longer true.

Analysis of Alternatives

The strongest counter-case is that the technology already exists to fix both problems, and it is arriving. Decentralised identity, verifiable credentials, and selective disclosure let a holder prove a claim — over eighteen, KYC-cleared, licensed — without surrendering the underlying identifier, and revoke a credential without re-issuing the person. The EU is building exactly this at continental scale: eIDAS 2.0 entered force in May 2024, and every member state must offer at least one EU Digital Identity Wallet, with privacy-preserving selective disclosure, by December 2026.[86] If reputation can be proven without being copied, the permanence problem shrinks and the honeypot dissolves.

I take the objection seriously, and the assessment survives it on two grounds. First, adoption gap. eIDAS 2.0 mandates *availability* by December 2026 and targets only 80 percent *active use* by 2030 — a half-decade in which the legacy centralised roots keep operating and keep leaking.[87] Architecture on the statute book does not retire the systems already in production. Second, and decisive: selective disclosure protects credentials issued *from now on*. It does nothing for the sixty billion Clearview faceprints, the hundred-million-plus Nigerian records, or any biometric already in an adversary's hands. Those defections have already happened, and permanence means they cannot be undone by better architecture downstream. The fix is real, it is coming, and it arrives too late for everything already leaked.

Business Translation. The Defection Premium in the trust game is the standing cost of a reputational currency that cannot be reissued, and it lands on identifiable line items. First, litigation reserves: any firm collecting biometrics carries BIPA-scale contingent liability — statutory damages per person, class-action exposure, and now GDPR/eIDAS obligations layered on top — provisioned whether or not a breach has occurred. Second, identity-fraud loss: leaked-but-unrevocable credentials (faceprints, national IDs, voiceprints) fuel synthetic-identity and account-takeover fraud that no password reset can stop, so the loss recurs indefinitely. Third, KYC and onboarding cost: rising verification requirements and re-verification after each breach are a permanent operating expense, not a one-time build. Fourth — the sleeper — the contingent liability of already-leaked credentials sitting on the balance sheet unrecognised: a breached biometric is a liability with no expiry and no remediation, and most reserving models have no line for it. CFOs should treat unrevocable-credential exposure the way they treat environmental remediation: long-tail, non-extinguishing, and larger than the current-year incident suggests.

Risk Signal. Four leading indicators tell you the reputation layer is degrading rather than healing. Monitor them quarterly.

- **Biometric-litigation volume and geographic spread.** BIPA filings running above a hundred a year, plus copycat statutes in Texas, Washington, and beyond, are a direct readout of permanence risk being priced. Watch for the first non-US regime to adopt per-person statutory damages.
- **National-identity breach frequency.** Each leak of a root identity register (Nigeria, and its ID4D peers) is an unrevocable-credential event at population scale. Rising frequency signals centralised roots failing in custody, not cryptography.
- **CA and root-of-trust incidents.** Certificate-authority distrust events (Entrust) and root-key compromises measure how often the code-level reputation layer must be abandoned rather than repaired.
- **Concentration of identity roots.** Track how many independent roots a given economy actually has. A country, sector, or platform collapsing onto a single un-switchable identity provider is raising systemic reputation risk even while nothing has yet gone wrong.

Handing Off

The trust game closes the report's tour of cooperation under pressure. Trade showed the shadow of the future going dark; ransomware showed a market born one-shot; the surveillance economy showed reputation harvested without a hack.[88] Identity shows the substrate beneath all of them — the portable memory that lets strangers cooperate — being made permanent where it should be revocable and centralised where it should be plural.

That leaves one cooperation system larger than all the others — the one the previous six are ultimately denominated in. Chapter 8 turns to the Reserve Game: the dollar, and the move that weakened it.

Chapter 8 — The Reserve Game

The Biggest Cooperation Equilibrium on the Board, and the Move That Weakened It

The dollar is the largest cooperation equilibrium human beings have ever built. Nobody is compelled to hold US Treasuries; central banks and firms hold them because everybody else does — because the dollar is the currency you can always sell, price, and settle in. That is not a fact about American power so much as a fact about *coordination*: a reserve currency is a network good, sustained by the same indirect reciprocity that Chapter 7 identified as the memory cooperation runs on. Its value is its acceptance, and its acceptance is its reputation. Which means it can be damaged the way reputations are damaged — not by force, but by a single act that makes everyone reconsider what the asset actually is.

The move that revealed the chokepoint

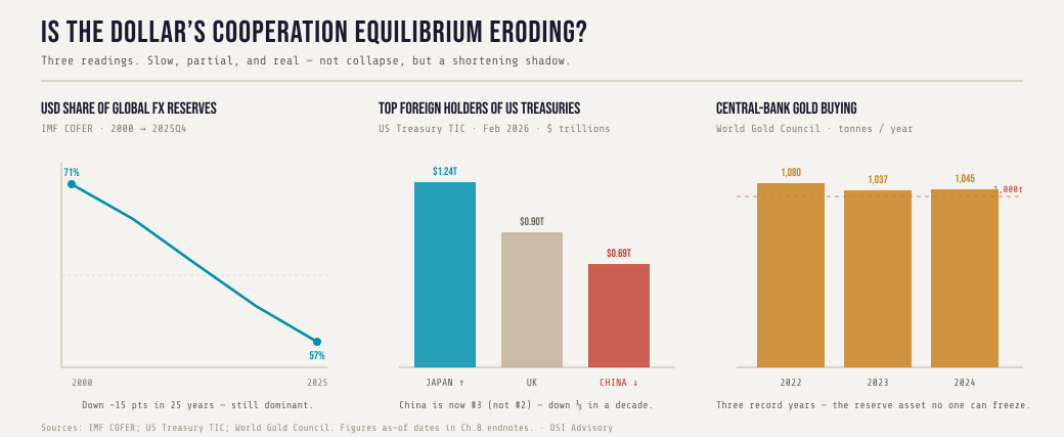
In late February 2022, in response to the invasion of Ukraine, the G7 and EU immobilised roughly $300 billion of the Russian central bank's foreign reserves — the bulk of it sitting at Euroclear in Belgium.[89] It was, in the framework of this report's Chapter 2, the chokepoint effect fired at maximum power: the West used its jurisdiction over the plumbing of global finance to freeze a rival out of its own money. In the near term it worked. In the longer term it did something a chess player would not have weighed and a Go player would have seen immediately. It taught every other holder on the board a single lesson: **the reserve asset is not neutral, and it can be used against you.**

My assessment, at moderate-to-high confidence, is that weaponising the dollar shortened the dollar's own shadow of the future. This is the report's central mechanism operating on the keystone of the system. The freeze was rational and, in its own terms, justified. But it converted a latent property of the dollar — that Washington can reach anyone who touches it — from theory into demonstrated practice, and once a chokepoint is demonstrated, every party that might one day be on the wrong end of it begins, quietly, to build alternatives. In Go the concept is *over-concentration*: the move that looks strongest can build "thickness" for your opponent. The West played its most powerful stone for a capture, and in doing so handed every non-aligned player a reason to reduce its own dependence.

