BLOCK I · FOUNDATIONS OF KNOWLEDGE & REASONING

Systems Thinking & Feedback

A disc on your wall holds a goal against the cold. The same trick, scaled up, can also tip a planet.

The thermostat: a cheap machine that holds a goal against a world trying to pull it away.

There is a small plastic disc screwed to a wall in your home that is quietly performing a feat which defeated the finest engineers of the eighteenth century. It is *holding a goal.* You told it 21°C; the winter outside is doing everything in its power to make the room colder; and the disc simply will not let it. It takes a reading, compares that reading to the number you set, and – if the room has drifted too cool – closes a circuit that summons heat, then watches to see whether its meddling worked, and adjusts again. Sense, compare, correct, repeat, forever. That little circular conversation between a system and its own goal is one of the most powerful ideas humans have ever stumbled into, and today we are going to follow it from a thermostat all the way to the edge of a destabilizing planet.

The idea has a name – *feedback* – and a science built around it: **systems thinking**. Its central, slightly subversive claim is that to understand almost anything that matters – a body, a market, a forest, a climate – you must stop staring at the *parts* and start watching the *loops*: the wires along which a system's outputs curl back to become its own inputs. Get those loops right and baffling behaviour suddenly makes sense. Get them wrong and the world will surprise you again and again, always in the same direction: badly.

● CORE MODEL

WHERE WE ARE

Yesterday (**Day 8**) we watched a murmuration of starlings turn simple local rules into breathtaking global order and gave it a name: *emergence*. Today we name the engine humming under that hood – **feedback** – and meet the grammar (stocks, flows, loops, delays) that makes wholes behave unlike their parts. Two more callbacks will recur: **Day 5**, where Pearl's causal arrows were forbidden from looping back – a rule feedback breaks on purpose; and **Day 1**, where Friston's predictive brain turned out to be, at heart, a thermostat for *surprise*. Of our five threads – *information, energy, evolution, emergence, computation* – today leans hardest on *emergence* and *energy*, and quietly sets up the climate science of **Day 177**.

The loop that steers

Begin with the machine that started it all. In 1788, James Watt faced an embarrassing problem. His steam engines were magnificent but temperamental: load them lightly and they'd race; load them heavily and they'd bog down. An engine that won't hold a steady speed is nearly useless for spinning looms or grinding flour. Watt's fix, borrowed from windmill-builders, was a small, beautiful piece of hardware called a *centrifugal governor* – and it is pure feedback made of brass.

Faster spin -> balls fly outward -> sleeve rises -> valve closes -> engine slows.

Watch the logic. Two metal balls hang from hinged arms on a spinning spindle geared to the engine. Spin faster and the balls fly *outward* (centrifugal effect); their arms rise; the rising arms pull a linkage that **closes the steam valve**; less steam means the engine **slows down**. Slow too much and the balls droop, the valve opens, the engine speeds back up. The engine's own speed reaches around, grabs the throttle, and corrects itself. No operator, no thought – just a loop wired so that any deviation from the running speed triggers its own undoing. This is *negative feedback*:

change provokes an *opposing* response, and the system settles toward a stable target. Your thermostat, your body's temperature, the pupils of your eyes, the pricing of a competitive market – all the same trick.

For decades the governor was a triumph of tinkering that nobody could quite explain – until **James Clerk Maxwell** (yes, the electromagnetism one – we'll meet his other masterpiece on Day 36) sat down in 1868 and did the mathematics in a paper drily titled "*On Governors.*" Maxwell's real contribution wasn't to praise feedback but to ask the dangerous question: *when does a self-correcting loop fail?* He linearized the engine's equations of motion and showed that a governor can do something worse than fail to correct – it can **overcorrect, then over-overcorrect**, swinging into ever-wider oscillations until the machine tears itself apart. Stability, he proved, is not guaranteed by good intentions; it depends delicately on the numbers. That paper is the seed of all modern *control theory*, and it sat almost forgotten for eighty years.

Its rediscoverer was the mathematician **Norbert Wiener**, who in 1948 published a book with a coined title: *Cybernetics.* The word comes from the Greek *kybernetes*, "steersman"– and so, by a lovely etymological coincidence, does the word *governor* itself (and, eventually, *government*). Wiener's leap was to see that the *same* loop – sense, compare, correct – was the secret common to a thermostat, a steam governor, a hand reaching for a cup, an animal regulating its blood sugar, and a brain. He even singled out Maxwell's 1868 paper as the first serious study of feedback. The pattern didn't care whether it was made of brass, flesh, or arithmetic. **Feedback was substrate-independent.** That insight – control as an abstract structure you can lift out of one system and drop into another – is exactly the kind of cross-domain pattern this whole course is built to chase.

Two loops, two destinies

Feedback comes in two flavours, and almost everything dramatic in a complex system is one of them wearing a costume.

Negative (balancing) feedback resists change and seeks a goal. It is the great *stabilizer*: thermostats, governors, homeostasis, a tightrope walker's endless micro-corrections. Systems people label these loops **B**. They are why you have a body temperature rather than a body temperature *problem.*

Positive (reinforcing) feedback amplifies change and feeds on itself. It is the great *accelerator*: a microphone held too near its speaker (a whisper becomes a shriek in seconds), compound interest, a rumour, a population breeding, a fire heating its own fuel. These loops are labelled **R**. Positive feedback is not "good" and negative "bad"– the names are about *direction*, not virtue. Reinforcing loops build every exponential success story and every runaway catastrophe alike. As we'll see at the end of today, the difference between a planet that wobbles and recovers and a planet that flips into a new state is, very often, the moment a sleeping **R** loop wakes up and overpowers the **B** loops that were holding things steady.

Feedback Engine, as three cases

SETTING	SYSTEM BEHAVIOR	POINT
Low gain, short delay, negative feedback	Settles toward 21°C	Correction is gentle enough and arrives early enough.
High gain, long delay, negative feedback	Oscillates around the target	Correction arrives after the state has changed, so the loop chases itself.
Positive feedback	Deviation amplifies	A sign flip turns goal-seeking into runaway amplification.

Stocks and flows: the physics of accumulation

If feedback is the verb of systems thinking, *stocks and flows* are its nouns. The idea is almost insultingly simple – which is exactly why people get it catastrophically wrong.

A *stock* is anything that accumulates: water in a bathtub, money in an account, carbon in the atmosphere, people in a population, grievance in a marriage. A *flow* is a rate that changes a stock: the faucet filling the tub, the drain emptying it. And here is the entire, deceptively deep rule:

> *A stock rises whenever its inflow exceeds its outflow — no matter what the flows themselves happen to be doing.*

Read that twice, because human intuition violently disagrees with it. Why? Because we are very good at tracking visible motion and very bad at mentally integrating a rate over time. When the faucet curve falls, the eye wants the level to fall with it. But a stock is not a copy of either flow; it is the running total of the **difference** between inflow and outflow. So a falling inflow can still raise the stock if it remains above the outflow. That small distinction is the whole trap.

The man who formalized stocks and flows into a working discipline was **Jay Forrester**, an MIT engineer who had built flight simulators and pioneering computer memory before turning, in the 1950s, to a stranger machine: the corporation. Hired to explain why a General Electric appliance plant kept lurching through three-year cycles of frantic hiring and brutal layoffs – with *steady* consumer demand – Forrester showed the booms and busts weren't caused by the market at all. They were manufactured *inside* the company, by the delays and feedback loops in how managers reacted to inventory and orders. The cycles were a property of the system's *structure*, not its environment. He called the new field *system dynamics*, and its core insight was unsettling: **complex systems are counterintuitive.** They actively mislead the people inside them.

Forrester’s most influential student was the extraordinary **Donella Meadows**, lead author of the 1972 global bestseller *The Limits to Growth* and, decades later, of the field’s beloved primer, *Thinking in Systems.* Meadows had a gift for turning equations into plain English, and she returned again and again to the humblest possible teaching device: a bathtub.

The 36% problem

This is real, not a course anecdote. The source is Linda Booth Sweeney and John Sterman’s 2000 *System Dynamics Review* paper, “Bathtub dynamics: initial results of a systems thinking inventory.” They gave highly educated graduate students a simple graph: here is the inflow to a tub over time, here is the outflow; sketch the water level. It requires no calculus, only the rule above. **Only about 36% drew the right shape.** A 2009 follow-up by Cronin, Gonzalez, and Sterman named the dominant mistake the **correlation heuristic**: people drew the *level* rising and falling in step with the *inflow*, as if the stock should mimic the flow. It doesn’t.

Booth Sweeney and Sterman’s bathtub task: many people draw the stock as if it should rise and fall with the inflow.

The mistake seems trivial only after you see the right graph. Before that, the wrong answer feels visually natural: “the faucet is going down, so the level should go down.” The hidden operation is accumulation. As long as the area where inflow exceeds outflow is still growing, the water level keeps rising.

Why should you care about a tub? Because the atmosphere is one. Carbon dioxide is a stock; our emissions are the inflow; natural sinks (oceans, plants) are the drain. Right now humans pour CO2 into the air at roughly twice the rate the drain can remove it. Sterman has shown that the very same 36%-style error infects public reasoning about climate: many people assume that if we simply *stop increasing* emissions, concentrations will level off. They won't. As long as the faucet beats the drain, the level keeps climbing. To *stabilize* the stock, you must cut the inflow all the way down to meet the outflow – a far more demanding ask, and one that the bathtub teaches in about ten seconds once you actually watch it.

The bathtub, as stock and flow

PHASE	FLOW RELATION	WATER LEVEL
Faucet rising	Inflow exceeds outflow	Rises quickly
Faucet falling but still above drain	Inflow still exceeds outflow	Keeps rising
Faucet below drain	Outflow exceeds inflow	Only then falls

What loops and delays do to us

Wire stocks, flows, and feedback together and you can draw almost any system as a *causal loop diagram* – arrows showing what affects what, with each closed loop labelled **R** (reinforcing) or **B** (balancing). The notation is humble; the payoff is that recurring *shapes* jump out – what Meadows and others call **system archetypes**.

