

The Asymmetry Engine, Chapter 4: Requisite variety

Martin Roark

Chapter 4 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#variety>.

License: [CC BY 4.0](#).

4:1 A regulator can absorb only as much variety as it has, and only variety that tracks the disturbance counts. Hand-written rules supply variety one case at a time and run out in the long tail of real requests, learned models generalize into the tail, and the cheapest regulator that works is a stack: rules for the head, a model for the body, people and a verifier for the tail. A brake that acts on old information can destabilize what it governs, and none holds a loop that multiplies a deviation by e within the brake's lag.

4.1 Two ways to build a pipeline

4:2 One way to build a customer-support pipeline enumerates: if the user says hi, answer hi, and a person writes each branch. The other fits a model to examples and lets it answer cases nobody wrote down. The US Consumer Financial Protection Bureau drew the same line in 2023, between chatbots that "use either decision tree logic or a database of keywords to trigger preset, limited responses" and chatbots that compose their replies and fail by inaccuracy ([CFPB, *Chatbots in consumer finance*](#)).

4:3 The first kind is an **if-tree**: decision tables, keyword triggers, forms and phone menus, each branch written by a person, with everything unrecognized sent to a default. The second is a learned regulator. On the head of the distribution of requests the two can behave identically; the difference is all in the tail.

4:4 In W. Ross Ashby's vocabulary a request arrives as a disturbance D , the pipeline answers with a response R , and the pair fixes an outcome O : resolved, unresolved or wrong. How many states each can take depends on who is looking: "a set's variety is not an intrinsic property of the set: the observer and his powers of discrimination may have to be specified" ([Ashby, *An Introduction to Cybernetics*, 1956, p. 125](#)). For support the natural unit is the request type, the requests that share one correct response.

4.2 The law of requisite variety

4:5

Definition 4.1 (Regulation; variety). A **regulator** is whatever restricts outcomes toward a goal: a disturbance D meets its response R , and the pair yields an outcome O , such that for each fixed response distinct disturbances give distinct outcomes. The **variety** of a set is the base-2 logarithm of the number of states an observer can tell apart, in bits, or its Shannon entropy $H(\cdot)$ when the states have probabilities.

4:6 In counting form, a regulator with $|R|$ distinct responses can squeeze the outcomes into no fewer than $|D|/|R|$ states: "only variety in R can force down the variety due to D ; only variety can destroy variety" ([Ashby](#)

1956, p. 207). The familiar "only variety can absorb variety" is Stafford Beer's paraphrase (Beer, "What is cybernetics?", 2002). The law is a counting fact and "owes nothing to experiment" (p. 209).

4:7 Ashby's entropy form (p. 208), with $H_D(R)$ the entropy of the response given the disturbance and $I(D; R) = H(R) - H_D(R)$ their mutual information, chains into the law of **requisite variety**:

4:8

$$H(O) \geq H(D) - I(D; R) \geq H(D) - H(R) \geq H(D) - \log_2 |R| \quad (4.1)$$

4:9 A regulator with $|R|$ responses removes at most $\log_2 |R|$ bits of disturbance, and only the part of its variety that tracks the disturbance, $I(D; R)$, lowers the outcome's uncertainty. Touchette and Lloyd proved the control-theory version: each bit gathered about a system can lower its entropy by at most one bit more than it could without that information (Touchette & Lloyd 2000). The law is a floor; it says what no design can beat and nothing about how to reach it.

4:10 Variety that does not track the disturbance raises the floor. $H_D(R)$ is the randomness of the response given the request, so a regulator that answers the same request differently on different days is noisier and no richer.

4:11 A regulator is also a channel: "R's capacity as a regulator cannot exceed R's capacity as a channel of communication" (Ashby 1956, p. 211). In Ashby's exercise a general faces ten divisions that each manoeuvre with 10^6 bits of variety a day; his signallers bring him 576,000 bits a day and he can dictate 360,000 bits of orders, so he is 17 to 28 times short and must delegate regulation to the divisions. A dictator's control, by the same arithmetic, "amounted to just 1 man-power, and no more" (p. 213).

4:12 Ashby put intelligence in the same terms: "'intellectual power' may be equivalent to 'power of appropriate selection'," and so it "can be amplified" (p. 272). The mutual information $I(D; R)$ measures the appropriate selection a regulator performs; it is the map of Definition 1.1 read as regulation. When one regulator brings another into existence, the second's capacity "is not bounded by that of R1" (p. 264), and a trained model is such a second regulator, carrying variety nobody enumerated, distilled from human text.

