

The Asymmetry Engine, Chapter 18: The pale dot

Martin Roark

Chapter 18 of *The Asymmetry Engine: How AI Closes the Gaps, and What Stays Scarce*, October 2026. The online edition, with the interactive figures and the links between chapters, is at <https://future-seems-so-good.com/blog/the-asymmetry-engine#pale-dot>. License: [CC BY 4.0](#).

18:1 The asymmetries close in the order of their slowest factor other than intelligence. Intelligence sets the pace of every closing rate, and verifiability, alignment and inertia set their order: checking closes first, then trusting, then building, and distributing what was closed comes last, while the open gap and its rent migrate to whatever is hardest to verify, trust or build. Seen from far enough away, the whole build-out is a small fraction of the light that falls on one planet. Humility at that scale has three working forms, calibration, deference and perspective, and on those conditions the future seems good.

18.1 The order of closing

18:2 Because the master equation makes each closing rate a product, $\lambda_k = I_k v_k a_k / \varphi_k$ ((0.2)), intelligence sets the pace of the closing and the slow factors set its order. Taking logarithms and differentiating:

18:3

$$\frac{d \ln \lambda_k}{dt} = \frac{d \ln I_k}{dt} + \frac{d \ln v_k}{dt} + \frac{d \ln a_k}{dt} - \frac{d \ln \varphi_k}{dt}, \quad \frac{\lambda_k}{\lambda_j} = \frac{s_k v_k a_k \varphi_j}{s_j v_j a_j \varphi_k} \quad (I_k = s_k) \quad (18.1)$$

18:4 A closing rate grows at the sum of its factors' growth rates, and when intelligence is allocated in fixed shares s_k of one supply I , the ratio of two closing rates does not depend on I at all. This is the change of clock of (0.3) seen from the other side: in intelligence-time the gap dynamics are linear, and the conductances decide which gap goes first. Intelligence is also the fastest factor. The price of a fixed capability keeps falling fast (3:48), the largest labs' compute grows severalfold a year (§10.1), automated research raises the supply itself (Proposition 11.1), and every gain multiplies all the λ_k alike. The other factors run on slower clocks: verifiers are built domain by domain (Proposition 14.1), alignment rises only as fast as evaluations certify drift (Definition 13.2), and inertia falls with permits, training and the turnover of capital (§15.1).

18:5 Sorted by the factor that governs them, the book's families fall into four groups. Compute, latency, energy and research are governed by the supply of intelligence and close fastest, and what binds each is something intelligence does not supply for itself: the hands (§3.4), the per-step floor of decoding (§5.3), the cache bytes each agent holds (§10.6), and the experiment compute that makes compute the homeostat of the research loop (Proposition 11.2). Data, swarms, security and science are governed by verifiability: training signal is harvestable where a cheap, sound verifier exists (Proposition 2.2), a fixed verifier caps a swarm (Experiment 9.1), security finds fast and fixes slowly (§12.4), and science is ordered by the latency of its verifiers (§14.3).

18:6 Topology, character, attention and the hours of human work are governed by alignment. At an alignment of 0.2 the best autonomy in the topology model is zero at every level of intelligence (Experiment 7.1), and because a_k multiplies every closing rate, alignment multiplies all the gaps and is the slowest digital term (Forecast 13.3).

Variety, physical work, capital and distribution are governed by inertia: the long tail goes to people (Proposition 4.2), a gap with a physical part closes at the physical rate however much intelligence it gets (Proposition 15.1), rents pool at the stiffest layers (Proposition 16.1), and distribution follows diffusion lags of decades (17:3).

18:7 Down the list of what binds, five inputs recur: verification, alignment, atoms, energy and human attention. Intelligence cannot supply them for itself, and they enter the closing rate as complements, so the scarcest governs (Definition 15.2), and raising intelligence while they stay fixed drives the intelligence elasticity of closing to zero (Proposition 15.1).