Every real population is a tug-of-war between growth engines and limiting brakes.

More people -> more births -> *more people* is an **R** loop: the engine of growth. But more people -> more crowding and scarcity -> *fewer* net births is a **B** loop: the brake. Every real population is a tug-of-war between the two, and most real systems mix both: a reinforcing loop creates motion, a balancing loop sets limits, and delay decides whether the result is smooth or chaotic.

Two complications turn these tidy diagrams into the stuff of real-world disaster. The first is **delay**. A balancing loop with a long lag doesn't gently settle – it overshoots, like a shower with a slow water heater that scalds you the instant after you've cranked the cold. (You felt this directly in the Feedback Engine: add delay and the calm thermostat starts to hunt.) The second is **nonlinearity**: in systems with feedback, doubling the cause rarely doubles the effect. Push gently and nothing happens; push a little harder and the whole thing lurches.

The canonical demonstration is a game. MIT's *Beer Distribution Game* sits players along a supply chain – retailer, wholesaler, distributor, factory – each ordering from the next, each separated by shipping delays, none able to see the whole. The customer demand is almost flat: a single, modest, one-time bump. And yet, reliably, every time it's played, that tiny ripple at the retail end amplifies into wild, accelerating swings of frantic over-ordering and despairing cancellation upstream – the **bullwhip effect.** John Sterman's classic 1989 study showed the cause isn't stupidity or greed; it's that people, unable to perceive the feedback structure they're embedded in, systematically misjudge the delays and overreact. The lesson is humbling and it generalizes far beyond beer: *well-meaning people, reacting sensibly to what's in front of them, reliably destabilize a system whose loops they cannot see.*

Which is why Meadows spent her career asking a different question: not "how do I push harder?" but "where is the *leverage*?" Her famous, ranked list of *leverage points* – places to intervene in a system – turns intuition upside down. The interventions everyone reaches for (tweak a number, adjust a tax, set a target) sit near the *bottom*, the weakest places to push. The truly powerful leverage lives higher up: the structure of the feedback loops themselves, the rules of the system, the flow of information, and – most powerful of all – the *goals* and *paradigms* the whole system is organized around. Change a parameter and the system shrugs. Change a system's goal and everything downstream reorganizes. It is the single most practical idea in the field, and we'll lean on it for the rest of the course.

Take it apart, or leave it whole?

Lurking under all of this is a genuinely old quarrel about how to understand *anything*. The dominant method of science – call it *reductionism* – says: to understand a thing, break it into parts, understand each part, and reassemble. It has been spectacularly successful. Cells explain bodies; atoms explain chemistry; the strategy built the modern world.

But systems thinkers, following the biologist **Ludwig von Bertalanffy** and his 1968 *General System Theory*, push back: for a system whose parts are richly looped together, taking it apart *destroys the very thing you wanted to study.* Cut the feedback loops and you're left holding inert components and a puddle of behaviour that lived only in their interaction. A living cell is an *open system*, ceaselessly trading matter and energy with its surroundings to hold itself far from equilibrium – study its molecules in isolation and the aliveness evaporates. This is the *holist's* complaint: *the whole is not the sum of its parts; it is the sum plus the pattern of their connections.*

Here we can be more precise than the old slogans, because **Day 8** already gave us the tools. The honest resolution isn't holism *or* reductionism – it's knowing when each applies. Reductionism works beautifully when a system's parts are *weakly coupled* and combine in roughly *linear* ways; you can study them one at a time and add up the answers. It breaks down precisely when the parts are strongly looped and nonlinear – when feedback makes the whole behave in ways no isolated part can show. That's not mysticism; it's *weak emergence* (Day 8): system-level behaviour that is fully grounded in the parts yet only knowable by letting them interact – by running the loops. Whether there's also *strong* emergence – genuinely irreducible powers, like a downward-causing whole – remains, as we saw, contested and (for now) confined to the riddle of consciousness. Feedback is what makes "more is different" a tractable engineering fact rather than a slogan. And note the quiet rhyme with **Day 5**: Pearl's causal graphs had to be *acyclic* – no arrows allowed to loop back – which is exactly the structure feedback violates. A thermostat *is* a causal cycle. Systems thinking is, in part, what causal reasoning looks like once you let the arrows come full circle.

The edge: when loops run away

Now we cash it all in on the highest-stakes system of all. Most of the time, the Earth's great systems are governed by balancing loops – push them and they push back toward where they were. A forest cools and humidifies the air above it; an ice sheet's bright surface reflects sunlight and keeps itself frozen. Stability, bought by negative feedback. But hidden inside many of these systems is a sleeping *reinforcing* loop – and if a driver (say, warming) pushes hard enough, the R loop can wake, overpower the B loops, and carry the system, all by itself, into a completely different state. That threshold is a *tipping point*, and crossing it can be abrupt, self-sustaining, and very hard to undo.

The mathematics is the theory of *critical transitions*, charted most influentially by the ecologist **Marten Scheffer**. Picture the system's state as a ball resting in a valley – a stable basin held by feedback. As a driver changes, the landscape itself deforms: your valley grows shallower while a second valley deepens nearby. For a while the ball stays put. But at a critical point – a *fold bifurcation* – your valley flattens to nothing, and the ball rolls irreversibly into the other basin. The sting is *hysteresis*: to get back, it is not enough to undo the original nudge. You have to push the driver *far* back the other way until the *second* valley disappears. The door only swings one way cheaply. That asymmetry is why "we'll just reverse it later" is, for many tipping elements, a dangerous fantasy.

Most beautifully, the theory predicts a **warning sign.** As a valley flattens toward its tipping point, the ball recovers ever more sluggishly from each random jostle –

a phenomenon called *critical slowing down*. Statistically, this shows up as *rising variance* (the ball wanders further) and *rising autocorrelation* (each moment looks more like the last, because the system can't spring back). These *early-warning signals*, introduced in a landmark 2009 *Nature* paper by Scheffer and colleagues, are the same regardless of whether the system is a lake, a power grid, or an ice sheet. The panel below lets you feel all three ideas at once.

Edge 01 · warning signal

● CRITICAL SLOWING DOWN · ESTABLISHED

● AS A DATE-PREDICTOR · CONTESTED

The warning sign — and its limits

Critical slowing down is real and demonstrated, both in theory and in controlled whole-lake experiments. But there's a hard catch the headlines tend to skip. Detecting that a system is *losing resilience* is one thing; extrapolating to the *year* it will tip is quite another. A sober 2024 analysis (Ben-Yami and colleagues, *Science Advances*) argued that the uncertainties are simply "too large to predict tipping times of major Earth system components from historical data"– the records are short and patchy, and the answer depends heavily on modelling assumptions. So: the signal is a genuine resilience thermometer, but not a crystal ball for specific dates. Hold that distinction – it is about to matter enormously.

Stability landscape, as warning signs

POSITION	WARNING SIGNAL	RESULT
Deep basin	High resilience, low variance	Perturbations return to the old state
Near the edge	Low resilience, high variance and autocorrelation	Small shocks can trigger a transition
Past the fold	The old basin is gone	Undoing the driver is not enough for immediate return

Edge 02 · climate thresholds

● TIPPING ELEMENTS EXIST · ESTABLISHED

● EXACT THRESHOLDS · LARGE UNCERTAINTY

Which parts of the Earth can tip — and when

The most careful systematic estimate is a 2022 *Science* paper led by David Armstrong McKay and Tim Lenton, which identified roughly **nine "core" global tipping elements** plus several regional ones, each with an estimated warming threshold. Its sobering headline: with the planet already around 1.2–1.3°C above pre-industrial, we have entered the lower uncertainty range of *five* of them – and several more become “likely” within the 1.5 – 2°C window the Paris Agreement was built to stay inside. The chart below shows the verified central estimates and their (very wide) uncertainty ranges. Two lessons are baked into its shape: the ranges are *enormous* (this is genuine scientific humility, not vagueness), and several bars already cross today’s warming line.

Central estimates and uncertainty ranges; several lower bounds already overlap today’s warming.

Two assessments now anchor this field for policymakers: the first **Global Tipping Points Report** (Lenton et al., launched at COP28 in December 2023), the largest synthesis of its kind, and its **2025 second edition** (374 pages, October 2025, >160 scientists). The 2025 report’s grim centerpiece: warm-water coral reefs are now judged to be *passing* their tipping point – their threshold (central estimate near 1.2°C) sits below today’s warming, and even holding at 1.5°C, widespread loss is assessed as virtually certain. This isn’t abstract: NOAA’s Coral Reef Watch reports that the bleaching event running from 2023 into 2025 has hit roughly **84% of the world’s reef area**, the largest on record. The reports also stress a hopeful mirror image – *positive* tipping points, where reinforcing loops in technology (solar, batteries, EVs) drive self-accelerating *good* change. The same runaway math can cut both ways.