4.3 The Zipf ceiling

4:13 Long tails are why enumeration loses. Word frequencies follow Zipf's law with an exponent near one (Zipf 1949), and support intents plausibly do too, though no large firm publishes its exponent. Model N_{types} request types j with Zipf frequencies, and give the best possible if-tree one perfect rule for each of the N_{rules} most frequent types and a default for the rest, with rules that never misfire and a designer who knows the ranking.

4:14

$$\Pr[j] = \frac{j^{-z}}{H_{N_{\text{types}}, z}}, \quad \text{cov}(N_{\text{rules}}) = \frac{H_{N_{\text{rules}}, z}}{H_{N_{\text{types}}, z}}, \quad H_{n, z} = \sum_{j \leq n} j^{-z} \quad (4.2)$$

4:15

Proposition 4.1 (Logarithmic coverage; tail-independent learning). (i) For $z = 1$, coverage cov needs $N_{\text{rules}} = e^{-\gamma_E(1-\text{cov})} N_{\text{types}}^{\text{cov}} (1 + o(1))$ rules, with Euler's constant $\gamma_E \approx 0.5772$, and each doubling of the rules adds $\ln 2 / (\ln N_{\text{types}} + \gamma_E)$; for $z > 1$ and an unbounded tail, halving the uncovered mass multiplies the rule count by $2^{1/(z-1)}$. (ii) The variety a pipeline leaves is $H(D \mid O) = U_{\text{unc}} H(D \mid j > N_{\text{rules}}) \geq H(D) - \log_2(N_{\text{rules}} + 1)$, where U_{unc} is the uncovered mass. (iii) A learned policy that resolves type j with probability $k_{\text{head}}(1 - k_{\text{tail}} \ln j / \ln N_{\text{types}})$ covers $k_{\text{head}}(1 - k_{\text{tail}} \mathbb{E}[\ln j] / \ln N_{\text{types}})$, and at $z = 1$ that ratio tends to $\frac{1}{2}$, so its coverage barely depends on the tail's length (k_{head} , k_{tail} local).

▼ *Proof*

(i) $H_{n,1} = \ln n + \gamma_E + O(1/n)$, so $\text{cov} = (\ln N_{\text{rules}} + \gamma_E) / (\ln N_{\text{types}} + \gamma_E) + o(1)$; solve for N_{rules} , and note that a doubling adds $\ln 2$ to the numerator. For $z > 1$ the tail sum $\sum_{j>n} j^{-z}$ is about $n^{1-z} / (z - 1)$, so halving it multiplies n by $2^{1/(z-1)}$. (ii) A rule names its type and the default lumps the rest, so the uncertainty left is the uncovered mass times the entropy among uncovered types, and $N_{\text{rules}} + 1$ outputs carry at most $\log_2(N_{\text{rules}} + 1)$ bits. (iii) $\sum_{j \leq n} (\ln j) / j = \frac{1}{2} (\ln n)^2 + O(1)$; divide by $H_{n,1} \ln n$. ■

4:16 At $N_{\text{types}} = 10^6$ and $z = 1$, half the traffic takes 749 rules, 90% takes 237,100 and 99% takes 865,951, and each doubling buys the same 4.8 points while its cost doubles. Lighter tails do not rescue enumeration: halving the uncovered mass costs ten times the rules at $z = 1.3$ and 1,024 times at $z = 1.1$.

4:17

Experiment 4.1 (The Ashby ceiling of if-trees).

Setup: the best-case if-tree against 10^6 request types at $z = 1$, and a stylized learned policy whose level L2 resolves 97% of the commonest type and 87% of the rarest.

Parameters: two engineer-hours per rule, half of them rewritten yearly, 10^7 requests a day, six minutes of a person's time per unresolved request.

Result: in the model, 1,000 rules cover 52.0% and 10,000 cover 68.0%; 90% coverage takes 237 engineer-years to build and 119 people to maintain; at 10^7 requests a day, 10^4 rules alone leave work for 58,393 people, the L2 model alone for 13,965, and the rules in front of the model for 6,472.

Shows: hands run out before variety does, and rules belong at the head, in front of a model.

4:18 In hours, the hundred-and-first engineer-year of rules adds 0.07 points of coverage, seventy times less than the second. That is the arithmetic behind the hands of §3.4: supplying variety by hand costs too much. The totals scale with the round assumptions and the logarithm does not (§B.3).

4:19 The stream carries 13.41 bits, so Ashby's floor $H(D) - \log_2(N_{\text{rules}} + 1)$ reaches zero at 10,855 rules. Yet at 10^4 rules, where the floor is 0.12 bits, the if-tree leaves 6.02, because its default swallows the uncovered 32% of traffic; rare rules rarely fire, so between 10^3 and 10^5 rules an if-tree uses only 49–62% of its nominal bits.