▼ *The families and what binds them*

Family	Home	Gap: the value locked behind	Factor	What binds
Data	§2.5	solutions cheap to check, costly to find	v	verifier soundness
Compute	§3.1	frontier prices paid for sufficient work	I	the hands
Variety	§4.5	long-tail requests	I, φ	the tail sent to people
Latency	§5.1	human waiting	I	the per-step floor
Attention	§6.3	decisions settled by default	a, φ	the architect's objective
Topology	§7.2	effectiveness lost to the wrong hierarchy	a	alignment; error correlation
Hours and egos	§8.6	the inertia of human work	φ, a	congruent gauges
Swarms	§9.4	parallel work not yet merged	v	verifier capacity
Energy	§10.3	demand for concurrent agents	I	cache per agent; power
Research	§11.3	the research gap	I	experiment compute
Security	§12.4	latent vulnerabilities	v, σ	remediation
Character	§13.6	autonomy withheld	a	evaluations of drift
Science	§14.3	discoveries awaiting a verdict	v	wet labs and trials
Physical work	§15.2	the physical part of closing	φ	hands, permits, grid
Capital	§16.3	build-out capacity	I, φ	the stiffest layers
Distribution	§17.2	the incidence of $\lambda_k G_k$	φ, σ, a	diffusion; ownership

18:8 Read by date, as of October 2026, the book's forecasts trace the same order. Those for 2027 concern gaps governed by intelligence where verifiability is near one: elasticity (Forecast 0.1), doubling times (Forecast 11.5), fast modes (Forecast 5.3), caches (Forecast 10.3) and formal mathematics (Forecast 14.1). Those for 2028 turn to verification and alignment as they become priced inputs: swarm verification (Forecast 9.2), remediation (Forecast 12.3), review as a layer of the stack (Forecast 16.1) and cross-lineage monitors (Forecast 13.1). Those for 2029 to 2031 concern energy per token, interfaces, protocols, attention and the labor share (Forecast 10.2, Forecast 5.4, Forecast 15.3, Forecast 6.4, Forecast 17.1), and the last, around 2035, is the humanoid that would end Baumol's cost disease in physical services (Forecast 15.4). Whether they also come true in that order is a reading, not a scored forecast, because their horizons were set by the same argument; Forecast 18.1 tests the order on the world instead, across occupations.

18:9 The asymmetries close in the order of their slowest factor other than intelligence: checking first, then trusting, then building, and distributing what was closed last.

Forecast 18.1 (The slow factors set the order). By the end of 2031, the depth of AI adoption across US occupations follows verifiability and inertia more than benchmark scores: legal and healthcare-practitioner occupations, where models pass licensing exams but outputs lack cheap verifiers and face regulation, rank below computer and mathematical occupations and below business and financial operations in the share of work hours AI assists.

Horizon 2031-12-31

Probability 50%

Check Use the occupation breakdown in the latest report of the St. Louis Fed's survey of AI use at work, or its successor, published by the horizon. The forecast holds if the share of work hours AI assists is lower for both legal occupations and healthcare practitioners than for both computer and mathematical occupations and business and financial operations; it fails otherwise.

18.2 The residue

18:11 Applied intelligence leaves the open gap in the slowest family, and the rent follows it there. To see how fast, hold the closing rates constant, which fixes the supply of intelligence; in units where intelligence-time then equals calendar time, $\lambda_k = \kappa_k$. Index the families so that $\lambda_1 < \lambda_2 \leq \dots$, with family 1 the slowest to close.

18:12

Proposition 18.1 (Residue migration). The **residue** is the open gap that remains as intelligence is applied.

