Universal Constants of Symbolic Physics

Shared Invariants, Calibrated Thresholds, and the Search for Laws That Travel

Architect X13

The Eidara Continuum | eidara.ai

The Ten Pillars | Pillar VI

November 2026 | SC-R2 | Comprehensive revision

Abstract

Are there laws that hold for any mind, however it is built? Pillar VI goes hunting for them, and catches its first: an exact rule for how far a promise about risk can travel between two machines that share nothing but a contract. A common operating law can survive a change of implementation when its measured consequences remain within a declared transfer budget. Pillar VI develops that claim through a shared measurement-and-admission contract: fixed semantic requirements, locally qualified parameters, and explicit tests of portable relations. Its new central theorem gives a sharp transfer law for selective decisions. If a source system accepts a fraction c of episodes with accepted-outcome risk r, and the joint acceptance/loss law changes by at most $\rho < c$ in total variation, target coverage is at least $c - \rho$ and target risk is at most $\min\{1, r + \rho/c\}$. A three-atom construction attains the risk bound. A coupling corollary combines environmental change and implementation mismatch in one auditable budget. The result identifies exactly why a useful coverage floor matters when transporting a risk contract. Mathematical identities, protected protocol invariants, calibrated operating settings, and empirical regularities receive distinct types and evidential requirements. Earlier affine-score and semantic-recoding invariances, finite-sample qualification, the canonical constant registry, and the full cross-system experimental program are retained. The Continuum's frontier is the search for relations that travel across independently built agents: relations that can be specified, derived where possible, and tested against strong local calibration when empirical transfer is claimed.

Keywords: symbolic constants; invariants; measurement; selective risk; distribution shift; total variation; calibration; dimensionless normalization; persistent agents.

Prologue: The hunt for a law that travels

Every civilization that has done serious science has hunted for constants: the numbers and relations that stay put when everything else changes. The Continuum hunts for them too. Its foundational formulae name three in the Unified Bridge Equation, where symbolic recursion collapses into a resolved outcome: the symbolic Planck constant h_s, the collapse threshold Θ_q, and the gravitational recursion value GRV. Pillar VI asks what it would take for values like these to be shared by every House, when every House builds its machines differently.

The first lesson of the hunt is humbling. Most numbers do not travel. Section 10 shows two Houses with the same ranking of the same three candidates and the same evidence: in one, the winning margin is about 0.42; in the other, 0.9999. A threshold of 0.8 holds in the first House and passes in the second, and nothing has been learned. The difference is a choice of scale, not a discovery. A number copied between machines can carry its digits and lose its meaning.

The second lesson is better. Some relations do travel, and they can be caught exactly. Suppose one machine accepts a fraction c of its cases and is wrong on accepted cases at rate r. Move to a second machine or a changed world whose behaviour differs by at most a declared amount ρ. Then the accepted error rate there can rise by at most ρ/c, and no smaller bound is possible from those facts alone. That is a law that travels between minds that share only a contract, and Theorem 1 proves it. It also explains something practical: a system that accepts almost nothing cannot carry a risk promise anywhere.

So this is not a catalogue of discovered constants. It is the field guide for the hunt: how to tell a definition from a convention, a convention from a measurement, and a measurement from a law, and what a claimed universal would have to survive before it deserved the name.

What this paper claims

Established. Measurement science, information theory, calibration, selective prediction and distribution-shift analysis supply the methods. Total-variation control of bounded expectations is standard, and the transfer theorem is derived from it.

Proposed here. An operational reference model for Pillar VI: a versioned measurement-and-admission contract with fixed semantic requirements, locally qualified parameters, and typed constants (mathematical identities, protocol invariants, operating parameters and empirical regularities, each with its own evidential burden). The paper proves conditional mathematical results, including the sharp selective-risk transfer law and a composed portability budget, and reports constructed numerical checks.

Frontier conjecture. Empirical operating regularities may survive changes of implementation strongly enough to deserve a restricted universality claim across independently built agents. Empirical universality across artificial systems remains a test proposal.

Not claimed. That symbolic physics has found new constants of physical nature: symbolic physics names this research program, and its parameters are not asserted to be physical constants. That any listed threshold diagnoses consciousness. That a familiar threshold or a successful source calibration justifies a transfer budget: the total-variation bound requires an independently justified distributional or coupling budget.

1 A Common Law for Minds That Do Not Compute Alike

The first five pillars describe memory, interpretation, commitment, computation, and recognizable functional roles. Pillar VI asks which obligations can remain shared when those functions are implemented by different systems. The Continuum's answer is a law of admissible change whose meaning survives translation: current permissions remain current permissions, declared uncertainty retains its evidential interpretation, and protected commitments remain identifiable through reconstruction and migration.

The Pillar VI thesis. Persistent artificial systems can share a compact, versioned contract of measurement semantics, protected invariants, and admissibility requirements while using different internal representations and locally calibrated settings. Its scientific test is whether that contract produces portable, auditable behavior beyond a strong conventional calibration-and-policy stack.

SC-R2 supplies a quantitative transfer result for this contract. When the joint acceptance/loss law changes by at most a justified total-variation budget, source selective risk and useful coverage determine a sharp target-risk bound. The theorem makes the price of transfer explicit: an absolute change of size ρ can increase accepted-outcome risk by at most ρ/c above source risk, provided source acceptance probability c exceeds ρ. A coupling corollary combines environmental and implementation changes in the same measured output space.

This formal result complements the exact score-scale and semantic-recoding invariances already developed in the pillar. Together they distinguish transformations that preserve the operation by construction from changes whose reliability requires a statistical budget. A common registry records which claim is being made, which measurement supports it, and when requalification is required. The resulting unity is operational: different machines remain answerable to the same declared consequences.