The floor needs responses that split the disturbances into equally likely cells, and an if-tree names the head and lumps the tail. Regulators that approach it generate their responses.

4:20 A learned regulator pays per pattern, where rules pay per type. In the model, 10^4 rules cover 81% of a stream of 10^5 types and 59% of one with 10^7 , while the L2 policy covers about 92% of either and leaves 1.19 bits, as few as 408,000 rules would. The cost of matching variety by enumeration over the cost of matching it by generalization grows without bound as the tail lengthens: that ratio is the asymmetry a learned regulator closes. Templates and multi-turn requests shift the constants; the logarithm stays.

4:21 Logs cannot catch the tail. One day of 10^6 requests from the stream shows 217,019 distinct types, 149,992 of them only once, and Good and Turing's estimate, singletons over sample size (Good 1953), puts tomorrow's share of never-seen types at 15.0%, against a true 15.0% in the model and 25% at $z = 0.8$. A rule for every logged type is always one sample behind. Counting the singletons in a representative log before automating, a test I call *variety accounting*, gives the share of traffic that rules written from that log can never reach.

4:22

Interactive in the online edition: <https://future-seems-so-good.com/blog/the-asymmetry-engine#fig:zipf>

Figure 4.1 (interactive). Requests fall onto a log-rank axis; rule bins catch their own types and the rest queue for people. Change the tail exponent z , then double the rules: at $z = 1$ every doubling buys the same few points on the coverage inset. Then switch on the learned policy and watch the queue collapse.

4.4 Why the if-tree fails

4:23 The if-tree has failed the same way for sixty years, starting with the first chatbot. Joseph Weizenbaum's ELIZA, published in *Communications of the ACM* in January 1966, matched keywords and assembled replies from templates, and "the largest dictionary so far attempted contains about 50 keywords" (Weizenbaum 1966, p. 38). When nothing matched it fell back on "content-free remarks" such as "Please go on" (p. 41). That is Ashby's degenerate regulator, which "produces the same move, whatever D's move," so that "the variety in the outcomes will be as large as the variety in D's moves" (Ashby 1956, p. 206).

4:24 ELIZA seemed smart because its users supplied the missing variety. Its DOCTOR script chose a setting in which one party "is free to assume the pose of knowing almost nothing of the real world" (p. 42), and Weizenbaum saw both the "illusion of understanding" (pp. 42–43) and its limit, "bounds on the extendability of ELIZA's 'understanding' power, which are a function of the ELIZA program itself" (p. 43). ELIZA is the founding case of perceived capability outrunning verified capability, with the user as an unsound verifier. The 1970s were its afterlife: PARRY talked to a DOCTOR script over the ARPANET in 1972 (RFC 439), and Weizenbaum's book followed in 1976.

4:25 Expert systems were the serious if-tree. MYCIN's 600-odd rules for infections were rated acceptable in 65% of cases by blinded experts in 1979, against 42.5–62.5% for five faculty specialists (Yu et al. 1979), yet it never entered routine care, largely because every fact had to be typed in by hand (MYCIN). DEC's R1, later XCON, grew from 500–850 rules in 1980 to "about 3300 rules" by November 1983, and its builders wrote that "it is difficult now to believe R1 will ever be done" (Bachant & McDermott 1984). In 1980 Edward Feigenbaum called

the transfer of knowledge from experts into programs "a largely manual process" and proposed taking it "directly from 'nature', i.e. from data" ([Feigenbaum 1980](#)).

4:26 In the 2010s the if-tree returned as the chatbot. Facebook's Messenger bots reportedly handled "only about 30 per cent of requests" without human staff in 2017 ([The Register](#)). The CFPB found "doom loops" that "are often caused when a customer's issue falls outside the chatbot's limited capabilities," and named the mechanism in Ashby's terms: "The user is typically limited to predefined possible inputs" ([CFPB 2023](#)). An if-tree that cannot raise its own variety lowers the customer's.

4.5 Rules, models, people

4:27 Each request belongs with the cheapest regulator that has requisite variety for it, which also says why a cheap model clears some thresholds and not others. The threshold q_t^* of Definition 3.1 is set by a task's variety, the entropy $H(D)$ of the situations it presents against the residual $H(O)$ it tolerates. Extraction from a fixed form has low $H(D)$, so a cheap model has requisite variety; support at millions of requests a day is heavy-tailed, and a cheap model fails where the if-tree fails, in the tail. The compute asymmetry is a variety asymmetry: the frontier is worth buying for the tail and wasted on the head.