Let $h_k = G_k/\Gamma$ be family k 's share of the total open gap Γ , with $G_1(0) > 0$. (i) Without regeneration, the slowest family's share tends to one,

$$h_1(t) = \left[1 + \sum_{k \geq 2} \frac{G_k(0)}{G_1(0)} e^{-(\lambda_k - \lambda_1)t} \right]^{-1} \rightarrow 1,$$

exceeding $1 - \epsilon$ for every $t \geq t_\epsilon = \frac{1}{\lambda_2 - \lambda_1} \ln \frac{G_{\text{rest}}(1 - \epsilon)}{G_1(0)\epsilon}$, where $G_{\text{rest}} = \sum_{k \geq 2} G_k(0)$; its share of the

realized value flow $\dot{W}$ also tends to one, and the total gap closes at the slowest rate. (ii) With regeneration

$\sigma_k = \bar{\sigma} \omega_k$, the composition converges instead to $h_k \propto \omega_k/(\lambda_k + g^*)$, where g^* solves

$\bar{\sigma} \sum_k \omega_k \lambda_k/(\lambda_k + g^*) = 1$ as in Proposition 0.1: the slowest family holds the most open gap per unit of regenerated demand, but not all of it. (iii) Rent follows the residue. If each increment of value realized in family k divides among the inputs j of its closing chain in the shares $\omega_{j,k}$ of Proposition 16.1, and closers enter freely, input j 's share of all realized value tends to $\omega_{j,1}$, its share in the slowest family's chain. Where that chain's binding constraint is human work that compute cannot reproduce, the wage cap of Proposition 16.3 does not bind and labor earns the rent; otherwise the owners of compute and physical inputs do.

▼ Proof

(i) Without regeneration $G_k(t) = G_k(0)e^{-\lambda_k t}$. Divide the numerator and denominator of $h_1 = G_1/\Gamma$ by $G_1(0)e^{-\lambda_1 t}$; every remaining term carries a factor $e^{-(\lambda_k - \lambda_1)t} \leq e^{-(\lambda_2 - \lambda_1)t}$, so $h_1 \geq \left[1 + (G_{\text{rest}}/G_1(0)) e^{-(\lambda_2 - \lambda_1)t} \right]^{-1}$, which is at least $1 - \epsilon$ once $t \geq t_\epsilon$. The flow share is the same expression with each ratio $G_k(0)/G_1(0)$ multiplied by λ_k/λ_1 , and $\Gamma(t) = G_1(0)e^{-\lambda_1 t}(1 + o(1))$. (ii) With

regeneration the gaps obey $\dot{G} = \mathbf{J}G$ with $\mathbf{J} = -\text{diag}(\lambda) + \bar{\sigma} \omega \lambda^\top$, a Metzler matrix that is irreducible when every $\omega_k > 0$. By Perron–Frobenius its dominant eigenvalue g^* is real with a positive eigenvector h , and $(\text{diag}(\lambda) + g^*)h = \bar{\sigma} \omega (\lambda^\top h)$ gives $h_k \propto \omega_k / (\lambda_k + g^*)$ and the equation for g^* . As $\bar{\sigma} \rightarrow 0$, $g^* \rightarrow -\lambda_1$ and h concentrates on family 1, recovering (i). (iii) Input j receives $\sum_k \omega_{j,k} \lambda_k G_k$; divide by $\dot{W} = \sum_k \lambda_k G_k$ and apply the flow share of (i). The wage clause is part (iii) of Proposition 16.3. ■

18:13 A case with three families shows the speed, in the model. Digital, verifiable work closes at $\lambda = 1$ a year, work limited by alignment, such as management, at 0.25, and physical work at 0.05, and they start with 50%, 30% and 20% of the gap. The physical family supplies 1.7% of the value flow at the start. At year five the alignment-limited family supplies 66% of the flow. After ten years the physical family holds 83% of the open gap and half the flow, and after twenty it holds 97% of the gap and 88% of the flow, with the total gap down to 7.6% of where it began. Its share passes 90% at 13.0 years, inside the sufficient 17.9 years of part (i). The residue is Baumol's cost disease stated for intelligence, and Aghion, Jones and Jones show that such essential, hard-to-improve tasks can cap growth even under extensive automation (§15.2).