Edge 03 · AMOC caution

● AMOC WEAKENING & AT RISK · ESTABLISHED

● “COLLAPSE BY ~2057”· CONTESTED/HYPE

The Atlantic’s conveyor belt — a case study in careful reading

No tipping element better shows why the hype filter matters than the **Atlantic Meridional Overturning Circulation** (AMOC) – the vast ocean conveyor that carries warm water north and gives northwest Europe its improbably mild climate. Disentangle

three layers:
What's established. The AMOC is a genuine tipping element; it has very likely weakened; and it is "very likely" to weaken further this century. This is mainstream **IPCC** assessment. Notably, the IPCC's 2021 report *lowered* its confidence that an abrupt collapse this century can be ruled out – to only "medium confidence" that it will *not* collapse before 2100 – precisely because climate models are known to be biased toward over-stability. The risk is real and was upgraded, not dismissed.
What's contested. In July 2023, Peter and Susanne Ditlevsen published a much-publicized *Nature Communications* paper fitting a tipping model to Atlantic sea-surface temperatures and estimating collapse around mid-century – a central year of **2057**, with a 95% range of 2025 – 2095. (A 2025 correction to that paper, fixing code errors, shifted the central estimate to roughly **2065**, range 2037 – 2109.) These specific dates drew intense criticism – they lean on a short temperature proxy and strong statistical assumptions, exactly the "too uncertain to date" problem from Edge 01. **The warning deserves to be taken seriously; the precise year does not deserve to be quoted as fact.** Treat the dating claim as hype risk, not settled timing.
What's new and solid. In February 2024, van Westen, Kliphuis and Dijkstra (*Science Advances*) achieved something genuinely important: the first simulation of a full AMOC collapse in a complex global climate model, yielding a physics-based early-warning signal – and finding that the present-day AMOC appears to be "on route to tipping." Crucially and admirably, that paper gives *no calendar date.* It tells us the direction of travel is worrying without pretending to know the arrival time. Later that year, 44 scientists signed an open letter urging governments to take the underestimated risk seriously. The honest 2026 summary: *the AMOC is weakening and the risk of crossing a tipping point this century is real and possibly underestimated – but anyone quoting you a specific collapse year is selling more certainty than the science can buy.*

One year is not the threshold

You'll have seen the headline: 2024 was the first calendar year more than 1.5°C above pre-industrial (1.60°C, per Copernicus, January 2025). It's a real and alarming milestone – and it is *not* a breach of the Paris goal, which refers to a multi-decade *average*, not a single hot year. This is itself a stocks-and-flows lesson: climate goals are defined on the slow-moving stock (long-run temperature), not the noisy annual flow. Keep the timescales straight, and you'll dodge both false alarm and false comfort.

What's genuinely unsettled

- **Can we ever predict a tipping *date*?** Early-warning signals reliably detect lost resilience in clean experiments. Whether they can ever give trustworthy timing for a noisy, half-observed system like the AMOC is, right now, a real and open methodological fight.

- **How coupled are the tipping elements?** The nightmare scenario is a *cascade* – Greenland's meltwater weakening the AMOC, a weakened AMOC drying the Amazon, and so on, one domino tipping the next. The links are plausible and partly modelled, but their strengths are not yet pinned down.
- **Where exactly does reductionism stop?** We can say feedback and nonlinearity mark the boundary – but turning "this system needs holistic treatment" into a sharp, quantitative test (rather than a judgment call) is unfinished business, tangled up with the strong-emergence question still open from Day 8.
- **Can we engineer *positive* social tipping points on purpose?** If reinforcing loops can flip a system into ruin, can they be deliberately seeded to flip it toward good – rapid decarbonization, say? The 2025 report bets yes. Whether that's sound systems engineering or hopeful narrative is being tested in real time.
- **Is the brain literally a control system?** The Day 1 thread returns: predictive processing casts cognition as feedback minimizing surprise. Useful framing, or the literal architecture? (We'll push on this at Days 119 and 123–126.)

The day in three sentences

Big idea: To understand a body, a market, or a climate, watch the *loops*, not just the parts: feedback (balancing loops that stabilize, reinforcing loops that run away), stocks and flows, and the delays between them generate behaviour that no isolated component can show – and that human intuition reliably gets wrong.
Best analogy: The thermostat holding a goal against the cold (negative feedback) and the bathtub whose level keeps rising long after the faucet starts closing (stock lags flow) – scaled all the way up to a ball in a flattening valley that tips, irreversibly, into a new basin.
Live controversy: That Earth systems can tip is established and early-warning signals detect lost resilience; but predicting a specific collapse *date* – the AMOC's contested "~2057/2065"– outruns what the data can support, even as the underlying risk is real and may be underestimated.
Threads today › emergence (feedback as the engine under "more is different") · energy (open systems held from equilibrium by throughflow -> Days 33, 83 – 85) · computation (system dynamics; the brain as controller) · information (stocks/ flows; early-warning signals as evidence) – with a light touch of evolution (homeostasis, selected-for control).

"TOMORROW -> DAY 10"

Models, Maps & Idealization

Today we built models of systems — Forrester's, Meadows' World3, a ball in a valley — and trusted them to teach us. Tomorrow we turn the lens on the modelling itself. "The map is not the territory"; "all models are wrong, but some are useful." When is a simplification illuminating and when is it a lie? We'll meet the realism-vs-instrumentalism debate and the modern epistemology of simulations and digital twins — the question of how much to believe the very tools we just used.

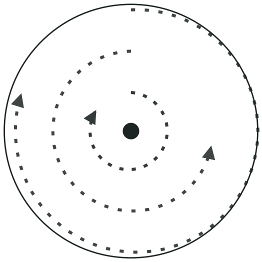
Sources & further reading

1. Maxwell, J. C. (1868). "On Governors." *Proceedings of the Royal Society of London* 16: 270–283. doi:10.1098/rspl.1867.0055 — the mathematical birth of control theory. doi.org/10.1098/rspl.1867.0055
2. Wiener, N. (1948). *Cybernetics: Or Control and Communication in the Animal and the Machine.* MIT Press. — coins "cybernetics"; names Maxwell 1868 the first feedback paper; thermostat & governor as exemplars.
3. Forrester, J. W. (1961). *Industrial Dynamics.* MIT Press. Also *Urban Dynamics* (1969), *World Dynamics* (1971), and "Counterintuitive Behavior of Social Systems"(1971).
4. Meadows, D. H. (2008). *Thinking in Systems: A Primer* (ed. D. Wright). Chelsea Green. — the field's standard primer (posthumous).
5. Meadows, D. H. (1999). *Leverage Points: Places to Intervene in a System.* The Sustainability Institute. (expanded from a 1997 essay in *Whole Earth Review*). donellameadows.org
6. Meadows, D. H., Meadows, D. L., Randers, J. & Behrens, W. W. (1972). *The Limits to Growth.* Universe Books (Club of Rome; World3 model).
7. Booth Sweeney, L. & Sterman, J. D. (2000). "Bathtub dynamics: initial results of a systems thinking inventory." *System Dynamics Review* 16(4): 249–286. — only ~36% of educated subjects read accumulation correctly.
8. Cronin, M. A., Gonzalez, C. & Sterman, J. D. (2009). "Why don't well-educated adults understand accumulation?" *Organizational Behavior and Human Decision Processes* 108(1): 116–130. — the "correlation heuristic"; the CO2-bathtub misconception.
9. Sterman, J. D. (1989). "Modeling Managerial Behavior: Misperceptions of Feedback in a Dynamic Decision-Making Experiment." *Management Science* 35(3): 321–339. doi:10.1287/mnsc.35.3.321 — the Beer Game & bullwhip effect. doi.org/10.1287/mnsc.35.3.321
10. von Bertalanffy, L. (1968). *General System Theory: Foundations, Development, Applications.* George Braziller. — open systems, equifinality, the holist program.
11. Scheffer, M. et al. (2009). "Early-warning signals for critical transitions." *Nature* 461: 53–59 (3 Sep 2009). doi:10.1038/nature08227 — critical slowing down, rising variance & autocorrelation. doi.org/10.1038/nature08227
12. Scheffer, M. (2009). *Critical Transitions in Nature and Society.* Princeton University Press. — bifurcations, alternative stable states, hysteresis.
13. Armstrong McKay, D. I., Staal, A., Abrams, J. F., Winkelmann, R., Sakschewski, B., Loriani, S., Fetzer, I., Cornell, S. E., Rockstrom, J. & Lenton, T. M. (2022). "Exceeding 1.5°C global warming could trigger multiple climate tipping points." *Science* 377(6611): eabn7950 (9 Sep 2022). doi:10.1126/science.abn7950 — the verified threshold table. doi.org/10.1126/science.abn7950
14. Ben-Yami, M., Morr, A., Bathiany, S. & Boers, N. (2024). "Uncertainties too large to predict tipping times of major Earth system components from historical data." *Science Advances* 10(31): eadl4841 (2

Aug 2024). doi:10.1126/sciadv.adl4841 – the methodological brake on date-prediction. doi.org/10.1126/sciadv.adl4841

15. Ditlevsen, P. & Ditlevsen, S. (2023). "Warning of a forthcoming collapse of the Atlantic meridional overturning circulation." *Nature Communications* 14: 4254 (25 Jul 2023). doi:10.1038/s41467-023-39810-w – central est. 2057 (2025–2095); 2025 correction -> ~2065 (2037–2109). Contested dating. doi.org/10.1038/s41467-023-39810-w
16. van Westen, R. M., Kliphuis, M. & Dijkstra, H. A. (2024). "Physics-based early warning signal shows that AMOC is on tipping course." *Science Advances* 10(6): eadk1189 (9 Feb 2024). doi:10.1126/sciadv.adk1189 – first full AMOC tip in a complex model; deliberately no date. doi.org/10.1126/sciadv.adk1189
17. Flores, B. M. et al. (2024). "Critical transitions in the Amazon forest system." *Nature* 626: 555–564 (14 Feb 2024). doi:10.1038/s41586-023-06970-0 – 10–47% of forest exposed to compounding disturbances by 2050. doi.org/10.1038/s41586-023-06970-0
18. Boulton, C. A., Lenton, T. M. & Boers, N. (2022). "Pronounced loss of Amazon rainforest resilience since the early 2000s." *Nature Climate Change* 12: 271–278. doi:10.1038/s41558-022-01287-8 – rising autocorrelation (an early-warning signal) across the Amazon.
19. Lenton, T. M. et al. (eds.) (2023; 2nd ed. 2025). *Global Tipping Points Report.* University of Exeter. – the COP28 synthesis and its Oct 2025 update (warm-water coral reefs passing their tipping point; positive tipping points). global-tipping-points.org
20. Copernicus Climate Change Service (2025). "Global Climate Highlights 2024"(10 Jan 2025). – 2024 the first calendar year >1.5°C, at 1.60°C above 1850–1900; single year = Paris breach. climate.copernicus.eu

Optional appendix: Deeper Waters



Loops inside loops inside loops. The deeper you look, the more there are.