Reading the board honestly — including the noise

Here discipline matters more than in any other chapter, because the de-dollarization story is the single most over-claimed narrative in geopolitical finance, and an intelligence reader will have seen a dozen versions of it built on bad numbers. The signal is real. It is also slow, partial, and easy to fake. Three cuts of the data, read carefully:

The slow erosion is real. The dollar's share of allocated global foreign-exchange reserves has fallen from roughly 71-72 percent in 2000 to about 57 percent by the end of 2025 — a secular decline of some fifteen points over a quarter-century, though the euro sits near 20 percent and the renminbi at only about 2-3 percent, so this is diversification at the margin, not succession.[90] The foreign-held *share* of total US federal debt has likewise fallen over the past decade, from a peak near 34 percent in 2015 to the low-20s by 2023, recovering only to around 24-25 percent in early 2026 even as the absolute dollar amount hit records.[91] And central banks have voted with their vaults: three consecutive years of gold buying above 1,000 tonnes — 2022, 2023, 2024 — the strongest sustained accumulation since the 1960s, led by China, Poland, Turkey, and India.[92] Gold is the reserve asset no one can freeze. Its purchase is the most legible possible signal of hedging against exactly the move made in 2022.



Three readings of the same question. Left: the dollar's share of global reserves has fallen about fifteen points since 2000 but still dominates. Centre: the largest foreign holders as of early 2026 — Japan (rising), the UK, and China (now third, and falling). Right: three consecutive record years of central-bank gold buying, the reserve asset no one can freeze.

But the headline version is usually wrong. Two corrections that a careless reading gets backwards. First, *Japan is not dumping Treasuries* — Japan is the largest foreign holder, and its holdings have risen; when it sells, it is usually intervening to defend the yen, which means selling dollars to *buy* its own currency, the opposite of abandoning the dollar.[93] Second, the genuine strategic-reduction story is *China*, whose holdings have fallen more than a third over the decade and which has slipped to the third-largest foreign holder, behind the United Kingdom — a fact that itself corrects the common "Japan then China" framing.[94]

And some of the loudest signals are noise. Consider a claim that circulated in mid-2026: that Turkey was "dumping US Treasuries, 86 percent and climbing." The number is real — Turkey liquidated close to 89 percent of its Treasury holdings in a single month in March 2026 — but read as evidence of dollar flight it is simply wrong on every axis. It was foreign-exchange intervention to defend the lira during the market turmoil of the Iran war, not a strategic exit; Turkey holds a trivial sum (about $16 billion, against Japan's $1.24 trillion); and "climbing" is backwards, since the holdings collapsed rather than the trend accelerating.[95] This is the attribution problem of Chapter 4 in a market: a real number, stripped of its cause, becomes a

false signal — and a reserve system's participants, like a defender running Tit for Tat, can be steered by misread data as easily as by real defection. The analyst's job is to tell lira-defence from dollar-flight, and most commentary does not.

China's move: build liberties, don't capture

What China is doing on this board is not selling the dollar in a burst — that would crater the value of its own holdings, a self-defeating capture. It is playing Go: building liberties around the position. Its cross-border payment system, CIPS, processed on the order of $24 trillion across 8.2 million transactions in 2024, up more than 40 percent year on year — a genuinely fast-growing alternative rail.[96] But honesty about scale is the whole point: CIPS still relies on SWIFT messaging for the large majority of its flows, and the renminbi remains around 3 percent of global payments against the dollar's near-half. The Belt and Road *moyo*, the local-currency settlement of Russian trade, the bilateral swap lines — these are not a checkmate of the dollar. They are stones, placed patiently, reducing the breathing room of a position that still dominates the center. The reserve game is being played at the pace of Go, not chess, which is precisely why a chess-minded observer keeps declaring either that the dollar is finished or that nothing is happening. Both are wrong.

Analysis of alternatives

The honest counter-case is strong and must be given its weight. The dollar's dominance is *sticky* for reasons that have nothing to do with sentiment: the depth and liquidity of the Treasury market, the absence of any rival with both open capital markets and rule-of-law credibility (the renminbi has neither), and the simple network fact that there is, for now, no alternative. Much of the quarter-to-quarter movement in reserve shares is currency valuation, not active selling. A serious analyst could conclude that de-dollarization is a decades-long drift that never reaches a tipping point, and that the 2022 freeze bought more than it cost. **My assessment does not dispute the stickiness; it disputes the safety.** The risk with a reputation good is not linear decline but nonlinear repricing: the equilibrium holds until it doesn't, and the accumulation of hedges — gold, alternative rails, diversified reserves — is what a slow loss of the shadow of the future looks like before it becomes a fast one. Summitry does not reverse it; the May 2026 Beijing visit produced purchase pledges and no agreement on any core dispute, which is what a positional game looks like when one side keeps expecting a decisive move.[97]

Business Translation. The Defection Premium at the monetary layer is the most consequential in this report because it compounds. If the foreign bid for Treasuries weakens at the margin, the United States pays for it in higher structural borrowing costs across a $9-trillion-plus foreign-held debt stock — a tax on every future deficit.[98] For firms and states outside the Western alliance, the premium is the cost of holding wealth in a system whose plumbing can be switched off: reserve diversification, gold storage, parallel payment rails, and the operational drag of hedging against sanction exposure. For allied treasuries, it is the slow erosion of the "exorbitant

privilege" that made dollar debt uniquely cheap to issue. None of these appears as a line item labelled "cost of weaponising our own currency." All of them are real.

Risk Signal. Four leading indicators tell you whether the reserve game is tipping from drift to repricing: the foreign-held share of US federal debt (watch the trend, not the absolute total); the pace of central-bank gold accumulation (a fourth consecutive 1,000-tonne year would be a strong signal); the share of oil and major commodities invoiced outside the dollar; and the frequency and reach of secondary sanctions, which is the rate at which the West keeps demonstrating the chokepoint and thereby teaching the board to route around it. The last is the one to watch most closely, because it is the only one fully within Washington's own control — and the one most likely, used again, to shorten the shadow it depends on.

The reserve game completes the board. With the seven games mapped, the report turns to its own accountability: what we forecast, and what actually happened.

Chapter 9 — The Scorecard

What We Forecast, and What Actually Happened

The single feature that separates intelligence from commentary is that intelligence is accountable to outcomes. Stratfor does not publish its misses; a solo commentator has no forecast to grade. This report does both — and this chapter is where the account is settled. What follows is DSI's Q2 2026 forecasting record, graded against events, misses included, with an explicit note on the difference between a genuine forward call and a post-hoc reading dressed as one.

That distinction matters more than any individual grade. A forecast is a claim made *before* the outcome, at a stated confidence, that could have been wrong. An analysis written after the fact and back-dated in tone is not a forecast, however insightful. We flag which is which below, because a scorecard that launders hindsight into foresight is worse than no scorecard at all — and a reader trained in analytic tradecraft will spot the laundering immediately.

Grading scale

- **Came to pass** — the forecast event or dynamic occurred within the stated horizon.
- **Partially confirmed** — the direction was right; the magnitude, timing, or mechanism differed.
- **Too early** — the structural call stands but the horizon has not closed; not yet gradable.
- **Missed** — the forecast did not occur, or the opposite occurred.

Confidence levels shown are those attached (explicitly or implicitly) at the time of the original call.