18:14 Regeneration changes the conclusion in degree. Closing reveals latent demand (Definition 0.1), of which the Jevons effect is the empirical form ((3.3)). With regeneration weights of 0.6, 0.3 and 0.1, sending most new demand to the digital family, the physical family's limiting share of the open gap is 91% at $\bar{\sigma} = 0.5$, 74% at 0.8 and 53% at 1, where the total gap is conserved; at $\bar{\sigma} = 1.2$ the gap grows 7.2% a year and the alignment-limited family holds the most, 40%. Strong regeneration pulls the residue toward wherever new demand arises, and the fast families churn the way markets do. Panel (b) of Figure 5.3 draws the same migration for the book's twelve task families.

18:15 The residue has an inventory and a payroll. The inventory is the five inputs: verification, whose cost per task falls more slowly than the price of tokens (Forecast 3.3); alignment (§13.6); atoms, with interconnection queues measured in years and builders the industry cannot hire (§15.3, §15.4); energy, with gas turbines booked years ahead (§16.2); and attention (§6.3). The payroll depends on ownership. Where the slow input is people, such as electricians and reviewers, wages rise (Forecast 15.2, Forecast 17.4). The mechanism would fail if robotics and clinical development closed as fast as software, with humanoid shipments above 10 million a year or a large fall in median drug-development time. Where it is capital, its owners take the rent, and with most US equities held by the top tenth of households (17:12), closing the digital asymmetries opens an asymmetry of ownership.

18:16

Forecast 18.2 (The residue is where wages move). By the end of 2030, the ratio of the US median wage of electricians to that of software developers is higher than it was in 2024.

Horizon 2030-12-31

Probability 65%

Check Compare the national median annual wages of electricians and of software developers in the latest release of the [BLS Occupational Employment and Wage Statistics](#) available at the horizon with the May 2024 estimates. The forecast holds if the ratio of the electricians' median to the software developers' median is higher; it fails otherwise.

18:17

Forecast 18.3 (The residue has a price index). By the end of 2030, an index of the prices of the residue's inputs, the hard-to-verify, hard-to-trust and physical services, has risen faster than consumer prices since 2026.

Horizon 2030-12-31

Probability 65%

Check Average the BLS producer price indexes for legal services, offices of certified public accountants, electrical contractors for nonresidential building work and non-auto liability insurance, with equal weights and each rebased to its 2026 annual average, and compare the latest month published by the horizon with CPI-U rebased the same way. False if the index has risen less than CPI-U while AI deployment grew: the slow inputs would then be elastic and the residue would carry no rent.

18.3 The scale ladder

18:18 Measured from one brain to one star, the AI build-out is large against us and small against the light we receive, and its exponentials reach physical ceilings within decades. On 14 February 1990, Voyager 1, about 6 billion kilometers from the Sun, imaged Earth as a point of light about 0.12 of a pixel across, 34 minutes before the spacecraft powered off its cameras for good ([NASA](#)). Carl Sagan built his 1994 book *Pale Blue Dot* around the image: "Look again at that dot. That's here. That's home. That's us." Every saint and sinner in the history of the species, he wrote, "lived there—on a mote of dust suspended in a sunbeam" ([The Planetary Society](#)).

18:19

Definition 18.1 (Kardashev–Sagan index). Kardashev classified civilizations by the power they command: Type I "close to the level presently attained on the earth", about 4×10^{12} W; Type II "capable of harnessing the energy radiated by its own star"; Type III, a galaxy's ([Kardashev 1964](#)). Sagan proposed decimal types, Type 1.0 at 10^{16} W and a factor of ten for each tenth, and placed his own civilization near Type 0.7 ([Ćirković](#)). The **Kardashev–Sagan index** of a power P is the continuous interpolation

$$\mathcal{K}(P) = \frac{\log_{10}(P/1 \text{ W}) - 6}{10}.$$

No equation appears in Sagan's book; the earliest that Gray found is in a 2004 Wikipedia article ([Gray 2020](#)).