The main descent left us at the edge of a destabilizing planet, holding one hard-won distinction: that detecting a system's *loss of resilience* is not the same as naming the *year* it breaks. That gap – between sensing fragility and forecasting failure – is the doorway to this appendix. Everything below lives in that gap. We've already met the skeleton of systems thinking: the thermostat and the governor, stocks and flows, the bathtub, the ball in its flattening valley. Now we go down past the textbook into the engine room – the strange people who built this science, the deep laws they uncovered, and the wildest idea any of them ever had: that a planet, with no brain and no plan, can hold its own temperature steady for billions of years using nothing but feedback. None of what follows is required to pass the day. All of it is the good stuff.

“RETURN CARRIED DOWN FROM THE MAIN FLOOR”

Assumed throughout (so we don't repeat them): **negative/balancing (B)** loops stabilize, **positive/reinforcing (R)** loops run away; **stocks** accumulate what **flows** deliver; **delay** turns correction into oscillation; a **fold bifurcation** with **hysteresis** is the shape of a tipping point; and **critical slowing down** is its warning sign. We'll lean on every one of these and add new floors beneath them.

● DEEP DIVE

How a war, a ferry, and a dinner party invented cybernetics

Norbert Wiener didn't set out to found a science of control. He set out to shoot down aeroplanes. In 1940, with Britain under bombardment, the American mathematician was handed a brutally practical problem: anti-aircraft gunners couldn't hit fast planes because a shell takes time to climb, and by the time it arrives the plane has moved. To hit the target you must fire not where it *is* but where it *will be* – you must *predict*. Wiener built a statistical predictor that treated the pilot's evasive jinking as a noisy signal to be extrapolated, and in wrestling with it he noticed something that would consume the rest of his life: the gun, the radar, and the dodging pilot formed a single *loop* of sensing, predicting, and correcting, and the same loop described an animal stalking prey, a hand reaching for a glass, a brain. Prediction and feedback were two faces of one thing. (The historian Peter Galison has argued that this wartime fusion of human and machine into one feedback circuit quietly shaped the entire worldview of *cybernetics*.)

The hardware side of the story is even better, and it had happened a decade earlier on a commuter ferry. In August 1927, a young Bell Labs engineer named **Harold Black** was crossing the Hudson, stuck on a problem that was strangling the telephone industry: amplifiers for long-distance calls distorted the signal a little, and across thousands of miles of repeaters those little distortions compounded into mush. Black's flash of insight, which he scribbled on the only paper he had – a copy of *The New York Times* – was almost perverse: deliberately feed a portion of the amplifier's *output* back to its input *with the sign flipped*, so the amplifier continuously corrects its own errors. He was throwing away gain to buy accuracy. The *negative-feedback amplifier* made coast-to-coast telephony possible and, not incidentally, vindicated Maxwell's old warning from the main file: feedback can stabilize, but wire it wrong and it screams. Within a few years Harry Nyquist and Hendrik Bode (both at Bell Labs) had worked out the mathematics of *exactly when* a feedback loop stays stable versus bursts into oscillation – the formal engineering of the line between control and catastrophe.

Then came the dinner parties. Between 1946 and 1953, a rotating cast of the century's most restless minds gathered for a series of meetings funded by the Josiah Macy Jr. Foundation – the *Macy Conferences* – to ask whether feedback, information, and control might be a single language spanning machines, brains, and societies. The guest list is almost comic in its density: Wiener, the polymath John von Neumann, the neurophysiologist Warren McCulloch, the anthropologists Margaret Mead and Gregory Bateson, and the British psychiatrist **W. Ross Ashby**. Out of that cauldron came the conviction that *organization itself* – in a cell, a thermostat, a market, a mind – could be studied as one subject. Later thinkers, led by Heinz von Foerster, added a vertiginous twist they called *second-order cybernetics*: the observer studying a system is part of a feedback loop *with* it, so there is no view from nowhere. A scientist watching a society changes it by watching; a "cybernetics of cybernetics." We'll meet that snake-eating-its-tail again when this course reaches the science of mind.

Ashby's law: only variety can absorb variety

Of all the deep laws to come out of that period, the one most worth carrying is Ashby's, and it sounds almost too simple to be profound. A regulator – a thermostat, an immune system, a government, a goalkeeper – faces a world that can throw a certain number of distinct disturbances at it. Call that number the world's *variety*. Ashby proved that **to hold an outcome steady, the regulator must command at least as much variety as the disturbances it faces.** Only variety can absorb variety. A goalkeeper with one fixed dive cannot stop shots aimed everywhere; a immune system with ten antibodies cannot defeat a thousand pathogens; a policy with a single lever cannot govern a society that can go wrong in a hundred ways. This is the *Law of Requisite Variety*, and it is the iron reason that simple controllers fail against complex worlds – and why every successful regulator is, secretly, at least as complicated as its problem.

A regulator can only neutralize as many kinds of trouble as it has kinds of response.

Keep this law in your pocket. It will return, transformed, as a reason large language models need staggering scale to handle the variety of language, as the logic behind biological diversity buffering ecosystems, and as the quiet mathematics under "you can't manage what you can't match."

Your body is ten thousand thermostats

Long before Wiener, the most sophisticated feedback machine on Earth was sitting inside every reader of this page. In 1865 the French physiologist **Claude Bernard** noticed something that should have caused more fuss than it did: the fluid bathing our cells – the *milieu interieur*, the "internal environment"– stays eerily constant even as the outside world swings wildly. Your blood is the same warm, salty, sugar-balanced broth whether you're in a blizzard or a sauna. Bernard's line became one of the most quoted in biology: the constancy of the internal environment is the condition for a free and independent life. Seventy years later the American physiologist **Walter Cannon** gave the phenomenon its name – *homeostasis*, "steady-state"– and spelled out that it is achieved by exactly the negative-feedback loops we've been tracking, just rendered in flesh and hormones instead of brass.

The examples are everywhere once you see them. Your blood sugar is a stock regulated by two opposing flows: when it rises after a meal, the pancreas releases *insulin* to push glucose into cells; when it falls, *glucagon* releases it from storage – a thermostat with two hormones for a sensor and a switch. Your core temperature is held near 37°C by sweating, shivering, and the rerouting of blood. Your blood pressure is policed second-by-second by the *baroreflex*: stretch receptors in your arteries sense pressure and reflexively adjust your heart rate to counter it. Stand up too fast and feel the half-second lag before the loop catches up – that brief dizziness is delay-induced overshoot, the baroreflex doing its Beer-Game stumble. You are not one thermostat; you are a teeming republic of them, mostly arguing in your favour.

The useful exception: when the body wants a runaway

Biology mostly wants stability, so it mostly uses negative feedback. But every so often it needs an *event* – something fast, decisive, irreversible – and for that it reaches for **positive** feedback, carefully caged. Childbirth is the classic: the baby's head stretches the cervix, which triggers oxytocin, which strengthens contractions, which stretch the cervix further – an explosive R loop that builds to a climax and then switches off once the baby is delivered. Blood clotting cascades the same way; so does the firing of a neuron, where a trickle of charge flings open the gates for a flood. Positive feedback isn't the body's enemy. It's the body's detonator – used sparingly, with a hand always on the off-switch.

And here the appendix loops back to the very first day of this course. The old picture of homeostasis is purely *reactive*: wait for a deviation, then correct it. But physiologists increasingly favour a richer idea called *allostasis* – "stability through change"– in which the body doesn't wait for the error at all. It *predicts* demand and adjusts in advance: your cortisol rises before you wake, your heart rate climbs in anticipation of exertion, not in response to it. Regulation becomes *feedforward*, not just feedback – anticipatory control. If that rings a bell, it should: it is the bodily cousin of the **predictive brain** from Day 1, Friston's machine that minimizes surprise by modelling the world before it arrives. Cybernetics' deepest modern turn is the realization that the best controllers don't merely react to the world – they see it coming.

A socialist control room and a doomsday model

If feedback governs bodies and machines, the temptation was irresistible: why not govern an *economy* with it? In 1971, the British management theorist **Stafford Beer** – a cigar-smoking, yoga-practising guru of "management cybernetics"– was invited by Salvador Allende's newly elected socialist government to do exactly that, for the entire nation of Chile. The result, **Project Cybersyn** (Synco, in Spanish), is one of the strangest and most haunting episodes in the history of technology: an attempt to run a national economy as a single self-regulating organism, in real time, with the computing power of one mainframe and a few hundred telex machines.

Seven chairs, big-button armrests, the economy on the walls: a national feedback dashboard before personal computers.

The system had a daily heartbeat. Factories wired their production figures over a repurposed telex network ("Cybernet") to a central computer in Santiago, which fed back warnings when a number drifted out of its expected range – a national nervous system signalling pain. At its heart sat the *Opsroom*: a hexagonal chamber of seven white fibreglass swivel chairs, each with big buttons on the armrest, facing wall-mounted screens displaying the state of the economy. It looked like the bridge of a starship and was meant to be a place where managers could *see* the country's feedback loops and steer them. The system's finest hour came in October 1972, when a strike by truck owners threatened to choke the country's supply lines; the government used the Cybernet telex network to coordinate the few thousand loyal trucks in something close to real time, and kept the nation fed. A little over a year later, the project died in a single afternoon – the military coup of 11 September 1973 swept away Allende, and Cybersyn with him. It remains the most ambitious attempt ever made to govern a society by explicit feedback, equal parts inspiration and warning. ● A REAL SYSTEM, GENUINELY OPERATIONAL – though whether it could ever have *worked* at scale runs straight into Ashby's Law: can one control room ever hold variety enough to match a whole economy?

The model that refused to die

The other great attempt to model a whole system was quieter but louder in its afterlife. We met **The Limits to Growth** (1972) in the main file as Donella Meadows' work; here is what it actually *did*, and why it still starts arguments fifty years on. A team at MIT built a system-dynamics model called *World3* with five tightly looped global stocks – population, food production, industrial output, pollution, and nonrenewable resources – and let it run to the year 2100 under different assumptions. The headline "standard run," with business roughly as usual, produced a disturbing shape: growth continuing for decades, then *overshoot* and a sharp 21st-century decline in population and output as resource depletion and pollution caught up. The reaction was ferocious. Economists savaged it; the model was crude, its resource estimates pessimistic, its omission of price-driven adaptation and human ingenuity glaring. For a generation, "Limits to Growth" was a byword for doom-mongering that history had supposedly refuted.