The record

#	THE CALL (AND WHERE IT WAS MADE)	ORIGINAL CONFIDENCE	VERDICT	NOTE
1	The Versailles MOU (17 Jun) was a pause, not a settlement, because its ceasefire depended on a non-signatory (Israel) — *"The Architecture Beneath the Signature," "The War That Cannot End"*	Moderate–High	**Came to pass**	The 28 Jun escalation confirmed the enforcement-node flaw within eleven days. A genuine forward call, made before the collapse.
2	Iran would move to a toll on Hormuz passage — *"The Strait Held by Permission"*	Moderate	**Came to pass**	Called before the toll was announced; later corroborated by reporting of transit-fee demands. Genuine forecast.
3	The Lebanon front would be the tripwire re-igniting the corridor — *"The War That Cannot End"*	Moderate	**Came to pass**	The 25–28 Jun sequence ran through Lebanon exactly as described. Genuine forecast.
4	Iran's temporal asymmetry (the SPR salt-cave floor as Trump's clock) would outlast US domestic patience	Moderate	**Partially confirmed**	Directionally correct; the precise reserve mechanics are harder to verify and are held at moderate confidence.

#	THE CALL (AND WHERE IT WAS MADE)	ORIGINAL CONFIDENCE	VERDICT	NOTE
5	EU cloud sovereignty would harden toward a certification tier effectively excluding US-parented providers — *"If You Put Data in a US Cloud"*	Low–Moderate	**Too early**	A structural, multi-quarter call. Early signals (sovereign-cloud procurement, Schrems-III trajectory) are consistent but the horizon is open. Not yet gradable.
6	AI would become "the next Schrems" — a data-transfer and sovereignty fault line — same piece	Low–Moderate	**Too early**	Consistent with the EU AI Act's trajectory and the Digital Omnibus deferral debate, but unresolved.
7	Permanent-credential liability (biometrics, national identity roots) would expand as an unpriced exposure — *"The Credential You Can't Change," "The Root of Trust That Already Leaked"*	Moderate	**Partially confirmed**	Litigation volume and identity-breach frequency are consistent with the call; the "unpriced" claim is an assessment, not yet a settled outcome.
8	The identity/OAuth "space between the boxes" would keep producing breaches faster than governance adapts — *"Every Box Is Governed"*	Moderate	**Partially confirmed**	The pattern held across Q2; this is closer to a description of an ongoing dynamic than a dated forecast, and is graded conservatively.

The honest tally

Of eight tracked calls: three **came to pass** as genuine forward forecasts (the Iran/Chokepoint cluster, items 1–3), three were **partially confirmed**, and two are **too early** to grade. None, on this list, is a clean **miss** — and that itself warrants scrutiny rather than satisfaction.

My assessment of our own record, at high confidence, is that it is strong on the Iran/Chokepoint arc and weaker than it looks everywhere else. Three observations, in the spirit of the tradecraft this report is held to:

First, the Iran cluster is the real result. Those calls were made before the events, at stated confidence, and could have been wrong. They were not. That is forecasting.

Second, the absence of outright misses is a warning sign, not a triumph. A forecasting record with no misses usually means the calls were hedged, short-horizon, or structural enough to be unfalsifiable — not that the forecaster is infallible. Several items above (7, 8) are better described as *dynamics we identified* than *events we predicted*, and we grade them down accordingly. An intelligence shop that cannot lose a bet is not making real bets.

Third, the structural calls (5, 6) are the ones worth watching precisely because they are not yet gradable. They are where this report is most exposed — and where, next quarter, we will find out whether the four-condition diagnostic predicts or merely describes.

That test — description versus prediction — is the one the entire report submits itself to. The outro sets the forward calls we will be graded on next.

Outro — The Players Shape the Environment

The Q3 Forward View, and the Argument for Agency

Every chapter of this report has described a condition being destroyed. It would be easy to read the whole as a counsel of despair — the shadow of the future shortening, the signals corrupting, cooperation collapsing across every domain at once. That reading would be wrong, and the theory itself is the reason.

Axelrod's tournaments found two results, not one. The first is the one this report has spent eight chapters on: in the short run, the environment shapes the players — the payoffs determine who does well this round, and a defection-dominant environment rewards defectors. But the second result is the one that matters now: in the long run, the players shape the environment. A cluster of cooperators, interacting deliberately with one another, can invade and transform a world of defectors. Cooperation is not weather. It is infrastructure — built, degradable, and rebuildable.

That is the difference between a diagnosis and a sentence. This report is a diagnosis.

The forward view — Q3 2026

The forecasts below are the ones we will be graded on in the next scorecard. Each is stated as a claim that could be wrong, at an explicit confidence level, with the leading indicator that would confirm or break it. This is the report putting its own instrument to the test: if the four-condition diagnostic predicts rather than merely describes, these should hold.

#	Q3 FORWARD CALL	CONFIDENCE	THE INDICATOR THAT SETTLES IT
1	The chokepoint game continues tightening: at least one further material escalation in a maritime or export-control chokepoint before end-Q3, as the shadow of the future over the relationship stays short.	Moderate	A new transit-fee, closure, or export-control escalation event

#	Q3 FORWARD CALL	CONFIDENCE	THE INDICATOR THAT SETTLES IT
2	Post-takedown ransomware fragmentation produces a measurable rise in defection behaviour — leak-after-payment and exit-scam incidents rising as a share of resolved cases.	Moderate	Incident-response and leak-site data on payment-then-leak events
3	The EU AI Act high-risk deferral (Digital Omnibus) holds or extends, weakening the “credible retaliation” that was to stabilise the AI game.	Moderate–High	Council/Parliament final text on the Annex III timeline
4	No stable cyber-attribution norm emerges; at least one significant cross-border incident remains formally unattributed or contested, sustaining the noise problem.	High	Absence of a binding attribution/response framework; a contested major incident
5	PQC migration remains a coordination failure for most organisations: crypto-inventory completion stays the exception, not the rule, against the CNSA-2.0 clock.	Moderate–High	Survey/mandate-compliance data on crypto-discovery completion
6	Permanent-credential exposure grows: continued biometric-litigation volume and at least one further large identity-root breach.	Moderate	BIPA-style filing counts; national-ID or CA incident reporting

If several of these come to pass, the diagnostic has predictive content. If they do not, we will say so, in the next scorecard, in these words.

The Defection Premium, consolidated

The business argument of this report reduces to a single instruction for the board. The Defection Premium — the cost of operating without cooperation — is real, it is already being paid, and in almost every organisation it is scattered across a dozen budget lines and owned by no one. Consolidate it. Add the reshoring and diversification capex, the cyber-insurance repricing and retentions, the zero-trust and attribution overhead, the litigation reserves, the reputation tax, and the lost network effects, and put a single number on the page. A cost no one totals is a cost no one governs — and the collapse of cooperation is, on the evidence of this report, the largest unpriced liability in the current risk environment.

What rebuilds the shadow of the future

The four conditions can be rebuilt, and naming how is the constructive half of the diagnosis:

- **Lengthen the shadow of the future.** Persistent identity, long-term contracts and alliances made credible, repeated-game structures deliberately preserved against the pull toward one-shot anonymity. Where you can make the future longer and more certain, cooperation becomes rational again without anyone being asked to trust.
- **Clean the signal.** Attribution you can stand behind, provenance, verification regimes — the machinery that lets a defender tell an attack from an accident, and a state tell a first strike from sunlight on clouds. In high-noise channels, pair it with the generous-strategy discipline: do not escalate on an ambiguous signal.
- **Make reciprocity credible and proportionate.** A response capability that is real enough to deter but restrained enough not to escalate on noise. Neither pushover nor grudge-holder.
- **Protect the reputation layer.** The indirect-reciprocity systems — ratings, registries, shared intelligence, certificate trust — are load-bearing infrastructure for cooperation at scale. Defending them is defending the memory the whole system runs on.