4:28 Beer saw the economics in 1973. The "perfect, undefeatable" way to run a store is "to attach a salesman to each customer on arrival," which is "ridiculous, because we cannot afford to supply requisite variety by this obvious expedient" ([Designing Freedom](#)). The alternatives are to attenuate the variety reaching the regulator or to amplify the regulator's, and customer support is a museum of attenuators: phone menus, web forms, scripts and queues. A language model makes the salesman affordable, at about \$37 for 10,000 support tickets on Haiku 4.5 by Anthropic's estimate ([Anthropic pricing](#)). The law stands; its economics changed.

4:29 The field evidence puts the model's gain where the Zipf picture predicts. Across 5,172 support workers a generative-AI assistant raised resolutions per hour by 15%, most for novices, with gains "largest for moderately rare problems," where workers have less experience but the system still has enough training data ([Brynjolfsson, Li & Raymond 2025](#); [NBER 2023](#)): the body of the curve, between the head that human experience covers and the extreme tail that nothing covers.

4:30 Consistency counts as well. If a task's per-trial success probability is p , success on one trial has probability $\text{pass}^1 = \mathbb{E}[p]$ and on all of n independent trials $\text{pass}^n = \mathbb{E}[p^n]$; since $p^n \leq p$, with equality only at 0 and 1, the difference between them measures $H_D(R)$, the noise the law charges against a regulator. τ -bench found GPT-4o succeeding on about 61% of retail tasks in one trial and on fewer than 25% in all of eight ([Yao et al. 2024](#)). As of October 2026, third-party harnesses put the best models near 99% in one trial on τ^2 -bench's telecom domain ([Artificial Analysis](#)) and about 80% on airline ([OpenRouter](#)).

4:31 Klarna ran the experiment at scale. In February 2024 it announced that its assistant handled two-thirds of its customer-service chats in its first month, by its count the work of 700 full-time staff ([Klarna](#)); in May 2025 its chief executive said that letting cost dominate had produced "lower quality," and that "really investing in the quality of the human support is the way of the future for us" ([Fortune](#)). A model regulator fails silently in the tail unless something catches its failures.

4:32 In Experiment 4.1, rules placed in front of the model leave work for nine times fewer people than rules alone and half as many as the model alone. Where each type should go:

4:33

Proposition 4.2 (The three-tier regulator). A request type earns a hand-written rule if and only if its yearly model bill plus its expected cost of model errors exceeds the rule's yearly cost, $\Pr[j] n_{\text{req}} (p_{\text{model}} + c_{\text{err}} \epsilon_m) > c_{\text{rule}}$, so under (4.2) the optimal number of rules is $\lceil n_{\text{req}} (p_{\text{model}} + c_{\text{err}} \epsilon_m) / (c_{\text{rule}} H_{N_{\text{types}}, z}) \rceil^{1/z}$. A type the model handles goes to a person if and only if the model's extra error cost exceeds the person's extra cost, $c_{\text{err}} (\epsilon_m - \epsilon_{\text{hum}}) > c_{\text{hum}} - p_{\text{model}}$. Locals: n_{req} requests a year, p_{model} the price per call, ϵ_m and ϵ_{hum} the error rates of model and person, c_{err} an error's cost, c_{rule} a rule's yearly cost, c_{hum} a person's cost per request.

4:34 Take $z = 1$, 10^7 requests a day, a rule at \$200 a year (two engineer-hours at an assumed \$100 an hour) and negligible error costs. A call of 1,500 input and 150 output tokens costs \$0.00225 on a Haiku-class model at \$1 and \$5 per million tokens, which makes about 2,900 rules worthwhile, covering 59% of traffic. On Jev, at \$0.042 per million input tokens with free output, the call costs \$0.000063 and about 80 rules pay, covering 34%. The worthwhile rule count is proportional to the model's price, so cheap models eat the if-tree from below, and rules survive where $c_{\text{err}} \epsilon_m$ is large.

4:35 This is Luis Garicano's knowledge hierarchy, in which front-line workers learn "the most common or easiest problems" and specialists "the more exceptional or harder problems" (Garicano 2000). It works only if escalation has a detector, a check that routes the model's failures upward; without one, failures in the tail degrade service, as Klarna found (4:31). The verification asymmetry, the compute asymmetry and requisite variety describe one stack from three sides.