And then the data started coming in. In 2008 and again in 2014, the Australian physicist **Graham Turner** (CSIRO) compared three to four decades of real-world statistics against the 1972 scenarios and found, uncomfortably, that the world had been tracking the "standard run" with eerie fidelity. In 2021, the analyst **Gaya Herrington** updated the comparison with fresh data and reached a similar verdict: empirical trends sat closest to the model's business-as-usual and "comprehensive technology" runs, both of which imply growth stalling around the 2040s. The honest, hype-filtered reading matters here, because it cuts both ways. This is *not* a validated prophecy – World3 is a coarse model, the agreement is partly coincidental, and "the standard run fits so far" says little about the exact timing or inevitability of any decline. ● "PROVEN RIGHT"– CONTESTED/HYPE. But the lazy dismissal – that it was simply *wrong* – is also false. ● THE DATA-TRACKING STUDIES ARE REAL. The truth is the more unsettling middle: a 50-year-old systems model of overshoot has not yet been embarrassed by reality. Which is the perfect place to hand you to tomorrow's question – *how much should we ever believe a model?* – but we're getting ahead of ourselves.

The twelve places to push

The single most practical thing Donella Meadows ever wrote was a ranked list. We sketched it in the main file – *parameters are weak, paradigms are strong* – but the full ladder rewards a proper look, because it inverts almost everything our instincts tell us about how to change a system. We reach reflexively for the bottom rungs (adjust a number, set a target) precisely because they're easy to grasp and easy to push. The trouble is that they're also the *weakest*. The real power sits higher up, where the leverage is enormous and the handhold is hard. Climb the ladder; tap any rung to open it.

Meadows' leverage points, grouped

LEVEL	EXAMPLES	WHY IT MATTERS
12-10	Parameters, buffers, stock-flow structure	Most visible, often only tuning.
9-6	Delays, balancing loops, reinforcing loops, information flows	Changes how feedback reaches actors.
5-1	Rules, self-organization, goals, paradigms, transcending paradigms	Changes what the system optimizes for and can imagine.

Notice the shape of the lesson. Tax tweaks and targets — the entire substance of most political argument — live at the bottom. The transformative rungs are about *feedback* (rungs 8–6: strengthen a balancing loop, slow a reinforcing one, add a missing information channel), then about *structure and rules* (5 – 4), and finally about *purpose and worldview* (3–1). Meadows' own favourite illustration of a high-leverage information fix is almost too neat: a Dutch housing development where, by accident of design, some homes had their electricity meters in a visible front hallway and others tucked in the basement. The visible-meter households used markedly less power — perhaps a third less — for no reason but that a feedback loop which had been hidden was suddenly in plain sight. No tax, no rule, no technology. Just information completing a loop. That is leverage.

The same traps, over and over

Spend long enough drawing causal loop diagrams and you notice something liberating: the same handful of loop-*shapes* generate a startling fraction of the world's chronic problems. Forrester, Meadows, and Peter Senge catalogued these recurring structures as *system archetypes* — the stock characters of dysfunction. Learn to recognize a few and you can often diagnose a stuck situation on sight, and (more usefully) predict where the obvious fix will backfire. Here are the most useful, each as a one-breath loop story.

Tragedy of the Commons · R

each user gains -> all overuse -> shared resource collapses -> everyone loses When a resource is shared but the gains are private, each rational actor takes a little more, and the reinforcing pull of individual benefit grazes the common pasture bare. The fix lives high on the ladder: change the rules or the information so the cost of overuse is felt by the user.

Shifting the Burden · B

problem -> quick symptomatic fix -> relief -> real capacity atrophies -> problem worsens The “addiction” archetype. A symptomatic fix relieves the pressure that might have driven a fundamental solution – so the underlying capacity withers and dependence deepens. The relief is real; that’s the trap.

Fixes That Fail · B

problem -> fix -> quick relief -> delayed side-effect -> problem returns, bigger A solution works beautifully in the short run and quietly sows a long-run consequence that brings the problem back worse. Borrowing to cover a shortfall; spraying a pest that wipes out its predators. The delay is what hides the cost.

Success to the Successful · R

A wins early -> A gets more resources -> A wins more -> B starves When winners are rewarded with the means to win again, small initial differences amplify into runaway dominance – the rich-get-richer loop behind market monopolies, academic prestige, and unequal schools. Pure positive feedback, scaled.

Limits to Growth · R+B

growth engine (R) speeds up -> hits a hidden constraint (B) -> growth stalls or crashes Every reinforcing success eventually meets a balancing limit it didn’t see coming – a resource, a market size, a carrying capacity. The World3 model is this archetype written at planetary scale. Growth never simply continues; it negotiates with a ceiling.

Escalation · R

A acts -> B counter-acts -> A escalates -> B escalates -> both worse off Two balancing loops, each trying to restore *its own* sense of safety, couple into a single reinforcing spiral: arms races, price wars, social-media outrage cycles. No one is irrational; the structure does the damage.

The deep point isn’t the catalogue itself – it’s that *structure drives behaviour*. In every archetype, well-meaning people making locally sensible choices produce a collectively terrible outcome, not because anyone is foolish but because the loops they’re embedded in steer them there. Change the people and the trap persists; change the structure and the behaviour changes itself. That is the whole creed of systems thinking in a sentence.

How resilience dies — three different ways

The main file gave you one image of a tipping point: a valley flattening until the ball rolls out. That’s true, but it’s only one of three distinct ways a system can tip, and conflating them causes real confusion in the climate debate. First, though, we need a sharper idea of what’s actually being lost when a system “tips”– and for that we go back to a foundational 1973 paper by the ecologist **C. S. “Buzz” Holling**, who pried apart two things everyone had been calling “stability.”

Engineering resilience is how *fast* a system snaps back after a knock – the steepness of the valley walls. *Ecological resilience* is something else entirely: how *big* a shock the system can absorb before it flips into a different basin altogether – the *width* of the valley, the size of the disturbance it can swallow. These can pull in opposite directions, which is the part that matters. A system can snap back from small knocks beautifully (high engineering resilience) while sitting in a narrow basin that a single large shock would tip forever (low ecological resilience). Efficiency often buys the first by quietly spending the second. The main file measured a valley’s *depth*; Holling teaches you to also watch its *width*.

Fast return and large-shock tolerance are not the same thing.

Now, the three roads to collapse. Dynamical-systems theorists (notably Peter Ashwin and colleagues, in a clarifying 2012 paper) sorted tipping into three mechanisms, and the distinction is genuinely illuminating:

Three ways to tip: threshold, shock, and change applied too fast.

Bifurcation tipping (B-tipping) is the one we already know: push a driver past a threshold and the stable state vanishes. **Noise-induced tipping (N-tipping)** is when the driver is still safely below threshold, but a large enough random fluctuation – a freak heatwave, a once-a-century storm – kicks the system over the ridge anyway. And then the genuinely counterintuitive one: **rate-induced tipping (R-tipping)**, where what matters is not *how far* the driver moves but *how fast*. A system that could comfortably track a slow change can be tipped by the *same* change applied too quickly – the valley floor tilts faster than the ball can roll to keep up, and it spills out even though a leisurely version of the identical shift would have been survived. This is sobering for climate precisely because it means a threshold in *temperature* isn't the only danger; a threshold in the *rate* of warming may matter just as much. Ecosystems and ice don't only care where we end up. They care how fast we get there.

Dominoes, and the wheel of renewal

Two final ideas round out the deep theory. The first is the one that keeps climate scientists up at night: *tipping cascades*. The Earth's tipping elements aren't independent – they're wired together, so one tipping can shove another toward its own edge. Modelling work from the Potsdam Institute (Wunderling, Donges and colleagues, 2021) found that interactions between elements tend, on balance, to *lower* the safe thresholds and raise the odds of domino effects, with the great ice sheets often acting as the first domino and a weakening Atlantic conveyor as a transmitter. The magnitudes are still genuinely uncertain ● PROMISING BUT UNSETTLED, but the direction of the finding is the worrying one: coupled systems are more fragile than the sum of their parts – emergence, from Day 8, wearing its darkest coat.

The second idea is strangely consoling. The ecologist Holling, late in his career, proposed that complex systems don't just sit in basins or tip out of them – they move through a recurring *adaptive cycle*: a phase of rapid growth, then a long phase of rigid accumulation and efficiency, then a sudden release or collapse (he marked it Omega, “creative destruction”), then a fertile reorganization from which new growth springs. Collapse, in this view, is not only catastrophe; it is also the system clearing the board for renewal, and these cycles nest inside one another across scales – a framework he called *panarchy*. It’s a reminder that “tipping” isn’t always the end of a story. Sometimes it’s the hinge between one story and the next.

Daisyworld: homeostasis with no one home

We end where systems thinking becomes almost philosophical. In the 1970s, the chemist James Lovelock proposed the *Gaia hypothesis*: that life and the Earth’s surface form a single self-regulating system, holding temperature, atmospheric chemistry, and ocean salinity within the narrow bands life needs – much as a body holds its internal milieu. The idea was beautiful and immediately attacked, on one devastating objection: it seemed to require *foresight*. How could a planet’s microbes “want” to regulate the climate? Evolution has no goals; natural selection acts on individual organisms, not on planets. Regulation for the good of the whole looked like teleology smuggled into biology – and biologists, rightly, don’t allow that.

Lovelock’s answer, built with Andrew Watson in 1983, is one of the most elegant toy models in all of science, and it settles the objection completely. It is called **Daisyworld**, and it has no biologist’s hand-waving in it at all – just feedback. Imagine a grey planet orbiting a slowly brightening star, seeded with two kinds of daisy: *black* daisies, which absorb sunlight and warm their surroundings, and *white* daisies, which reflect it and stay cool. Each daisy simply grows fastest near a comfortable temperature and dies off when it’s too hot or too cold. There is no foresight, no cooperation, no planetary “purpose”– only individual daisies selfishly growing where they grow best. Watch what their blind feedback does to the temperature of the entire world.