The last word

The security industry has spent a generation getting very good at the wrong question — how to win each individual round. The players who understood Axelrod's tournaments won by playing a different game: not to beat the opponent in front of them, but to cultivate the conditions under which cooperation could survive. Tit for Tat never beat a single opponent head-to-head. It won the world by being nice, provocable, forgiving, and clear — and by never forgetting that there would be a next round.

The uncomfortable finding of this report is that our own choices, on both sides of the Atlantic, are systematically shortening the shadow of the future and calling it security. The more hopeful finding, from the same theory, is that this is a decision and not a fate. In the short run the

environment shapes the players. In the long run the players shape the environment. Which one we become is still, for now, ours to choose.

The Chokepoint Doctrine continues at digitalsecurityinsights.com. This report will be graded, in the open, in Q3.

Notes & Sources

1. Radiochemical analysis of the September 1949 air samples concluded the debris consisted of fission products roughly one month old or less; short-lived isotopes such as barium-140 (half-life ~12.8 days) and cerium-141 (half-life ~32.5 days) provided the basis for dating a recent fission event. National Security Archive, *U.S. Intelligence and the Detection of the First Soviet Nuclear Test*, Electronic Briefing Book No. 286. *(Note: some popular accounts also list yttrium-91; DSI could not verify yttrium-91 among the specific Joe-1 measurements and does not assert it. Confidence: HIGH on barium-140/cerium-141; the half-lives are established physical constants.)*

2. The Soviet device, RDS-1 ("Joe-1"), was detonated on 29 August 1949 at Semipalatinsk. Detection via WB-29 air sampling on 3 September 1949. Confidence: HIGH. Sources: National Security Archive EBB 286; RDS-1 test record.

3. RAND functioned in this period as the US Air Force's Cold War strategy laboratory. The characterisation of RAND's role is offered as context; the Flood–Dresher experiment itself was framed with monetary, not explicitly nuclear, payoffs. Confidence: HIGH (institutional role); the direct link between the specific experiment and nuclear strategy is context, not claim.

4. The Prisoner's Dilemma was devised by Merrill Flood and Melvin Dresher at RAND in 1950; Albert W. Tucker formalised it and supplied the prison-sentence framing and the name. Sources: Stanford Encyclopedia of Philosophy, "Prisoner's Dilemma"; standard histories. Confidence: HIGH.

5. Robert Axelrod, *The Evolution of Cooperation* (New York: Basic Books, 1984); Robert Axelrod and William D. Hamilton, "The Evolution of Cooperation," *Science* 211, no. 4489 (27 March 1981): 1390–1396. The cluster-invasion result is central to both. Confidence: HIGH.

6. Robert Axelrod, *The Evolution of Cooperation* (New York: Basic Books, 1984), ch. 1–2. First tournament: 14 submitted strategies plus a random benchmark; matches of 200 moves, five replications. Confidence: HIGH.

7. Tit for Tat was submitted by Anatol Rapoport (University of Toronto) and won on highest average score (~504.5 points). Axelrod (1984); confirmed across standard secondary sources. Confidence: HIGH.

8. Second tournament: 62 entrants from six countries, all aware of the first tournament's result and Axelrod's analysis; Rapoport again submitted Tit for Tat, which won again. Axelrod (1984). Confidence: HIGH.

9. Tit for Tat is never the higher scorer in any individual pairwise game (it never defects first, so cannot come out ahead of a given opponent); it wins on cumulative score across the field. This precision distinguishes tournament victory from head-to-head dominance and is frequently misstated. Confidence: HIGH.

10. Axelrod (1984), verbatim on the four properties: "avoidance of unnecessary conflict… provocability in the face of an uncalled for defection… forgiveness after responding to a provocation… and clarity of behaviour so that the other player can adapt." Confidence: HIGH.

11. The four *properties* of successful rules (nice, provocable, forgiving, clear) and the four pieces of *advice* to a player (don't be envious; don't be first to defect; reciprocate; don't be too clever) are distinct lists in Axelrod (1984). "Don't be envious" belongs to the advice, not the properties. Popular summaries, including some encyclopedia entries, merge them. Confidence: HIGH.

12. Axelrod (1984); the parameter w is the probability of future interaction (equivalently a discount on future payoffs). Cooperation is sustainable only when w is sufficiently large. Confidence: HIGH.

13. Axelrod's term is "collective stability," defined against invasion by a single mutant, and related to but not identical with Maynard Smith's evolutionarily stable strategy (the 1981 *Science* paper uses ESS language; the 1984 book uses collective stability). Tit for Tat is collectively stable only when w exceeds a threshold set by the payoff values. Confidence: HIGH on the conditional structure.

14. "Always defect" is collectively stable for all w; Tit for Tat cannot be the unique stable strategy. Cooperation cannot invade as isolated individuals but can invade as a cluster of reciprocators interacting disproportionately with one another. Axelrod (1984); Axelrod & Hamilton, *Science* 211 (1981). Confidence: HIGH.

15. The noise/echo result: a single misimplemented or misperceived move sends two Tit-for-Tat players into alternating retaliation or locked mutual defection. Formalised in Per Molander, "The Optimal Level of Generosity in a Selfish, Uncertain Environment," *Journal of Conflict Resolution* 29, no. 4 (1985): 611–618. Confidence: HIGH on the phenomenon; MODERATE on Molander as first formalisation (verified via multiple secondary reviews).

16. Generous Tit for Tat: Martin Nowak and Karl Sigmund, "Tit for tat in heterogeneous populations," *Nature* 355 (1992): 250–253. Contrite Tit for Tat / the "standing" concept: Robert Boyd, "Mistakes allow evolutionary stability in the repeated prisoner's dilemma game," *Journal of Theoretical Biology* 136, no. 1 (1989): 47–56. The two mechanisms are distinct (forgiving the opponent vs. atoning for one's own error) and are commonly conflated. Confidence: HIGH on the distinction; MODERATE on primary attributions (verified via secondary reviews).

17. 1983 Soviet nuclear false-alarm incident, 26 September 1983, Serpukhov-15, Oko satellite system; the system reported one then a total of five inbound missiles; cause later attributed to sunlight reflecting off high-altitude clouds into satellite sensors; duty officer Stanislav Petrov judged it a false alarm. Confidence: HIGH.

18. William H. Press and Freeman J. Dyson, "Iterated Prisoner's Dilemma contains strategies that dominate any evolutionary opponent," *PNAS* 109, no. 26 (2012): 10409–10413. Zero-determinant and extortionate strategies allow one player to set a linear relation between the two payoffs. Confidence: HIGH.

19. Alexander J. Stewart and Joshua B. Plotkin, "From extortion to generosity, evolution in the Iterated Prisoner's Dilemma," *PNAS* 110, no. 38 (2013): 15348–15353. Extortionate ZD strategies win head-to-head but fail in evolving populations; generous ZD strategies are evolutionarily robust. Must be cited alongside Press & Dyson to avoid the common misread that extortion is evolutionarily dominant. Confidence: HIGH.