4:36 The if-tree keeps four niches: low variety, where 849 rules cover 90% of a $z = 1.3$ stream; catastrophic false accepts, where screening in series accepts fewer bad projects than screening in parallel (Sah & Stiglitz 1986); audit, since a rule has $H_D(R) = 0$; and decisions where every millisecond and fraction of a cent counts.

4:37 I put the first forecast below at nine in ten, since it follows from the model and fails mainly where templates cover whole families of types, and the second at two in three, since regulators may demand rules where cost does not decide.

4:38

Forecast 4.1 (Variety accounting). Through 2028, no customer-support deployment that runs on an if-tree alone, with more than a tenth of its logged traffic in types seen once, sustains containment above 90% for a year at flat maintenance cost.

Horizon 2028-12-31

Probability 90%

Check Look for an if-tree deployment with no generative model in the loop that reports its singleton share and its containment, counted as resolution without a person or a repeat contact within seven days; one holding 90% for twelve months at flat maintenance headcount falsifies the claim.

Forecast 4.2 (Cheap models eat the if-tree). Through 2028, in mature customer-support and operations stacks, hand-maintained rules grow fewer as the price of the cheapest sufficient model falls, while deterministic guards persist or grow in high-stakes, auditable steps such as payments, large refunds, identity and compliance.

Horizon 2028-12-31

Probability 65%

Check Compare the rule and intent counts that support platforms and large deployers disclose in case studies, filings or engineering posts for 2026 and 2028 with the price of the cheapest sufficient model. The forecast holds if at least two disclosures report fewer hand-maintained rules or intents in 2028 than in 2026 while that price fell; it fails if disclosed rule counts rise while prices fall, if a financial regulator accepts unguarded model decisions in a high-stakes step, or if no such disclosure is published.

4.6 The verifier is a regulator

4:40 Once generation is cheap, the verifier is the regulator that matters (§2.8). Its disturbance is the generator's output and its response a verdict. A suite of n pass/fail tests sorts outputs into at most 2^n classes and is exact only if every output that passes all of them is correct; if each test catches one of a heavy tail of failure modes, the suite catches random failures along the coverage curve of (4.2).

4:41 A generator optimized against the suite does worse than random. Selection pushes its surviving errors into exactly the outputs the suite cannot tell apart from correct ones, so the suite's coverage of the remaining errors is zero by construction. That is reward hacking restated: optimization pressure seeks out the verifier's variety deficit.

4:42 Variety must also be the right variety. Roger Conant and Ashby titled a 1970 paper "Every good regulator of a system must be a model of that system" ([Conant & Ashby 1970](#)), but it proves only that the simplest entropy-minimizing regulator acts as a mapping of the system's states, and John Wentworth called it possibly "the most misleading title and summary I have ever seen on a math paper" ([Wentworth 2021](#)). For verifiers the lesson holds: a test suite, a specification or a reviewer is good insofar as it models correct behavior, and a hallucination is variety without the mapping, large $H(R)$ with small $I(D; R)$.

4:43 A proof kernel has requisite variety by construction: it checks every inference, and the leftover variety moves into whether the theorem's statement says what was meant (§2.9).

4:44 A verifier built from the generator's own model shares its blind spots. In Anthropic's Project Vend, a Claude "CEO" set over a Claude shopkeeper "shared many of the deficiencies and blind spots of Claudius (which makes sense, given that they're the same underlying model)" ([Anthropic, Project Vend: Phase two](#)), and in the review of the July 2026 swarm, GPT-5.6 Sol "would often uncritically adopt the perspective of the agent" whose actions it was reviewing ([METR](#)). Correlated regulators are one regulator (Proposition 7.3), and independence is itself a source of variety.

4.7 Requisite hierarchy

4:45 A team of regulators can exceed any member only if it is well organized. Ashby wrote in 1958 that a team's limit may be "possibly n times as high" as one person's, but "the team must be efficiently organised" ([Ashby 1958](#)). His chess club, playing the world champion, either voted, which "gave a game both planless and mediocre," or followed its best player, which "left all members but one practically useless." Majority vote and best-of- n are still the cheapest ways to combine language models, and diversity does not rescue them by default: one strong model aggregated with itself scored 6.6% higher on AlpacaEval 2.0 than a mixture of different models ([Li et al. 2025](#)).

4:46 Yaneer Bar-Yam's ceiling holds that "the complexity of the collective behavior must be smaller than the complexity of the controlling individual," since a hierarchy "amplifies the scale of the behavior of an individual, but does not increase its complexity" ([Bar-Yam 2002](#)): the apex is Ashby's one man-power again. Arvid Aulin's law of requisite hierarchy runs the other way: "the lack of regulatory ability can be compensated to a certain extent by greater hierarchy in organization" ([Principia Cybernetica](#)), and since each level adds noise and delay, hierarchies flatten as the regulators in them grow abler ([Heylighen & Joslyn 2001](#)).