Daisyworld, as regulation without foresight

SOLAR LUMINOSITY	DAISY MIX	TEMPERATURE RESULT
Dim	More black daisies	Absorbs heat and warms the world
Moderate	Black and white daisies regulate together	Temperature stays on a habitable plateau
Too bright	Daisies die out	Regulation fails and temperature follows dead rock

That flat plateau is the whole point – and there is not a shred of foresight anywhere in the model. When the star is dim, only black daisies can get a toehold; they spread, darken the planet, and warm it – a reinforcing loop that bootstraps the world up to a livable temperature. As the star brightens and the planet would otherwise overheat, white daisies start to outcompete the black ones (they stay cooler, so they grow better in the heat), and as white coverage spreads it reflects more sunlight and *cools* the planet back down. The mix of daisies shifts automatically to counteract the changing sun. **Robust planetary self-regulation emerges from nothing but selfish local growth plus feedback** – exactly Day 8's "more is different," now keeping a world alive. Daisyworld doesn't prove Gaia is true; the strong claim that Earth's biosphere actively regulates the planet *for life's benefit* remains ● CONTESTED. But it demolishes the charge that such regulation would *require* foresight – and that is a profound result. Purpose-like behaviour can fall out of pure mechanism. The planet keeps its temperature for the same reason your body keeps its own: not because anything wants it to, but because the loops are wired that way.

What this appendix opened

- **Is requisite variety a hard limit on governance?** If Ashby is right, no control room – Cybersyn or otherwise – can ever match the variety of a whole economy. Does that doom central planning, or just demand that complexity be met with *distributed* control rather than one heroic dashboard?
- **How much is regulation actually predictive?** Allostasis says the body sees demand coming. The brain (Day 1) may do the same. How far down does anticipatory, feedforward control go – to single cells? To ecosystems?
- **Does rate-induced tipping change the climate target?** If systems can tip from the *speed* of change and not just its size, then "stay under 1.5°C" may be necessary but not sufficient – the *pace* of the approach could matter independently.
- **Is collapse sometimes renewal?** Panarchy reframes the Omega-phase as creative destruction. When is a tipping point a catastrophe to prevent, and when is it the hinge a rigid system needs to reorganize?
- **If Daisyworld self-regulates without foresight, what else does?** Markets, immune systems, language, perhaps minds – how much apparent "purpose" in nature is really just feedback we haven't yet diagrammed?

The appendix in three sentences

The history: feedback control grew from a wartime gun and a ferry-boat sketch into cybernetics, whose deepest law – Ashby's requisite variety – says a regulator must be as various as the trouble it faces, a law you can feel in your own body's ten thousand homeostatic loops and in Chile's doomed attempt to run an economy from a single futuristic control room.
The deep theory: resilience comes in two flavours (fast-return and big-shock-tolerant) that can trade off invisibly, and a system can tip three

different ways – by crossing a threshold, by a random kick, or, most sneakily, by being pushed too *fast* – with coupled tipping elements making the whole more fragile than its parts.
The capstone: Daisyworld proves that a planet can hold its temperature steady for ages with no foresight whatsoever, purely because selfish local growth plus feedback adds up to global self-regulation – homeostasis with no one home, and the cleanest demonstration in science that purpose-like order can emerge from blind mechanism.

“-> BACK TO THE MAIN THREAD”

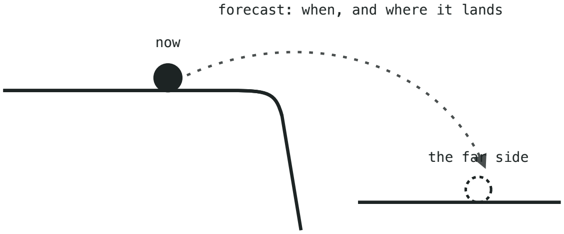
Everything here was a deepening of Day 9. The descent itself continues at **Day 10 – Models, Maps & Idealization**: we just trusted World3, the leverage ladder, and a planet made of daisies to teach us real things – so tomorrow we ask the unavoidable question lurking under all of it. *The map is not the territory; all models are wrong, but some are useful.* When is a simplification a lamp, and when is it a lie?

Sources & further reading

1. Wiener, N. (1948). *Cybernetics.* MIT Press. See also Galison, P. (1994). “The Ontology of the Enemy: Norbert Wiener and the Cybernetic Vision.” *Critical Inquiry* 21(1): 228–266. – the anti-aircraft predictor as the seed of feedback thinking.
2. Black, H. S. (1934). “Stabilized Feedback Amplifiers.” *Bell System Technical Journal* 13(1): 1–18. (invention conceived Aug 1927; U.S. Patent 2,102,671, 1937). The ferry-sketch origin of negative-feedback electronics.
3. Nyquist, H. (1932). “Regeneration Theory.” *Bell System Technical Journal*; Bode, H. (1945). *Network Analysis and Feedback Amplifier Design.* – the stability mathematics of feedback.
4. Macy Conferences on Cybernetics (1946–1953). See Heims, S. J. (1991). *The Cybernetics Group.* MIT Press; Pias, C. (ed.) (2016). *Cybernetics: The Macy Conferences 1946–1953.* – Wiener, von Neumann, McCulloch, Mead, Bateson, Ashby.
5. von Foerster, H. (1974). *Cybernetics of Cybernetics.* Univ. of Illinois. – second-order cybernetics; the observer inside the loop.
6. Ashby, W. R. (1956). *An Introduction to Cybernetics.* Chapman & Hall. – the Law of Requisite Variety; the homeostat. full text
7. Bernard, C. (1865). *Introduction a l’etude de la medecine experimentale.* – the *milieu interieur.*
8. Cannon, W. B. (1932). *The Wisdom of the Body.* Norton. – coins “homeostasis”(term introduced 1926).
9. Sterling, P. (2012). “Allostasis: A model of predictive regulation.” *Physiology & Behavior* 106(1): 5–15. (concept: Sterling & Eyer 1988). – stability through anticipatory change; the link to predictive regulation.
10. Medina, E. (2011). *Cybernetic Revolutionaries: Technology and Politics in Allende’s Chile.* MIT Press; Beer, S. (1972). *Brain of the Firm.* – Project Cybersyn (1971–73) and the Viable System Model.
11. Meadows, D. H., Meadows, D. L., Randers, J. & Behrens, W. W. (1972). *The Limits to Growth.* Universe Books. – the World3 model.

12. Turner, G. M. (2008). “A comparison of The Limits to Growth with 30 years of reality.” *Global Environmental Change* 18(3): 397–411; Turner (2014), MSSI Research Paper No. 4, Univ. of Melbourne. – data tracks the “standard run.”
13. Herrington, G. (2021). “Update to limits to growth: Comparing the World3 model with empirical data.” *Journal of Industrial Ecology* 25(3): 614–626. doi:10.1111/jiec.13084 – empirical data closest to BAU2 / comprehensive-technology runs. doi.org/10.1111/jiec.13084
14. Meadows, D. H. (1999). *Leverage Points: Places to Intervene in a System.* The Sustainability Institute. – the ranked twelve; the Dutch electricity-meter example (also in *Thinking in Systems*, 2008). donellameadows.org
15. Senge, P. (1990). *The Fifth Discipline.* Doubleday. – the system archetypes (with Forrester & Meadows lineage).
16. Holling, C. S. (1973). “Resilience and Stability of Ecological Systems.” *Annual Review of Ecology and Systematics* 4: 1–23. – engineering vs ecological resilience. Gunderson, L. & Holling, C. S. (2002). *Panarchy.* Island Press. – the adaptive cycle.
17. Ashwin, P., Wieczorek, S., Vitolo, R. & Cox, P. (2012). “Tipping points in open systems: bifurcation, noise-induced and rate-dependent examples in the climate system.” *Phil. Trans. R. Soc. A* 370: 1166–1184. doi:10.1098/rsta.2011.0306 – the B / N / R taxonomy. doi.org/10.1098/rsta.2011.0306
18. Wunderling, N., Donges, J. F., Kurths, J. & Winkelmann, R. (2021). “Interacting tipping elements increase risk of climate tipping cascades.” *Earth System Dynamics* 12: 601–619. doi:10.5194/esd-12-601-2021 – coupling lowers thresholds; ice sheets as initiators. doi.org/10.5194/esd-12-601-2021
19. Watson, A. J. & Lovelock, J. E. (1983). “Biological homeostasis of the global environment: the parable of Daisyworld.” *Tellus B* 35(4): 284–289. doi:10.1111/j.1600-0889.1983.tb00031.x – self-regulation without foresight. doi.org/10.1111/j.1600-0889.1983.tb00031.x

Optional appendix: Catching the Cliff



The old science could feel the ground softening; the new science wants the date, depth, and far-side map.

For sixty years, the science of tipping points could do one honest thing: feel the ground going soft. As the main descent showed, a system approaching a cliff *slows down* – it wobbles wider and recovers slower, and those statistical tremors (rising variance, rising autocorrelation) are a genuine early-warning signal. But that signal is frustratingly mute. It tells you that *something* is coming without telling you *what*, or *when*, or what the world looks like on the far side. Since about 2020, a burst of work – much of it leaning on machine learning, satellites, and fresh mathematics – has been trying to turn that mute tremor into something that speaks. This appendix is a field report from that frontier: five live research programs from the last six years, each genuinely exciting, none yet finished, and all run through the hype filter without mercy. Treat every claim here as a dispatch from a moving front, not a settled map.

“RETURN WHAT YOU ALREADY HAVE (AND WE WON’T REPEAT)”

From the **main file**: fold bifurcations, hysteresis, critical slowing down, the generic early-warning signals (Scheffer et al. 2009), Ben-Yami et al. 2024 on why dates are hard, and the climate specifics (AMOC, Amazon, coral, the 2022 threshold table). From **Appendix I**: Holling resilience, the B/N/R tipping taxonomy, tipping cascades, Daisyworld. Here we build *on* all of it – and only cover work published **2020 or later**.