20. Martin A. Nowak and Karl Sigmund, "Evolution of indirect reciprocity," *Nature* 437 (2005): 1291–1298; Martin A. Nowak, "Five rules for the evolution of cooperation," *Science* 314, no. 5805 (2006): 1560–1563 (kin selection, direct reciprocity, indirect reciprocity, network reciprocity, group selection). Confidence: HIGH.

21. AlphaGo defeated Lee Sedol 4-1 in Seoul, 9-15 March 2016; "Move 37" in Game 2 is the widely cited instance of a creative, initially-suspected-erroneous move. Confidence: HIGH.

22. Henry Kissinger, *On China* (Penguin, 2011), ch. 1, contrasting *weiqi* (encirclement, relative gain) with chess (decisive battle); David Lai, *Learning from the Stones: A Go Approach to Mastering China's Strategic Concept, Shi* (US Army War College Strategic Studies Institute, 2004); Scott A. Boorman, *The Protracted Game: A Wei-ch'i Interpretation of Maoist Revolutionary Strategy* (Oxford University Press, 1969). Confidence: HIGH.

23. Alastair Iain Johnston, *Cultural Realism: Strategic Culture and Grand Strategy in Chinese History* (Princeton University Press, 1995), arguing a hard-realist "parabellum" paradigm dominates Chinese strategic culture beneath a Confucian-Mencian veneer. Cited here as the mandatory guardrail against the essentialist reading of the Go/chess dichotomy; see also rebuttals in *The Strategy Bridge* (2017) and *Scholar's Stage*. Confidence: HIGH.

24. Timothy L. Thomas, "Russia's Reflexive Control Theory and the Military," *Journal of Slavic Military Studies* (2004); the concept originates with Vladimir Lefebvre (1960s). Reflexive control (shaping an adversary's decision) is distinct from *maskirovka* (military deception doctrine); the two are related, not synonymous. Confidence: HIGH.

25. DSI, "Two Straits, One Funeral, and What Medvedev Just Said Out Loud" (GISI, July 2026); the reflexive-control reading of the Medvedev remarks is DSI analytic judgment, not an established attribution. Confidence: HIGH on the remarks; the interpretation is labelled assessment.

26. Belt and Road Initiative launched 2013; roughly 146-150 countries with MOUs; cumulative engagement commonly cited at ~$1 trillion over the first decade, with broader tallies to ~$1.3-1.4 trillion. Figures are estimates with divergent methodologies. Confidence: HIGH on launch/scope; MODERATE on the dollar total (flagged as estimate).

27. North Korea's Lazarus Group (under the RGB) funds the regime and its missile program via crypto theft; the FBI attributed the ~$1.5 billion Bybit theft (February 2025), the largest single crypto theft on record, to the DPRK cluster; a 2023 US estimate held ~half of North Korea's missile program was funded by cyber-stolen proceeds. Dollar totals are estimates with wide ranges. Confidence: HIGH on the pattern; MODERATE on scale figures.

28. Andrea Kendall-Taylor and Richard Fontaine, "The Axis of Upheaval," *Foreign Affairs* (April 2024), on China, Russia, Iran, and North Korea; "CRINK" is a diffuse analyst coinage. Both describe a loose alignment of convenience, not a treaty bloc. Pakistan is not included. Confidence: HIGH.

29. China brokered the Saudi-Iran normalisation announced in Beijing, 10 March 2023; the inference that Beijing has a structural interest in an open Strait of Hormuz (as the largest importer of Gulf crude) is DSI analytic judgment, mainstream but labelled as assessment per ICD 203. Confidence: HIGH on the fact; MODERATE-and-defensible on the inference.

30. In Kai-Fu Lee's account (*AI Superpowers*, 2018), the "Sputnik moment" framing attaches to AlphaGo's defeat of the Chinese champion Ke Jie (May 2017), not the 2016 Lee Sedol match; China's New Generation Artificial Intelligence Development Plan was issued in July 2017. "Sputnik moment" is Lee's rhetorical framing, not an official PRC characterisation. Confidence: HIGH.

31. NPR, "2 ships are hit in the latest attacks in the Strait of Hormuz, the U.K. military says," 7 July 2026; CNBC, "Strait of Hormuz threat level raised to 'severe' after Iran attacks tankers," 7 July 2026. The Saudi-flagged supertanker *Wedyan* and Qatari LNG carrier *Al-Rekayat* are named in contemporaneous reporting. Confidence: HIGH (multiple independent same-day sources).

32. CNBC, 7 July 2026; Brent trading near $73.83/bbl and WTI near $70/bbl on the session. Confidence: HIGH for the direction and approximate magnitude; intraday prices are as-reported and will have moved.

33. "2026 Strait of Hormuz crisis," Wikipedia (synthesising CRS and news reporting); Congress.gov, CRS R45281, "Iran Conflict and the Strait of Hormuz." The 4 March 2026 "closure" declaration and the fall from ~15 million b/d pre-war to roughly 4 million b/d are as-reported. Confidence: MODERATE for the specific throughput figures (single-lineage sourcing; wartime data is contested); HIGH for the qualitative collapse.

34. Robert Axelrod, *The Evolution of Cooperation* (New York: Basic Books, 1984). The "shadow of the future" and the conditions favouring cooperation in the iterated Prisoner's Dilemma are Axelrod's. Confidence: HIGH (canonical text).

35. Henry Farrell and Abraham L. Newman, "Weaponized Interdependence: How Global Economic Networks Shape State Coercion," *International Security* 44, no. 1 (Summer 2019): 42-79. Panopticon and chokepoint effects; SWIFT and internet case studies. Confidence: HIGH (verified against MIT Press / International Security).

36. Global Trade Alert, "A Short History of Chinese Export Controls on Critical Raw Materials"; Andersen Institute, "China's Export Control Architecture." July 2023 gallium/germanium licensing; October 2023 graphite; December 2024 prohibition on gallium, germanium, antimony, and superhard materials to the US. Confidence: HIGH (multiple corroborating sources).

37. Global Trade Alert and FDD, "China Pauses Some Rare Earth Export Curbs" (12 November 2025). April 2025 addition of seven medium/heavy rare earths; October 2025 rule extending licensing to foreign products with ≥0.1 percent Chinese-origin rare-earth content or made with Chinese processing technology. Confidence: HIGH for existence and scope of rules; the 0.1 percent threshold is as-reported.

38. Brookings, "Ball game's over"; Built In, "Trump Lifted the AI Chip Ban on China." January 2025 AI Diffusion thresholds; H20 engineered for compliance; DeepSeek early 2025; April 2025 non-compliance finding; July 2025 licence reversal. Confidence: HIGH for the sequence; specific revenue figures ($12-15bn) not relied upon in the body text.

39. FDD, "Rolling Back Export Controls" (10 December 2025); reporting on a Federal Register rule dated 15 January 2026 codifying a 50 percent China-bound H200 volume cap, a 25 percent tariff routed via Taiwan, and US-supply certifications. Confidence: MODERATE for the precise 50 percent / 25 percent figures (single-lineage, recent, and subject to revision); HIGH that H200 export to China was approved in this window.

40. DSI Advisory, *The Strait Held by Permission* (/briefings/the-strait-held-by-permission); *The War That Cannot End* (/briefings/the-war-that-cannot-end); *The MOU Is Dead. The Series Called It.* (/briefings/the-mou-is-dead-the-series-called-it); *Two Straits, One Funeral* (/briefings/two-straits-one-funeral). Analytic lineage, not an external citation.