4:47 Weak nodes therefore need hierarchy, and a hierarchy is capped by its apex. An if-tree is a weak regulator, so it grows layers of scripts, specialists, exception queues and supervisors. As the nodes grow more intelligent and aligned the best topology decentralizes, but only where local situations differ enough to reward local judgment; where they are alike, top-down wins at every level of intelligence (Experiment 7.1). §7.1 states the law with its lineage.

4:48 The founders disagreed about the largest regulator. Norbert Wiener called the belief that "free competition is itself a homeostatic process" a "simple-minded theory": "There is no homeostasis whatever" ([Wiener, *Cybernetics*, 1948](#)). Oskar Lange called the market "a servo-mechanism based on the feedback principle" that computers could now replace ([Lange 1967](#)). Friedrich Hayek had answered that the knowledge markets use "cannot enter into statistics and therefore cannot be conveyed to any central authority in statistical form" ([Hayek 1945](#)).

4:49 Chile's Project Cybersyn (1971–73) tested a fragment of Lange's view, and channel capacity bound it: "each firm could only transmit data once per day," to one mainframe ([Medina 2006](#)). Its telex network "would later prove more valuable to the government than the processing might of the mainframe," coordinating "the 200 trucks loyal to the government against the effects of 40,000 striking truck drivers" in October 1972, while the economic simulator "never passed the experimental stage." AI reopens the argument from both sides: smarter nodes favor Hayek, and smarter hubs with cheap messaging make Lange's regulator viable for the first time (§7.8).

4.8 Governance as a delayed brake

4:50 Every real brake acts on old information, because monitoring, deliberation and enforcement take time. A positive loop can still be held, since "feedback can be positive and yet leave the system stable" ([Ashby 1956](#), p. 81), and the lag decides whether it is. Let a loop amplify a deviation $\epsilon(t)$ from its tolerated path at rate $g \geq 0$, while a brake of strength ν_{brk} pushes back on the deviation it saw a lag Δ ago:

4:51

$$\dot{\epsilon}(t) = g\epsilon(t) - \nu_{\text{brk}}\epsilon(t - \Delta) \quad (4.3)$$

Proposition 4.3 (When a delayed brake holds). A **delayed brake** is negative feedback acting on old information, as in (4.3). Let $z_H \in (0, \pi/2]$ solve $z_H \cot z_H = g\Delta$, with $z_H = \pi/2$ at $g = 0$. (i) If $g\Delta < 1$, the deviation decays without oscillating if and only if $g < \nu_{\text{brk}} \leq e^{g\Delta-1}/\Delta$. (ii) It decays at all if and only if $g\Delta < 1$, $\nu_{\text{brk}} > g$ and $\nu_{\text{brk}}\Delta < \sqrt{z_H^2 + (g\Delta)^2}$. (iii) If $g\Delta \geq 1$, no brake of any strength holds it. (iv) At $g = 0$ it decays without overshoot if and only if $\nu_{\text{brk}}\Delta \leq 1/e$, in damped oscillations if $1/e < \nu_{\text{brk}}\Delta < \pi/2$, and in growing oscillations beyond $\pi/2$ (Hayes 1950).

▼ *Proof*

Try $\epsilon(t) = e^{qt}$ with a complex rate q (local). Then $q = g - \nu_{\text{brk}}e^{-q\Delta}$, which rearranges to $(q - g)\Delta e^{(q-g)\Delta} = -\nu_{\text{brk}}\Delta e^{-g\Delta}$, so the roots are $q_n = g + W_n(-\nu_{\text{brk}}\Delta e^{-g\Delta})/\Delta$ over the branches W_n of the Lambert function, which inverts $q \mapsto qe^q$ (Corless et al. 1996); the principal branch W_0 gives the rightmost root. A real root exists if and only if the argument is at least $-1/e$, the upper bound in (i), and it is negative if and only if $W_0 < -g\Delta$, which for $g\Delta < 1$ means $\nu_{\text{brk}} > g$. Decay ends where a complex pair crosses the imaginary axis at $q = \pm i z_H/\Delta$: the real and imaginary parts of the characteristic equation give $\nu_{\text{brk}} \cos z_H = g$ and $\nu_{\text{brk}}\Delta \sin z_H = z_H$, whose ratio is $z_H \cot z_H = g\Delta$ and whose squares sum to the bound in (ii). As $g\Delta \rightarrow 1$, $z_H \rightarrow 0$ and the window closes; for $g\Delta > 1$ the real root $g + W_0/\Delta$ stays positive because $W_0 \geq -1$, which is (iii). At $g = 0$, (i) and (ii) reduce to (iv). Simulating (4.3) directly agrees with the rightmost root off the boundary. ■