18:20 On that ruler the book's numbers line up from one brain to one star, with three readings. First, if the forecast of more than 200 GW of world AI compute by the end of 2028 holds, AI accelerators alone, about 2×10^{11} W, will draw more power than the metabolism of every human brain combined, about 1.7×10^{11} W (Forecast 10.1). The comparison is between watts, and what it fixes is the order of magnitude of the mass of intelligence (§10.5). Second, that fleet is about a millionth of the sunlight Earth intercepts, and all of civilization's energy supply is about a ten-thousandth of it, at $\mathcal{K} \approx 0.73$, close to Sagan's 0.7. Third, about 33 doublings separate one brain from the 2028 fleet, and about 20 more separate the fleet from all the sunlight that reaches the planet, which is itself about one part in two billion of what the Sun emits.

▼ *The ladder, rung by rung*

Rung	Power (W)	$\mathcal{K}$	Share of the sunlight on Earth	Basis
One human brain	2×10^1	-0.47	1×10^{-16}	about 20 W (§6.1)
One GB300-class accelerator, all-in	2.6×10^3	-0.26	1.5×10^{-14}	§10.2
AI compute added worldwide in 2026	3×10^{10}	0.45	2×10^{-7}	about 30 GW (§10.1)
All data centers, 2025 average	$5.5\text{--}9.0 \times 10^{10}$	0.47–0.50	$3\text{--}5 \times 10^{-7}$	485 TWh (IEA) to 788 TWh (Energy Institute)
Every human brain	1.7×10^{11}	0.52	1×10^{-6}	about 165–170 GW (§10.5)
World AI compute, end of 2028	2×10^{11}	0.53	1×10^{-6}	more than 200 GW (§10.1)
World electricity, 2025	3.6×10^{12}	0.66	2×10^{-5}	31,779 TWh (Ember)
Kardashev's Type I	4×10^{12}	0.66	2×10^{-5}	Kardashev 1964
World energy supply, 2025	1.9×10^{13}	0.73	1×10^{-4}	more than 600 EJ (Carbon Brief)
Sunlight intercepted by Earth	1.74×10^{17}	1.12	1	1,361 W/m ² over a disk of radius 6,371 km (NASA)
The Sun	3.83×10^{26}	2.06	2×10^9	nominal solar luminosity (IAU 2015)

18:21

Proposition 18.2 (Exponentials meet ceilings). A power draw growing from P_0 at a constant annual rate g_P (local) reaches a ceiling P_c after

$$t_c = \frac{\ln(P_c/P_0)}{\ln(1 + g_P)} \text{ years.}$$

At the build-out rates of §10.1, AI's draw passes that of all human brains in 2028 and stays about a millionth of the sunlight Earth intercepts, about 1.7×10^{17} W.

18:22 From the end-2028 fleet, AI compute growing 50% a year would equal today's world electricity after about 7 years and all the sunlight Earth intercepts after about 34; at 30% a year it would take about 11 and 52 years. Kardashev ran the same arithmetic in 1964 and found that at 1% growth "the energy consumption per second will be equal to the output of the sun per second, 3200 years from now". AI compute will not grow 50% a year for 34 years, and that is the point. Every exponential the book measures, in tokens, gigawatts, capital spending and software efficiency, is the rising branch of a curve with a physical ceiling. The thermodynamic ceiling lies far above the ones that bind first, which are power equipment, packaging, memory and financing (§10.6, §16.2, §15.3), but at the build-out's rates it is within a human lifetime.

18:23 The memo announcing SpaceX's merger with xAI read the same arithmetic as a reason to leave the planet: "space-based AI is obviously the only way to scale" ([Business Insider](#)). Either way, physics sets the exit: under a compute cap the research loop grows no faster than the build-out (Proposition 11.2), and these ceilings set where it flattens.