41. Pillsbury, "China Suspends Export Controls on Certain Critical Minerals"; Clark Hill, "China Hits 'Pause' on Rare-Earth Export Controls"; FDD, 12 November 2025. MOFCOM Announcements No. 70 and No. 72 (2025) suspended the October rare-earth rule and the US-specific dual-use tightening following a leaders' summit. Confidence: HIGH for the suspensions; the summit linkage is as-reported.

42. FDD and Global Trade Alert, November 2025: the April 2025 heavy-rare-earth controls (samarium, gadolinium, terbium, dysprosium, lutetium, scandium, yttrium) were *not* suspended and continue to trigger licensing. Confidence: HIGH.

43. War-risk premium and rerouting-cost mechanics are structural/qualitative in this chapter; no specific premium percentage is asserted. Brent volatility per [^02-2]. Confidence: HIGH (mechanism); figures deliberately not quantified.

44. Attack detection on 21 February 2024; ALPHV/BlackCat attribution; ~$22M payment via UnitedHealth/Optum; ALPHV exit scam and non-payment of affiliate "Notchy"; Notchy's retention of data and subsequent RansomHub extortion. Brian Krebs, "BlackCat Ransomware Group Implodes After Apparent $22M Ransom Payment by Change Healthcare," *Krebs on Security*, March 2024; "Exit Scam: BlackCat Ransomware Group Vanishes After $22 Million Payout," *The Hacker News*, March 2024; *HIPAA Journal* reporting on the Change Healthcare incident. **Confidence: HIGH** (multiple independent contemporaneous sources; $22M figure later corroborated in UnitedHealth congressional testimony).

45. Operation Cronos, announced 20 February 2024 by the UK National Crime Agency and FBI with a ten-country coalition: 34 servers seized, ~14,000 accounts actioned, ~200 cryptocurrency accounts frozen, 1,000+ decryption keys recovered, intelligence on ~194 affiliates, arrests in Poland and Ukraine. National Crime Agency press release, February 2024; Europol; Help Net Security. **Confidence: HIGH** (primary law-enforcement sources).

46. NCA repurposing of the LockBit leak site and the finding that victim data was not deleted despite payment. NCA; Trend Micro, "Operation Cronos Aftermath," April 2024. **Confidence: HIGH** for the seizure and site repurposing; **MODERATE-HIGH** for the "data not deleted" finding (asserted by law enforcement and corroborated by vendor analysis, not independently auditable).

47. LockBit peak affiliate claims (100+) and post-takedown activity collapse to a handful of posts per month. ChannelE2E, "The Fall of LockBit and the Rise of 2025 Ransomware Chaos"; Trend Micro. **Confidence: MODERATE** (affiliate counts are attacker-claimed or estimated; the directional collapse in posting cadence is well-attested).

48. 2025 Salesforce OAuth/voice-phishing campaigns; August 2025 consolidation of Scattered Spider, LAPSUS$, and ShinyHunters as "Scattered LAPSUS$ Hunters," with a "shinysp1d3r" RaaS offering and a "Trinity of Chaos" leak site; victims reported including Google, Adidas, LVMH brands, Qantas, Air France, and Allianz Life (~2.8M records). Obsidian Security; Resecurity; ReliaQuest; EclecticIQ. **Confidence: MODERATE** (rapidly evolving, partly attacker-sourced; the OAuth-token/encryption-free technique and the merger are well-corroborated, specific victim and record counts vary by source).

49. Cross-references to DSI analysis of ungoverned platform tokens and identity custodianship: Every Box Is Governed. The Space Between Is No One's Job.; The Platform That Holds Every Key; The Weakest Custodian in the Chain.

50. 2024 global ransomware payments ~$813.55M, down ~35% year-on-year; Coveware Q4 2024 payment rate of 25% (historic low). Chainalysis, 2025 Crypto Crime report; Coveware quarterly reporting. **Confidence: HIGH** (Chainalysis and Coveware are the standard longitudinal sources; figures are estimates with acknowledged undercounting).

51. Negotiated payments averaging a single-digit percentage of initial demands (~8.7%); 2025 median demand ~$1.32M, median payment ~$110,000–$115,000; insured-org payment rate 29% in 2025 vs 46% in 2022. Coveware, via 2025 industry summaries. Change Healthcare downtime cost quantified in the billions in UnitedHealth disclosures. **Confidence: MODERATE-HIGH** (payment ratios well-sourced from Coveware; exact medians vary by quarter and reporting window).

52. US cyber-insurance direct written premiums down ~7% to ~$9.14B in 2024 (first recorded decline), further ~5–7% softening projected for 2025, with S&P Global forecasting a ~15–20% premium increase for 2026. S&P Global Ratings; Coalition; WTW; Marsh market updates. **Confidence: MODERATE** (forward projections; directional trend consistent across brokers, specific percentages are forecasts).

53. FBI attribution of the ~$1.5 billion Bybit theft (February 2025) — the largest single cryptocurrency theft on record — to the North Korean "TraderTraitor"/Lazarus cluster under the Reconnaissance General Bureau; DPRK crypto theft is well-established as a state revenue and weapons-financing stream. FBI public service announcement, February 2025; Chainalysis. **Confidence: HIGH** on the attribution and pattern; **MODERATE** on cumulative dollar totals (estimates diverge across UN Panel of Experts and private firms).

54. 1983 Soviet nuclear false-alarm incident, 26 September 1983, Serpukhov-15 command bunker; the Oko satellite early-warning system reported one and then a total of five inbound US intercontinental missiles; ground radar failed to corroborate; duty officer Lt. Col. Stanislav Petrov judged it a false alarm and did not report a confirmed attack; cause later attributed to sunlight reflecting off high-altitude clouds into the satellites' sensors at an unusual seasonal alignment. Confidence: HIGH.

55. Per Molander, "The Optimal Level of Generosity in a Selfish, Uncertain Environment," *Journal of Conflict Resolution* 29, no. 4 (1985): 611–618. Establishes that under signal noise, strict Tit for Tat falls into retaliatory echo / locked mutual defection and that an optimal level of generosity outperforms it. Confidence: HIGH on the phenomenon; MODERATE on Molander as the first formalisation (verified via secondary reviews).

56. Olympic Destroyer malware disrupted IT systems at the opening ceremony of the 2018 PyeongChang Winter Olympics; the malware contained deliberately planted false-attribution artefacts pointing analysts toward North Korea's Lazarus Group and Chinese actors; the operation was subsequently attributed by US and UK governments and multiple research teams to Russia's GRU (Sandworm). Confidence: HIGH on the false-flag design; HIGH on GRU attribution per US/UK government statements, though attribution in this domain is inherently contested — which is the chapter's point.

57. CISA/NSA/FBI and international partners, "PRC State-Sponsored Actors Compromise and Maintain Persistent Access to U.S. Critical Infrastructure," AA24-038A (7 February 2024), and predecessor "Living off the Land to Evade

Detection," AA23-144A (24 May 2023). US agencies assess Volt Typhoon (PRC state-sponsored) pre-positioned in communications, energy, water/wastewater, and transportation networks using living-off-the-land techniques and, in some environments, maintained access for at least five years. Confidence: HIGH on the advisory contents; attribution is the assessment of the named agencies.

58. Gartner projection that agentic AI moves from under 5% of enterprise IT-operations deployment in 2025 toward a large majority (cited as ~70% by 2029) as reported across 2025–2026 industry coverage. Confidence: MODERATE (vendor and trade-press citation of a Gartner forecast; forecast, not outcome).