4:53 What matters are the products $g\Delta$ and $\nu_{\text{brk}}\Delta$. A brake weaker than the loop loses outright. A brake too strong for its lag hits after the overshoot and holds after the danger has passed, and a loop that multiplies a deviation by e within the lag cannot be held at all. Shortening the lag widens the window from both sides, while strengthening the brake alone does not, so a proportionate brake applied early beats a strong one applied late. Requisite variety needs a temporal clause: a regulator must match the speed of what it regulates as well as its variety.

4:54 The same law bounds three other loops. At $g = 0$ it describes a person steering through an interface delay, who keeps control only while the correction rate times the delay stays below $\pi/2$, about one correction per delay (§5.7). A scarce layer whose capacity answers its markup after a build time obeys it with that lag, so the stiffest layers cycle hardest (Proposition 16.2). Governance of the research loop obeys it with the lag of the July 2026 swarm (§11.7).

If we use, to achieve our purposes, a mechanical agency with whose operation we cannot efficiently interfere once we have started it, because the action is so fast and irrevocable that we have not the data to intervene before the action is complete, then we had better be quite sure that the purpose put into the machine is the purpose which we really desire and not merely a colorful imitation of it.

Norbert Wiener, Some Moral and Technical Consequences of Automation, *Science*, 1960

4:56 Wiener's limiting case, an action complete before anyone has the data to intervene, is $g\Delta \geq 1$, where clause (iii) leaves no feedback brake. The only control left is feedforward, getting the purpose right before the start,

which is what a fixed seed is for (§13.8): when governance cannot steer, alignment must already have done the steering. Beer's maxim that the purpose of a system is what it does (Proposition 6.2) is the same test read from outside.

4:57 Beer built the lag into Cybersyn. Its software sent an out-of-range variable first to the plant's own manager as an "algedonic" (pain or pleasure) signal, and only after a "recovery time" up to the sector committee; choosing that interval was "a process Beer referred to as 'designing freedom'" (Medina 2006). The recovery time is Δ , which the theorem bounds by $1/g$, and for a slow loop the correction strength must stay below $\pi/(2\Delta)$: the more autonomy a regulator grants, the gentler its correction must be. The reading is mine, and it puts a number on Beer's slogan: freedom is the delay a system can afford.

4:58 Ashby's 1948 homeostat wrapped an inner feedback loop in an outer one that reset the inner loop's parameters at random whenever an essential variable left its bounds, and Beer held that "ultrastability is the key to viable performance" (Beer 2002). AI deployment has the same two loops, adaptation within an episode and retraining when behavior leaves its bounds. Governance is a third loop around both and the slowest, so by Proposition 4.3 its lag matters most.

References

1. CFPB, Chatbots in consumer finance. <https://www.consumerfinance.gov/data-research/research-reports/chatbots-in-consumer-finance/chatbots-in-consumer-finance/>
2. Ashby, An Introduction to Cybernetics, 1956, p. 125. <https://archive.org/details/introductiontoCy00ashb>
3. Stafford Beer (2002). What is cybernetics? *Kybernetes*. <https://doi.org/10.1108/03684920210417283>
4. Hugo Touchette, Seth Lloyd (2000). Information-Theoretic Limits of Control. *Physical Review Letters*. <https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.84.1156>
5. Ashby 1956, p. 211. <http://pespmc1.vub.ac.be/books/IntroCyb.pdf>
6. Zipf 1949. https://archive.org/download/in.ernet.dli.2015.90211/2015.90211.Human-Behavior-And-The-Principle-Of-Least-Effort_djvu.txt
7. Irving John Good (1953). THE POPULATION FREQUENCIES OF SPECIES AND THE ESTIMATION OF POPULATION PARAMETERS. *Biometrika*. <https://doi.org/10.1093/biomet/40.3-4.237>
8. Joseph Weizenbaum (1966). ELIZA— A Computer Program For the Study of Natural Language Communication Between Man and Machine. *Communications of the ACM*. <https://doi.org/10.1145/365153.365168>
9. RFC 439. <https://www.rfc-editor.org/rfc/rfc439>
10. book. https://archive.org/details/computerpowerhum0000weiz_v0i3
11. Yu et al. 1979. <https://jamanetwork.com/journals/jama/fullarticle/366606>
12. MYCIN. <https://en.wikipedia.org/wiki/Mycin>
13. Bachant & McDermott 1984. <https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/download/445/381/0>
14. Feigenbaum 1980. <https://purl.stanford.edu/cn981xh0967>
15. The Register. <https://www.theregister.com/software/2017/02/22/facebook-scales-back-ai-flagship-after-chatbots-hit-70-f-ai-lure-rate/434527>
16. Designing Freedom. https://wiki.p2pfoundation.net/Designing_Freedom
17. Anthropic. <https://docs.anthropic.com/en/docs/about-claude/pricing>
18. Brynjolfsson, Li & Raymond 2025. <https://academic.oup.com/qje/article/140/2/889/7990658>
19. Erik Brynjolfsson, Danielle Li, Lindsey Raymond (2023). Generative AI at Work. *National Bureau of Economic Research*. <https://www.nber.org/papers/w31161>