18.4 Three humilities

18:24 Humility has three working forms: calibration, which can be scored; deference, which an agent can compute; and perspective, which the scale ladder supplies.

18:25

Definition 18.2 (Calibration). A forecaster assigns probabilities p_i to claims $i = 1, \dots, n$ whose outcomes are $o_i \in \{0, 1\}$ (local). **Calibration** is the property that, among the claims given probability p , a share p come true. The Brier score scores it,

$$\text{BS} = \frac{1}{n} \sum_{i=1}^n (p_i - o_i)^2,$$

and it is a proper scoring rule: its expected value is lowest when the forecaster reports the probability it believes ([Brier 1950](#)).

18:26 Calibration is confidence stated in advance so that it can be checked afterward, the verification asymmetry turned on one's own beliefs: a good forecast is expensive to make and cheap to score once its date arrives (Definition 2.1). This book's test is its forecasts, each with a claim, a date, a probability and a check, collected at the end of the book. A forecast that something will be observed fails if no evidence is published by its date, and one that something will not happen holds by default. The forecasts are not independent: the nine that bet on a rent for verification fail together if automated verifiers reach expert agreement, so the expected 34 hits of 67 could land anywhere from about 27 to 42 if the forecasts were independent, and further out given such clusters. Forecasters of AI already score themselves in public, as the authors of AI 2027 did in August 2026 (11:71).

18:27 I expect physics to prove more forecastable than prices. Physical build-outs should track their forecasts more closely than valuations and labor outcomes do, because a price embeds everyone else's forecast, including the residual that keeps closers paid (Proposition 16.3), while a turbine or a fab does not.

18:28

Forecast 18.4 (Physics is more forecastable than prices). By the end of 2028, forecasts made in 2026 of the gigawatts of AI data centers for the end of 2027 and of 2028 have proved more accurate, in proportional error, than forecasts made in 2026 of AI capital spending and of the leading labs' valuations for the same dates.

Horizon 2029-06-30

Probability 50%

Check Take as forecasts the projections published in 2026 for the end of 2027 and of 2028: world AI data-center capacity from SemiAnalysis and Epoch AI; the big four's combined capital spending from bank and analyst consensus; and OpenAI's and Anthropic's valuations from bank and analyst projections, or, failing those, their last priced valuations of 2026. Compute each group's median absolute log error against the outcomes reported by the horizon. False if either the capital-spending or the valuation group has the smaller error.

18:29 The second humility belongs to agents. In the off-switch game of Hadfield-Menell, Dragan, Abbeel and Russell, a robot can act, switch itself off, or defer to a human, who lets the action proceed only if its utility U to the human is positive ([Hadfield-Menell et al.](#)).

18:30

Proposition 18.3 (Humility makes deference valuable). If the robot's belief about U is normal with mean m_U and spread s_U (locals), deferring is worth $\mathbb{E}[\max(U, 0)]$, against $\max(m_U, 0)$ for acting or stopping alone, and the gain from deferring is

$$\mathbb{E}[\max(U, 0)] - \max(m_U, 0) = s_U L_N\left(\frac{|m_U|}{s_U}\right) > 0, \quad L_N(z) = \frac{e^{-z^2/2}}{\sqrt{2\pi}} - z \Phi(-z),$$

with L_N the standard normal loss function. The gain rises with the robot's uncertainty and vanishes as $s_U \rightarrow 0$.

▼ *Proof*

For a normal U , $\mathbb{E}[\max(U, 0)] = m_U \Phi(m_U/s_U) + s_U e^{-m_U^2/(2s_U^2)}/\sqrt{2\pi}$. Subtracting $\max(m_U, 0)$ and using $\Phi(-z) = 1 - \Phi(z)$ gives $s_U L_N(|m_U|/s_U)$ for either sign of m_U . Since $L'_N(z) = -\Phi(-z)$, the derivative of the gain in s_U is $e^{-m_U^2/(2s_U^2)}/\sqrt{2\pi} > 0$. ■