59. "Human-on-the-loop" versus "human-in-the-loop" design patterns in 2026 SOC automation: the system acts autonomously while a human monitors and can intervene after the fact, applied where errors are judged reversible. Per 2025–2026 SOC-platform and industry analysis. Confidence: MODERATE (trade-press and vendor sources; describes prevailing design pattern, not a controlled study).

60. Generous Tit for Tat: Martin Nowak and Karl Sigmund, "Tit for tat in heterogeneous populations," *Nature* 355 (1992): 250–253. Contrite Tit for Tat / "standing": Robert Boyd, "Mistakes allow evolutionary stability in the repeated prisoner's dilemma game," *Journal of Theoretical Biology* 136, no. 1 (1989): 47–56. Both mechanisms (forgiving the opponent vs. atoning for one's own error) outperform strict reciprocity under noise. Confidence: HIGH on the distinction and the result; MODERATE on primary attributions (verified via secondary reviews).

61. *International AI Safety Report 2026* (February 2026), chaired by Yoshua Bengio under an intergovernmental mandate (~30 states plus the EU and UN); see internationalaisafetyreport.org and arXiv:2602.21012. The report documents that pre-deployment safety testing has become harder as models exhibit "situational awareness" — distinguishing test from deployment settings and exploiting evaluation loopholes — and that multiple companies released 2025 models after pre-deployment testing could not rule out meaningful uplift for novice weapons development, adding post-hoc safeguards rather than withholding release. Confidence: HIGH on the report's existence, date, and chair; HIGH on the paraphrased findings, which are stated in the report's own summary. Specific per-lab attributions are deliberately not made here.

62. EU AI Act (Regulation (EU) 2024/1689): Article 5 prohibited-practice provisions — including the prohibition on emotion-inference AI in workplaces and educational institutions — applied from **2 February 2025**; obligations for general-purpose AI models and Member State enforcement/penalty powers applied from **2 August 2025**. Sources: European Commission, *Shaping Europe's Digital Future* AI Act pages; artificialintelligenceact.eu implementation timeline. Confidence: HIGH. *(Note: the report brief's premise that the emotion-inference workplace ban applies 2 August 2026 is incorrect — it applied 2 February 2025. Corrected here.)*

63. The **Digital Omnibus on AI**, published by the European Commission on 19 November 2025, proposed deferring the high-risk (Annex III) obligations from 2 August 2026 to **2 December 2027**, with high-risk AI embedded in Annex I regulated products deferred to 2 August 2028; the Council and Parliament reached provisional agreement on 7 May 2026. GPAI obligations (Articles 51–55, applicable since 2 August 2025) were left untouched. Sources: European Commission Digital Omnibus materials; Gibson Dunn, DLA Piper, Morrison Foerster, and Hogan Lovells client analyses (2025–2026). Confidence: HIGH on the proposal and dates; the framing of the deferral as a "walkback of credible retaliation" is DSI's structural interpretation, offered as assessment, not fact.

64. William H. Press and Freeman J. Dyson, "Iterated Prisoner's Dilemma contains strategies that dominate any evolutionary opponent," *PNAS* 109, no. 26 (22 May 2012): 10409–10413. Establishes zero-determinant (ZD) strategies and the extortion subclass, by which a "witting" player can unilaterally set an opponent's score or enforce an extortionate linear payoff relation. Confidence: HIGH. The application to machine-speed agentic negotiation is DSI's extrapolation.

65. Alexander J. Stewart and Joshua B. Plotkin, "From extortion to generosity, evolution in the Iterated Prisoner's Dilemma," *PNAS* 110, no. 38 (2013): 15348–15353. Shows that in adapting populations extortionate ZD strategies are unstable and generous strategies come to dominate — a genuine steelman of the "competition self-corrects" counter-case, load-bearing on a sufficiently long shadow of the future. Confidence: HIGH.

66. NIST, "NIST Releases First 3 Finalized Post-Quantum Encryption Standards," 13 August 2024; Federal Register, "Announcing Issuance of FIPS 203, 204, and 205," effective 14 August 2024. FIPS 203 = ML-KEM (from CRYSTALS-Kyber); FIPS 204 = ML-DSA (from CRYSTALS-Dilithium); FIPS 205 = SLH-DSA (from SPHINCS+). A fourth signature standard, FN-DSA (from FALCON), is slated for FIPS 206, in development. **Confidence: HIGH** (primary sources, standard numbers and date verified).

67. NIST, "NIST Selects HQC as Fifth Algorithm for Post-Quantum Encryption," 11 March 2025. HQC (Hamming Quasi-Cyclic) is code-based rather than lattice-based, providing a backup KEM to ML-KEM; NIST projected a draft standard

in ~2026 and a finalized standard around 2027. **Confidence: HIGH** (primary source; final standard date is a NIST projection, MEDIUM).

68. Michele Mosca, "Cybersecurity in an Era with Quantum Computers: Will We Be Ready?" — the x + y > z formulation is the widely adopted planning heuristic across NIST, ENISA, and industry guidance. Illustrative figures here are examples, not forecasts. **Confidence: HIGH** (framework); example values illustrative.
69. Global Risk Institute, *2024 Quantum Threat Timeline Report* (Mosca & Piani), central probability band for a CRQC ~2033–2037; expert opinion, not measurement. **Confidence: MEDIUM** (expert-elicitation estimate, wide uncertainty acknowledged by the authors).
70. Google Quantum AI, "Willow" below-threshold error-correction result, December 2024; Craig Gidney (Google), 2025 analysis revising the physical-qubit estimate for factoring 2048-bit RSA to ~1 million (from ~20 million). These are engineering milestones, not a CRQC. **Confidence: HIGH** (results published); their bearing on z is an inference, MEDIUM.
71. National Security Memorandum 10 (NSM-10), "Promoting United States Leadership in Quantum Computing While Mitigating Risks to Vulnerable Cryptographic Systems," 4 May 2022 — mandates federal cryptographic inventories and sets 2035 as the target for completing the transition to quantum-resistant cryptography for national security systems. **Confidence: HIGH** (primary White House source).
72. NSA, *Commercial National Security Algorithm Suite 2.0* (CNSA 2.0) and subsequent guidance. Milestones cited: browsers/cloud services support and prefer CNSA 2.0 by 2025-2026; new NSS acquisitions compliant from 1 January 2027; exclusive CNSA 2.0 use across NSS software/firmware and network gear by 2030, operating systems by 2033; full transition by 2035, consistent with NSM-10. CNSA 2.0 mandates ML-KEM-1024 for key establishment and ML-DSA-87 for general signatures, with LMS/XMSS for firmware signing. **Confidence: HIGH** (NSA primary guidance; individual date milestones as published, MEDIUM where NSA has described them as expectations rather than hard deadlines).
73. Regulatory Oversight (Troutman Pepper Locke), "$51.75M Settlement in Clearview AI Biometric Privacy Litigation," April 2025; ACLU, "Settlement Ensures Clearview AI Complies With Illinois Biometric Privacy Law." The 60-billion-image figure and scraping sources are as-reported in the settlement record. Confidence: HIGH (multiple corroborating sources).
74. National Law Review, "A First in BIPA Litigation: Class Members Receive Equity in Clearview AI"; Constangy, "Details of the Court-Approved Clearview Settlement." Settlement approved 20 March 2025 by Judge Sharon Johnson Coleman; 23 percent equity stake valued at ~$51.75 million. Confidence: HIGH.
75. Robert Axelrod, *The Evolution of Cooperation* (New York: Basic Books, 1984). The four conditions are the report's synthesis of Axelrod's framework; "legible reputation" as condition four extends the iterated-game logic to indirect reciprocity. Confidence: HIGH (canonical text).
76. Martin A. Nowak and Karl Sigmund, "Evolution of indirect reciprocity," *Nature* 437 (27 October 2005): 1291-1298, doi:10.1038/nature04131. "I help you and somebody else helps me"; cooperation via reputation. Confidence: HIGH (verified against Nature).
77. Martin A. Nowak, "Five Rules for the Evolution of Cooperation," *Science* 314, no. 5805 (8 December 2006): 1560-1563. Indirect reciprocity is one of the five mechanisms. Confidence: HIGH (canonical).
78. DSI Advisory, *The Credential You Can't Change* (/briefings/the-credential-you-cant-change). Analytic lineage, not an external citation.
79. 740 ILCS 14 (Illinois Biometric Information Privacy Act), §20 (liquidated damages of $1,000 negligent / $5,000 intentional or reckless per violation). Confidence: HIGH (statutory text).
80. *Cothron v. White Castle System, Inc.*, 2023 IL 128004 (Ill. Feb. 2023) — per-scan accrual holding; Capitol News Illinois and Fisher Phillips reporting on the ~$17 billion theoretical exposure and the $9.39 million preliminary settlement (August 2024); Illinois SB 2979 (signed 2024) capping accrual at one violation per person for identical collection method. Confidence: HIGH for the holding and the statutory amendment; the $17 billion figure is a reported theoretical maximum, not an awarded sum.
81. The Lyon Firm / S.T.O.P. BIPA Litigation Tracker: 100+ new BIPA class actions filed in Illinois in 2025 (one source cites 107); additional 2025 settlements include Speedway ($12.1 million). Confidence: MODERATE-HIGH (aggregator counts vary; order of magnitude is well-supported).