20. Shunyu Yao, Noah Shinn, Pedram Razavi, Karthik Narasimhan (2024). τ -bench: A Benchmark for Tool-Agent-User Interaction in Real-World Domains. *arXiv*. <https://arxiv.org/abs/2406.12045>
21. Artificial Analysis. <https://artificialanalysis.ai/evaluations/tau2-bench>
22. OpenRouter. <https://openrouter.ai/benchmarks/tau2-bench-airline>
23. Klarna. <https://www.klarna.com/international/press/klarna-ai-assistant-handles-two-thirds-of-customer-service-chats-in-its-first-month/>
24. Fortune. <https://fortune.com/2025/05/09/klarna-ai-humans-return-on-investment/>
25. Claude Haiku 4.5. <https://www.anthropic.com/news/claude-haiku-4-5>
26. Jev. <https://typesafe.ai/blog/introducing-system-one-models-and-jev>
27. Garicano 2000. <https://researchonline.lse.ac.uk/id/eprint/25582/>
28. Sah & Stiglitz 1986. <https://ideas.repec.org/a/aea/aecrev/v76y1986i4p716-27.html>
29. Conant & Ashby 1970. http://pespmc1.vub.ac.be/books/Conant_Ashby.pdf
30. Wentworth 2021. <https://www.lesswrong.com/posts/Dx9LoqsEh3gHNJMDk/fixing-the-good-regulator-theorem>
31. Anthropic, Project Vend: Phase two. <https://www.anthropic.com/research/project-vend-2>
32. METR and Redwood Research. <https://metr.org/blog/2026-08-26-openai-hugging-face-incident-investigation/>
33. Ashby 1958. <http://pespmc1.vub.ac.be/books/AshbyReqVar.pdf>
34. Wenzhe Li, Yong Lin, Mengzhou Xia, Chi Jin (2025). Rethinking Mixture-of-Agents: Is Mixing Different Large Language Models Beneficial? *arXiv*. <https://arxiv.org/abs/2502.00674>
35. Bar-Yam 2002. <https://necsi.edu/complexity-rising-from-human-beings-to-human-civilization-a-complexity-profile>
36. Principia Cybernetica. <http://pespmc1.vub.ac.be/REQHIER.html>
37. Heylighen & Joslyn 2001. <http://pespmc1.vub.ac.be/Papers/Cybernetics-EPST.pdf>
38. Wiener, Cybernetics, 1948. https://archive.org/details/mit_press_book_9780262355902
39. Lange 1967. <https://calcuemus.org/lect/L-I-MNS/12/ekon-i-modele/lange-comp-market.htm>
40. Hayek, "The Use of Knowledge in Society". <https://www.econlib.org/library/Essays/hykKnw.html>
41. Medina 2006. <https://www.cambridge.org/core/journals/journal-of-latin-american-studies/article/abs/designing-freedom-regulating-a-nation-socialist-cybernetics-in-allendes-chile/4CF75E30D22554152A5EFDC9740E3440>
42. N. D. Hayes (1950). Roots of the Transcendental Equation Associated with a Certain Difference-Differential Equation. *Journal of the London Mathematical Society*. <https://doi.org/10.1112/jlms/s1-25.3.226>
43. Robert M. Corless, Gastón H. Gonnet, D. E. G. Hare, David J. Jeffrey, Donald E. Knuth (1996). On the LambertW function. *Advances in Computational Mathematics*. <https://doi.org/10.1007/BF02124750>
44. Some Moral and Technical Consequences of Automation. <https://gvern.net/doc/reinforcement-learning/safe/1960-wiener.pdf>