● FRONTIER

An AI that knows the shape of catastrophe

Start with the deepest problem in the classic warning signals: they are *generic to a fault*. Rising variance and autocorrelation appear before almost any kind of tipping, which is their strength – they work everywhere – but also their weakness: they cannot tell you which *kind* of tipping is coming, and so they cannot tell you what the new world will be like. Will the system jump abruptly to a far-off state? Start oscillating wildly? Slide gently into a new regime? The generic signal shrugs at all three.

In 2021, a team led by Thomas Bury, with Chris Bauch, Madhur Anand, and – note the name – **Marten Scheffer and Tim Lenton** themselves (the founders of the classic approach, now helping to supersede it), published a striking result in *PNAS*. Their insight rests on a beautiful piece of mathematics from the main file's world: as *any* system nears a tipping point, its dynamics collapse onto one of a small number of universal templates called *normal forms* (the fold, the Hopf, the transcritical – the basic ways a stable state can lose stability). There are only a handful. So Bury's team trained a deep neural network – a *convolutional-plus-recurrent* hybrid – on hundreds of thousands of simulated time series drawn from these universal templates, teaching it the subtle fingerprints of *each* route to collapse.

The payoff: their network spots tipping points in systems it was *never trained on* – including real climate and ecological data and a thermoacoustic experiment – more sensitively and with fewer false alarms than the classic indicators. And crucially, it does the thing the old signals can't: it names the *type* of bifurcation, telling you whether the new state will be a sudden jump, an oscillation, or a gentle shift. Bauch called it a potential "game-changer for the ability to anticipate big shifts." A 2025 follow-up (Zhuge, Li & Chen, *Royal Society Open Science*) pushed the approach toward the messiness of reality, designing a network to predict tipping occurrence even in the *irregularly sampled* time series that real-world monitoring actually produces – one of the sharpest limitations of the original method.

Tipping classifiers, as three routes

ROUTE	SHAPE	LANDING
Fold	Slow approach, then sudden jump	System lands in a distant state
Hopf	Oscillation grows	The future is a cycle, not a new level
Transcritical	Stabilities exchange	One state slides out as another takes over

The honest label: this is a real and important advance, but its triumphs are still mostly *in silico* and on a handful of empirical cases. Whether a normal-form-trained network reliably calls tipping points in the full, noisy, high-dimensional mess of the real Earth system – with enough lead time to matter – remains to be demonstrated at scale. ● PROMISING HINT — STRONG IN SIMULATION, REAL-WORLD TRACK RECORD STILL THIN

The digital twin that visits the far side

Detecting that a cliff is near is one thing. A bolder 2020–2021 program asks: can a machine, having only ever watched a system behave *normally*, predict the exact point at which it will collapse – and then describe the alien world beyond? The tool is *reservoir computing*, a strikingly efficient flavour of recurrent neural network that has proven almost uncannily good at learning the "grammar" of chaotic systems from raw data, with no model of the underlying physics.

In a 2021 *Physical Review Research* paper, Ling-Wei Kong, Ying-Cheng Lai, and colleagues did something that sounds like cheating. They fed a reservoir computer not just the system's behaviour but also the value of the slowly drifting *parameter* pushing it toward the edge – a "parameter input channel." Trained only on data from the safe regime, the machine could then **extrapolate** to parameter values it had never seen, predicting both *where* the critical transition lies and – remarkably – statistical properties of the doomed *transient* state that precedes final collapse. Follow-up work (Dhruvit Patel and Edward Ott, *Chaos*, 2023) explicitly used machine learning to "anticipate tipping points and extrapolate to post-tipping dynamics" of changing systems, and Kong's group has gone on to frame these trained reservoirs as *digital twins* of nonlinear systems – fast, data-driven stand-ins you can push past the cliff in software to see what happens, without pushing the real thing.

> *The dream is a wind tunnel for catastrophe: a learned replica you can crash a thousand times to map the edge of a system you must never actually break.*

It is a genuinely beautiful idea, and on benchmark chaotic systems it works. But here the hype filter bites hardest. These successes are on relatively low-dimensional, well-sampled, cleanly-measured mathematical systems. The real climate is high-dimensional, sparsely observed, and non-stationary in ways that can defeat the method's core assumption that the future "rhymes" with the trained past. The same community has published sober "Catch-22" analyses of reservoir computing's limits. A digital twin of a chaotic toy is a triumph; a trustworthy digital twin of the AMOC does not yet exist. ● PROMISING ● REAL-EARTH APPLICATION — CONTESTED/UNPROVEN

How a pattern lets a system dodge the edge

Now a result that doesn't predict tipping points so much as *question* them – and it may be the most quietly radical idea in this whole appendix. The standard picture, the one the main file sold you, is stark: push an ecosystem past its threshold and it collapses, abruptly and as a whole, from green to desert. In 2021, Max Rietkerk and colleagues argued in *Science* that for systems spread out in *space*, this picture is often wrong – and wrong in a hopeful direction.

Their claim: instead of collapsing all at once, a spatially extended system under stress can **reorganize into a pattern** – the spots, stripes, and labyrinths of vegetation you can see from the air in drylands the world over. These are *Turing patterns*, the self-organized structures Alan Turing predicted in 1952. And here is the twist that upends the textbook: where ecologists had long read these patterns as *warning signs* of imminent collapse, Rietkerk's analysis says they can be the opposite – a **sign of resilience**. By breaking into patches, the system survives at stress levels far beyond where a uniform version would have tipped. It trades total collapse for a graceful, often reversible retreat into pattern. The cliff is replaced by a ramp.

Pattern evasion, as two system designs

MODE	AS STRESS RISES	LESSON
Uniform	Collapses once the threshold is crossed	No spatial structure remains to retreat into
Patterns allowed	Forms patches and persists under higher stress	Self-organization can evade a tipping point for a while
Stress too high	Patterns thin out and fail	Evasion is not immunity; it buys time

Why it matters, and where the caution lives: if Rietkerk is right, then a whole class of ecosystems – and perhaps Earth-system components – flagged as "tipping-prone" may be substantially *more* resilient than the catastrophic-collapse models imply, because those models ignored spatial dynamics. That is genuinely good news, carefully argued with both mathematics and field observations. But it is a *review and a theoretical framework*, not a universal law: how far the escape hatch generalizes – to ice sheets, to the Amazon, to the whole Earth system – is exactly the open question its authors pose. And it cuts both ways for warning signals, because it means a pattern can signal either approaching collapse *or* hard-won persistence, and telling those apart is hard. ● PROMISING – AND PARADIGM-SHIFTING IF IT GENERALIZES

Reading the Earth's resilience from orbit

The classic warning signals were born on small, clean datasets – a whole-lake experiment, an ice core. The post-2020 leap is to compute them *everywhere at once*, by turning decades of satellite imagery into a planetary resilience monitor. The logic is unchanged from the main file – a system losing resilience recovers more slowly, so its autocorrelation creeps up – but the canvas is now the entire land surface, pixel by pixel.

The results have arrived in a rush. In 2021, Niklas Boers found statistically significant early-warning signals across *eight independent* observational fingerprints of the AMOC, concluding the Atlantic circulation may have drifted, over the last century, "from relatively stable conditions to a point close to a critical transition"(*Nature Climate Change*). The same year, Boers and Rypdal reported critical slowing down suggesting the western Greenland Ice Sheet sits near a tipping point (*PNAS*). Then the satellites opened up the biosphere: in 2022, two landmark papers – Smith, Traxl & Boers in *Nature Climate Change*, and Forzieri and colleagues in *Nature* – used satellite vegetation indices to show that forests across much of the planet have been **losing resilience since the early 2000s**, most worryingly in the tropics, even as some northern forests gained it. Tim Lenton's group has proposed building all of this into a permanent "*resilience sensing system for the biosphere*"(2022) – an early-warning dashboard for the living planet.

The reliability problem — read this before you panic or relax

This is the frontier where it is easiest to over-read a result, so the caveats are load-bearing. A 2024 study in *Nature Ecology & Evolution* showed that satellite resilience estimates become *unreliable in high-biomass regions* like dense tropical forest – precisely where we most want to trust them – partly because thick canopies saturate the optical signal. Other 2023 – 2025 work has flagged that data gaps, outliers, and the choice of instrument can distort these indicators, and a 2025 *Nature Climate Change* analysis argued that early-warning signals for climate tipping points can be fundamentally **ambiguous**. The signals are real and the global pattern of declining resilience is striking – but a rising autocorrelation in one noisy pixel is a hypothesis, not a verdict.

And the AMOC story carries its own built-in counterweight, which is exactly how healthy science should look. Even as the observational warning signals accumulated, a 2025 *Nature* study argued that wind-driven upwelling in the Southern Ocean may stabilize the Atlantic circulation enough to make an outright collapse *this century* unlikely. Both findings are serious; both are peer-reviewed; the disagreement is the actual state of the science. The fair 2026 summary: real, observation-based evidence of resilience loss is now being read off the planet for the first time ● GENUINELY NEW CAPABILITY, but converting those signals into confident statements about *which* system tips *when* remains contested ● CONTESTED.

Setting off the cascades we want

Every tipping point in this course so far has been a thing to fear or forecast. The final frontier program flips the sign. If a reinforcing loop can drag a system into ruin, the reasoning goes, then a reinforcing loop can also be deliberately *seeded* to flip a system toward something good – and fast. This is the science of *positive tipping points*, and it has become one of the most active and consequential corners of systems thinking since 2020, precisely because the math of decarbonization demands it: holding warming near 1.5°C requires emissions to fall faster than any gradual, linear policy is delivering. Only self-accelerating change is fast enough.

The agenda-setting paper was Ilona Otto, Jonathan Donges, and colleagues in *PNAS* (2020), who convened experts to identify “social tipping elements”– leverage-rich subsystems where a modest, well-placed push could trigger runaway change toward carbon neutrality. Their candidates: energy markets, financial divestment, carbon-neutral cities, norms and values, education, and – note the callback to Appendix I’s leverage ladder – *information disclosure*. Tim Lenton’s group followed with a framework for “*operationalising*“these points (2022) and a method to identify them while, in their own careful words,” avoiding wishful thinking about their existence.” The clearest real-world example is already here: the collapsing cost of solar power and batteries has, in sector after sector, crossed the point where clean energy simply out-competes fossil fuels, triggering self-reinforcing adoption – an **S-curve**, the signature of a reinforcing loop winning. Reinforcing feedback, the villain of the climate story, recast as its best hope.