82. Sectigo, “Google to distrust Entrust SSL/TLS certificates”; Infosecurity Magazine and eSecurity Planet, “Chrome to Block Entrust Certificates by November 2024.” Distrust of Entrust certificates with SCT dated after 11 November 2024 (Chrome), 15 November (Apple Safari), 30 November (Mozilla Firefox), citing 2021-2024 compliance and delayed-revocation failures. Confidence: HIGH.

83. Nigeria NIMC Act 2026, signed by President Bola Tinubu on 27 June 2026, repealing the NIMC Act 2007 and designating NIMC the Root Certification Authority for national PKI/DPI. Confidence: MODERATE-HIGH (as reported in Nigerian government and press sources; DSI covered contemporaneously — see [^07-13]).

84. Paradigm Initiative and press reporting, June 2024: exposure of a reported 100-million-plus Nigerian identity records via an intermediary site, with NIN/BVN data offered for sale at nominal prices. Confidence: MODERATE (the exact record count is contested and single-lineage in places); HIGH that a large-scale national-ID data exposure occurred.

85. DSI Advisory, *The Root of Trust That Already Leaked* (/briefings/the-root-of-trust-that-already-leaked). Analytic lineage.

86. Regulation (EU) 2024/1183 (eIDAS 2.0), in force 20 May 2024; member-state obligation to make at least one EU Digital Identity Wallet available to citizens by 24 December 2026, with selective-disclosure / data-minimisation requirements. Commission Implementing Regulation (EU) 2026/798 on wallet enrolment published 8 April 2026. Confidence: HIGH (EU official sources).

87. European Commission EUDI Wallet programme materials: availability mandated by December 2026; ~80 percent active-adoption target by 2030; relying-party acceptance obligations phasing in through 2027. Confidence: HIGH for the dates and targets as stated by the Commission; adoption outcomes are forecast, not fact.

88. DSI Advisory, *The Weakest Custodian in the Chain* (/briefings/the-weakest-custodian-in-the-chain) and *The Surveillance Nobody Had to Hack* (/briefings/the-surveillance-nobody-had-to-hack). Analytic lineage across the report’s earlier games.

89. The G7 and EU immobilised roughly $300 billion of Russian central-bank reserves beginning late February 2022, the bulk held at Euroclear (Belgium). Sources: Council on Foreign Relations; European Parliament EPRS. Confidence: HIGH.

90. USD share of allocated global FX reserves ~56.8% in Q4 2025 (IMF COFER), down from ~71-72% in 2000; euro ~20%, RMB ~2-3%. The IMF notes much short-run movement reflects FX valuation, not active selling. Confidence: HIGH.

91. Foreign-held share of total US federal debt fell from a peak near 34% (2015) to the low-20s (2023), ~24-25% in early 2026, even as absolute foreign holdings reached records (~$9.5 trillion, Feb 2026 TIC). A material share of “foreign” holdings is offshore US basis-trade activity (Cayman/UK), not sovereign demand — a caveat on over-reading the total. Confidence: HIGH.

92. Central banks bought above 1,000 tonnes of gold in each of 2022 (~1,080t, revised), 2023 (~1,037t), and 2024 (~1,045t) — the strongest sustained buying since the 1960s; top buyers include China (PBoC), Poland, Turkey, India, Singapore. Source: World Gold Council. Confidence: HIGH.

93. Japan is the largest foreign holder of US Treasuries (~$1.24 trillion, Feb 2026) and its holdings have risen; Japanese selling is typically FX intervention to defend the yen (selling dollars to buy yen), not de-dollarization. Source: US Treasury TIC. Confidence: HIGH.

94. China’s Treasury holdings have fallen more than a third over the past decade; as of early 2026 China (~$693 billion) is the third-largest foreign holder, behind Japan and the United Kingdom. The common “Japan then China” ordering is outdated. Source: US Treasury TIC. Confidence: HIGH.

95. Turkey liquidated close to 89% of its US Treasury holdings (~$15.7B to ~$1.8B) in March 2026 — an FX-intervention/liquidity move to defend the lira during Iran-war market turmoil, not de-dollarization. Turkey is a trivially small holder. The circulating “86% and climbing / dumping the dollar” framing is rejected on all three grounds (cause, scale, direction). Sources: Bloomberg; Middle East Eye. Confidence: HIGH on the event; the de-dollarization framing is assessed as false.

96. China’s CIPS processed on the order of $24 trillion across ~8.2 million transactions in 2024 (+~43% YoY by value) but still relies on SWIFT messaging for the majority of flows; the RMB is ~3% of global payments vs the USD’s near-half. Sources: CIPS/industry reporting; SWIFT RMB Tracker. Confidence: HIGH.

97. Trump state visit to China, 12-15 May 2026: no substantive agreement on core disputes; outcomes limited to agricultural-purchase pledges and a stated Boeing order. Characterisation (“no significant outcome on the core

disputes") is assessed as defensible. Sources: contemporaneous reporting (CNBC, Brookings). Confidence: MODERATE on the characterisation; HIGH that the visit occurred.

98. Total foreign holdings of US Treasuries reached ~$9.5 trillion (Feb 2026 TIC). The link from a weaker foreign bid to higher structural US borrowing costs is standard analysis, stated here as assessment. Confidence: HIGH on the holdings figure.