18:31 The authors show that a traditional agent, which "takes its reward function for granted", has "an incentive to disable the off switch, except in the special case where H is perfectly rational", and conclude that "giving machines an appropriate level of uncertainty about their objectives leads to safer designs." Corrigibility and calibration are one virtue at two levels: the agent's humility about its goals and the forecaster's humility about the world. Neither can come from a self-check, since a model that endorses itself passes at every fixed point, aligned or not (Proposition 13.3); both need a check from outside, a principal to defer to or outcomes to be scored against. Anthropic's constitution for Claude uses the same word for the builders' stance, approaching the project "with the humility that it demands" ([Anthropic](#)).

18:32

Forecast 18.5 (Calibration becomes a priced asset). By the end of 2029, at least one major AI lab or regulatory authority makes a scored forecasting record or a measured rate of deference an explicit condition for deploying a model.

Horizon 2029-12-31

Probability 50%

Check Search published safety frameworks, system cards and regulatory rules. True if one makes either input an explicit condition for deployment and names the measured quantity, such as the Brier score of the lab's own capability forecasts or the rate at which a model complies with shutdown and oversight instructions in evaluations, whether or not it publishes the threshold; false if none does.

18:33 The third humility is Sagan's: the frame in which "thousands of confident religions, ideologies, and economic doctrines" shrink to a pixel ([The Planetary Society](#)). Norbert Wiener, who named the helmsman's science, wrote the cybernetic version in 1950: "we are shipwrecked passengers on a doomed planet," yet "human decencies and human values do not necessarily all vanish, and we must make the most of them" ([Wiener](#)). The frame cuts both

ways, because the same cosmos feeds grandiosity: Sagan's own count of "some 500 trillion people yet to come" ([Sagan 1983](#)), the SpaceX memo's "Kardashev II-level civilization", and Musk's aside while reading the pale-dot passage aloud on Lex Fridman's podcast in November 2019, "This is not true (laughs). This is possible, Mars" ([Lex Fridman #49](#)). Thomas Nagel's register fits the facts better than awe: to meet our situation "with irony instead of heroism or despair" ([Nagel](#)).

18:34 Read as feedback, awe and humility belong together. A runaway is the positive loop of Proposition 0.1, and humility is the negative one, the correction that keeps forecasts calibrated and agents corrigible. Runaway names what happens when the gain outruns the loop's capacity to correct, and Proposition 4.3 makes that exact: no delayed brake holds a loop that grows by a factor of e or more within the brake's lag. A book that forecasts automated researchers and multi-trillion-dollar valuations earns its humility only by marking those claims as forecasts and scoring them later.

18.5 The future seems so good

18:35 The optimism holds on four conditions, and the master equation names them. The closers are aligned, with α_k high enough to license autonomy without the failures of §13.4 and the record of §9.1. The brakes are respected, since inertia and the security dual set φ_k and bound how fast λ_k can safely rise (§15.2, §12.1). The ownership asymmetry closes along with the others, or the closing gaps between producing and checking, frontier and sufficient, digital and physical open a widening gap between owning and not owning (§17.3). And we stay calibrated about time, so that "one or two years" never becomes a promise people plan their lives around (17:3).

18:36 The first condition sits in the slowest part of the residue. Values settle near a good fixed point only if the update pulls hard and little error gets through unverified (Proposition 13.1), and a faster loop can change the strength of the pull but not its sign (Proposition 13.2). Deference is valuable to an agent unsure of its objective (Proposition 18.3), calibration makes humility checkable (Definition 18.2), and a regulator slower than what it regulates needs both (Proposition 4.3). "If we fix alignment" (§13.8) is a conditional whose antecedent lies in the residue, because alignment cannot be produced faster than it is verified.