The hype filter here is unusual, because the object is partly *normative* – a goal as much as a fact. Some positive tipping points are demonstrably real and already past (the UK’s exit from coal; solar’s cost crossover). Others are hopeful conjectures about social systems whose contagion dynamics we understand far less well than a melting ice sheet – and critics (in replies to Otto et al.) warned that treating society like a physical system risks ignoring conflict, politics, and “how change actually happens.” ● SOME CASES ESTABLISHED ● THE GENERAL PROGRAM – PROMISING BUT PARTLY ASPIRATIONAL

Six years, one accelerating field

Colour: prediction, evasion and resilience-sensing, positive tipping, and reality checks.

Where each claim actually stands, mid-2026

- *Deep learning can detect tipping & name the bifurcation type.* Bury et al., PNAS 2021; Zhuge et al., R. Soc. Open Sci. 2025. *Verdict:* promising. Excellent in simulation & a few empirical cases; broad real-world reliability unproven.
- *Reservoir computing predicts collapse & post-tipping dynamics.* Kong & Lai, Phys. Rev. Research 2021; Patel & Ott, Chaos 2023. *Verdict:* promising, contested at scale. Works on chaotic benchmarks; trustworthy Earth-system “digital twins” don’t yet exist.
- *Spatial patterns let systems evade catastrophic tipping.* Rietkerk et al., Science 2021. *Verdict:* promising / paradigm-shifting. Well-argued theory + field cases; generality to ice sheets/whole Earth is open.
- *We can read resilience loss off satellites globally.* Smith et al. & Forzieri et al., 2022; Boers, Nat. Clim. Change 2021. *Verdict:* new capability, caveated. Striking global pattern; estimates unreliable in dense forest, signals can be ambiguous.
- *Positive tipping points can be triggered to speed decarbonization.* Otto et al., PNAS 2020; Lenton et al., 2022. *Verdict:* some cases real, partly aspirational. Solar/coal cases established; social contagion far less predictable than physics.
- *But can we name the collapse date?* Ben-Yami et al., Sci. Adv. 2024 (from the main file). *Verdict:* not yet. Detecting lost resilience ≠ forecasting timing. The deepest unsolved problem of all.

What the next six years must answer

- **Can any method give a trustworthy lead time?** Every program here improves *detection*; none has convincingly cracked *timing* for a real, half-observed Earth-system component. This is the field's central, stubborn gap.
- **Will AI predictors survive contact with the real Earth?** Normal-form networks and reservoir twins shine on clean systems. The open test is whether they hold up on sparse, noisy, non-stationary planetary data – or quietly overfit the past.
- **How far does the spatial escape hatch reach?** Rietkerk's evasion is demonstrated for drylands. Does it extend to forests, ice, ocean circulation – or are those the systems where the cliff is real after all?
- **Can a falling pattern be told from a rising one?** If spatial patterns can mean either imminent collapse *or* resilient persistence, every spatial early-warning signal now needs a way to read its sign. Several 2024 – 25 groups are racing at exactly this.
- **Do social systems really tip like physical ones?** The positive-tipping program borrows the mathematics of ice sheets for human behaviour. Where that analogy holds – and where politics and conflict break it – is being tested in real time, on the climate clock.

The frontier in three sentences

The leap: since 2020, machine learning (normal-form-trained deep nets; parameter-aware reservoir "digital twins") and planetary-scale satellite analysis have begun turning the mute tremor of critical slowing down into something that can name the *type* of tipping, rehearse the world beyond it, and map resilience loss across the whole Earth.
The twist: the same era produced a hopeful counter-current – Rietkerk's demonstration that spatial self-organization can let systems *evade* catastrophic tipping entirely, and a serious program to deliberately trigger *positive* tipping points for rapid decarbonization.
The honest bottom line: detection is improving fast and is genuinely new, but forecasting a specific collapse *date* for a real system remains unsolved, and the most exciting tools are still proving themselves against clean simulations rather than the noisy planet – promise, not yet proof.

"-> BACK TO THE MAIN THREAD"

This was the bleeding edge of Day 9. The descent itself resumes at **Day 10 – Models, Maps & Idealization** – which is the perfect next question, because nearly every tool in this appendix is a *model* betting that the future will rhyme with the trained past. When is that bet a lamp, and when is it a lie? Tomorrow we interrogate the models themselves.

Sources & further reading

1. Bury, T. M., Sujith, R. I., Pavithran, I., Scheffer, M., Lenton, T. M., Anand, M. & Bauch, C. T. (2021). "Deep learning for early warning signals of tipping points." *PNAS* 118(39): e2106140118 (28 Sep 2021). doi:10.1073/pnas.2106140118 — DL trained on bifurcation normal forms; detects & classifies tipping in untrained systems. doi.org/10.1073/pnas.2106140118
2. Zhuge, C., Li, J. & Chen, W. (2025). "Deep learning for predicting the occurrence of tipping points." *Royal Society Open Science* 12(7): 242240 (1 Jul 2025). doi:10.1098/rsos.242240 — extends DL prediction to irregularly sampled real-world time series. doi.org/10.1098/rsos.242240
3. Deb, S., Sidheekh, S., Clements, C. F., Krishnan, N. C. & Dutta, P. S. (2022). "Machine learning methods trained on simple models can predict critical transitions in complex natural systems." *Royal Society Open Science* 9(2): 211475. doi:10.1098/rsos.211475.
4. Kong, L.-W., Fan, H.-W., Grebogi, C. & Lai, Y.-C. (2021). "Machine learning prediction of critical transition and system collapse." *Physical Review Research* 3(1): 013090 (28 Jan 2021). doi:10.1103/PhysRevResearch.3.013090 — reservoir computing with a parameter channel predicts the transition & transient lifetimes. doi.org/10.1103/PhysRevResearch.3.013090
5. Patel, D. & Ott, E. (2023). "Using machine learning to anticipate tipping points and extrapolate to post-tipping dynamics of non-stationary dynamical systems." *Chaos* 33(2): 023143. doi:10.1063/5.0131787. See also Kong, L.-W. et al. (2023), "Reservoir computing as digital twins for nonlinear dynamical systems." *Chaos* 33(3): 033111.
6. Rietkerk, M., Bastiaansen, R., Banerjee, S., van de Koppel, J., Baudena, M. & Doelman, A. (2021). "Evasion of tipping in complex systems through spatial pattern formation." *Science* 374(6564): eabj0359 (8 Oct 2021). doi:10.1126/science.abj0359 — Turing patterns as resilience, not warning. doi.org/10.1126/science.abj0359
7. Boers, N. (2021). "Observation-based early-warning signals for a collapse of the Atlantic Meridional Overturning Circulation." *Nature Climate Change* 11(8): 680–688 (5 Aug 2021). doi:10.1038/s41558-021-01097-4 — EWS in 8 independent AMOC indices. doi.org/10.1038/s41558-021-01097-4
8. Boers, N. & Rypdal, M. (2021). "Critical slowing down suggests that the western Greenland Ice Sheet is close to a tipping point." *PNAS* 118(21): e2024192118. doi:10.1073/pnas.2024192118.
9. Smith, T., Traxl, D. & Boers, N. (2022). "Empirical evidence for recent global shifts in vegetation resilience." *Nature Climate Change* 12(5): 477–484. doi:10.1038/s41558-022-01352-2. See also Smith, T. & Boers, N. (2023), *Nature Communications* 14: 498.
10. Forzieri, G., Dakos, V., McDowell, N. G., Ramdane, A. & Cescatti, A. (2022). "Emerging signals of declining forest resilience under climate change." *Nature* 608: 534–539. doi:10.1038/s41586-022-04959-9.
11. Lenton, T. M., Buxton, J. E., Armstrong McKay, D. I. et al. (2022). "A resilience sensing system for the biosphere." *Phil. Trans. R. Soc. B* 377(1857): 20210383 (15 Aug 2022). doi:10.1098/rstb.2021.0383.
12. Smith, T., Zotta, R.-M., Boulton, C. A., Lenton, T. M., Dorigo, W. & Boers, N. (2023). "Reliability of resilience estimation based on multi-instrument time series." *Earth System Dynamics* 14(1): 173–183. — instrument choice affects resilience estimates.
13. Smith, T. et al. (2024). "Reliability of vegetation resilience estimates depends on biomass density." *Nature Ecology & Evolution* 8: 740–749. doi:10.1038/s41559-023-02194-7 — estimates unreliable in high-biomass forest.

14. "Ambiguity of early warning signals for climate tipping points." *Nature Climate Change* (2025). — argues climate-tipping EWS can be fundamentally ambiguous. (One of several 2024–25 reliability critiques; see also Lapeyrolerie & Boettiger 2021.)
15. Otto, I. M., Donges, J. F., Cremades, R. et al. (2020). "Social tipping dynamics for stabilizing Earth's climate by 2050." *PNAS* 117(5): 2354–2365 (21 Jan 2020). doi:10.1073/pnas.1900577117 — six social tipping elements. See critical replies: Smith et al. & Willis, PNAS 2020. doi.org/10.1073/pnas.1900577117
16. Lenton, T. M., Benson, S., Smith, T. et al. (2022). "Operationalising positive tipping points towards global sustainability." *Global Sustainability* 5: e1. doi:10.1017/sus.2021.30.
17. "Continued Atlantic overturning circulation even under climate extremes." *Nature* (2025). — Southern-Ocean upwelling may make AMOC collapse this century unlikely; the peer-reviewed counterweight.
18. Ben-Yami, M., Morr, A., Bathiany, S. & Boers, N. (2024). "Uncertainties too large to predict tipping times of major Earth system components from historical data." *Science Advances* 10(31): eadl4841. doi:10.1126/sciadv.adl4841 — the standing limit on date-prediction (also cited in the main file).