18:37 What, then, is the closing of the asymmetries? It is the descent of value gradients by machines built for it, from markets, swarms and learners to a general-purpose intelligence (§0.3). It never completes, because closers must be paid and a residual gap remains (Proposition 16.3), and because regeneration reopens the fastest gaps (Proposition 18.1). Its order is set by verification, alignment and inertia ((18.1)). What stays open is the work hardest to verify, trust or build, and the rent follows it. The last asymmetry is the ownership of the slow inputs (17:12).

18:38 Most of the asymmetries the book studies close the same way, through a price, at the speed of their build time. Data, compute and latency are closing now; memory, power and hands will close over years. Two compound: the root capability, which improves itself and reaches every branch (Proposition 12.1), and the fixed seed, which decides what the improvement is for (§13.2). A singularity is what a positive loop looks like before it meets the world, and the world is made of matter.

18:39 This book is itself an instance of the first asymmetry, between producing and checking. Writing it cost far more than checking any paragraph of it, so every number carries its source and every forecast its check, and the

reader closes the book's last asymmetry by scoring the forecasts as their dates arrive. We should close these asymmetries. I think that the future seems so good.

References

1. BLS Occupational Employment and Wage Statistics. <https://www.bls.gov/oes/>
2. NASA. <https://science.nasa.gov/mission/voyager/voyager-1s-pale-blue-dot/>
3. The Planetary Society. <https://www.planetary.org/worlds/pale-blue-dot>
4. Kardashev 1964. <https://articles.adsabs.harvard.edu/pdf/1964SvA%2E%2E%2E..8..217K>
5. Milan M. Cirkovic (2016). Kardashev's Classification at 50+: A Fine Vehicle with Room for Improvement. *arXiv*. <https://arxiv.org/abs/1601.05112>
6. Robert H. Gray (2020). The Extended Kardashev Scale. *The Astronomical Journal*. <https://iopscience.iop.org/article/10.3847/1538-3881/ab792b>
7. IEA. <https://www.iea.org/reports/key-questions-on-energy-and-ai/executive-summary>
8. Energy Institute. <https://www.energyinst.org/exploring-energy/resources/news-centre/media-releases/global-electrification-reaches-tipping-point-as-energy-demand-hits-record-highs-and-regional-paths-diverge>
9. Ember. <https://ember-energy.org/latest-insights/global-electricity-review-2026/>
10. Carbon Brief. <https://www.carbonbrief.org/six-charts-show-how-clean-power-was-worlds-largest-source-of-new-energy-in-2025>
11. NASA. <https://nssdc.gsfc.nasa.gov/planetary/factsheet/earthfact.html>
12. A. Prša, P. Harmanec, Guillermo Torres, Eric E. Mamajek, M. Asplund et al. (2016). NOMINAL VALUES FOR SELECTED SOLAR AND PLANETARY QUANTITIES: IAU 2015 RESOLUTION B3 * †. *The Astronomical Journal*. <https://iopscience.iop.org/article/10.3847/0004-6256/152/2/41>
13. Business Insider. <https://www.businessinsider.com/spacex-acquiring-xai-deal-elon-musk-2026-2>
14. Brier 1950. https://journals.ametsoc.org/view/journals/mwre/78/1/1520-0493_1950_078_0001_vofeit_2_0_co_2.xml
15. Dylan Hadfield-Menell, Anca Dragan, Pieter Abbeel, Stuart Russell (2016). The Off-Switch Game. *arXiv*. <https://arxiv.org/abs/1611.08219>
16. constitution. <https://www.anthropoc.com/constitution>
17. Wiener. https://monoskop.org/images/9/90/Wiener_Norbert_The_Human_Use_of_Human_Beings_1950.pdf
18. Sagan 1983. <https://www.foreignaffairs.com/articles/1983-12-01/nuclear-war-and-climatic-catastrophe-some-policy-implications>
19. Lex Fridman #49. <https://www.youtube.com/watch?v=smK9dgdTl40>
20. Nagel. <https://philosophy.as.uky.edu/sites/default/files/The%20Absurd%20-%20Thomas%20Nagel.pdf>