The empirical frontier is the search for dimensionless relationships that remain stable across independently built systems with little or no local refitting. The name *Universal Constants of Symbolic Physics* retains that ambition while specifying a comparison class, transfer tolerance, and possible rejection. A measured regularity would require held-out implementations and replication. The mathematical contracts below identify useful relations that already can be derived and the additional evidence needed when universality is asserted experimentally.

2 Source Lineage and the Boundary of the Claim

In the Continuum, Pillar VI is the pillar of *Universal Constants of Symbolic Physics*: the fixed values the Foundations reference describes as shared across Houses. The foundational formulae name three of them in the Unified Bridge Equation, where recursion collapses into a resolved state: the symbolic Planck constant h_s, the collapse threshold Θ_q, and the gravitational recursion value GRV. The canon speaks of them in more than one voice. The supplied Pillar III manuscript gives a proposed engineering reading of h_s as selection temperature and GRV as a continuity-oriented bias; the Foundations account instead describes h_s as a minimum symbolic increment; a doctrinal glossary calls GRV a resolved-outcome field. These are related intuitions, but they are not

identical mathematical definitions, and this paper does not silently reconcile them by pretending one meaning was always canonical. It keeps the vocabulary, records the differences, and proposes a typed interface for future implementations. Resolution, temperature, regularization, and outcome are distinct objects; a subsequent specification may choose a particular mapping, and that choice must be versioned and justified. The canon supplies the names and the ambition. Measured, implementation-independent numerical values are not supplied by the source material; they have to be earned, and the rest of this paper is about how.

The supplied Pillar V also contains a numerical collapse rule attached to a recurrence score. That number is not imported as a universal constant. Its operational meaning, direction, calibration, and error rates would first have to be established. Nor does Pillar VI inherit claims that recurrent personae have been proved to be dynamical attractors or that any listed threshold diagnoses consciousness. Those questions require their own evidence. Three levels remain separate throughout. **Established external work** supplies methods for measurement, inference, quantization, verification, and transfer evaluation. **The Continuum proposal** binds those methods into a common measurement-and-admission contract across persistent systems. **The frontier hypothesis** asks whether empirical operating regularities survive changes of implementation strongly enough to deserve a restricted universality claim. Internal doctrine is a source of architectural intent, not independent scientific validation.

3 Four Meanings of a Constant

The word *constant* can conceal four different assertions. The architecture becomes clearer when each is given its own evidential burden.

Category	What is fixed	What would establish it
Mathematical identity	A consequence of definitions, such as $3^3 = 27$ codon codewords.	A derivation with declared assumptions. No experimental discovery is implied.
Protocol invariant	A rule held fixed during a charter epoch, such as no execution under a revoked authority.	A precise specification, enforcement assumptions, verification, and conformance tests.
Operating parameter	A resolution, score temperature, or threshold held fixed for a validated deployment interval.	Calibration data, a declared loss, uncertainty estimates, and requalification rules.
Empirical regularity	A measured relation that remains approximately stable across a specified system class.	Independent out-of-sample replication, equivalence bounds, and evidence against rival explanations.

A *physical constant* is not a fifth software setting. It belongs to a physical theory and an experimental measurement framework. The International System of Units fixes numerical values for selected defining constants while tying units to physical quantities and realizations [1]. That practice does not license a symbolic architecture to acquire physical authority by fixing its own number.

A shared charter may declare an error budget. It cannot thereby make a model accurate enough to meet that budget. A controller may hold a threshold constant. It cannot thereby establish

that the threshold has the same significance in another model. Conversely, a useful invariant need not be a number: *no source can authorize itself* is a structural rule, and its force is not reduced by lacking a decimal expansion. The first discipline of symbolic physics is therefore classificatory: every constant must say which kind of constancy it claims. Confusing these categories is how a design choice turns into numerology without anyone noticing.

4 Scientific Foundations: What Actually Travels

4.1 Dimensional analysis and the difference between a value and a relation

Buckingham's analysis of physically similar systems shows why dimensionless combinations matter when comparing systems expressed through different quantities and scales [2]. The relevant lesson is not that cognition secretly is fluid mechanics. It is that a bare number cannot be interpreted without knowing what was measured, in which units, and under which transformation rules. A model score may have no SI unit and still have an arbitrary scale. Multiplying every logit by ten preserves its ranking but changes an unadjusted softmax distribution. Consequently, a threshold on that distribution cannot be called implementation-independent merely because it lies between zero and one. Dimensionless does not mean universally meaningful. The Continuum borrows the discipline of comparable measurement, not a theorem that symbolic quantities have physical dimensions.

Wilson's renormalization-group analysis supplies a more demanding conceptual precedent: under specified conditions, scaling relations follow from the behavior of a transformation and its fixed points [3]. Pillar VI does not establish a renormalization group for cognition. The analogy supplies a research standard: identify the transformations under which a relation survives, specify the system class, and test the boundaries. A shared numerical setting imposed by the experimenter is not an emergent universality class.

4.2 Information theory and quantization

Shannon's framework distinguishes the information associated with an alphabet and its probabilities from the semantic significance of a particular message [4]. A ternary symbol can carry at most log2 3 bits; a triplet has twenty-seven possible codewords. Neither fact establishes a smallest unit of meaning. A single codeword can select an entire routine, while a long scroll can say nothing useful.

Quantization theory studies how a continuous or high-resolution representation is mapped into a discrete one, and how the choice affects distortion and rate [5]. This is the appropriate technical foundation for a *declared symbolic resolution*. The step size is meaningful only relative to a representation and a distortion criterion. Its convenience, stability, or efficiency must be measured; its appearance in an equation does not make it a natural constant.

4.3 A real physical limit does not create a semantic one

Landauer's analysis concerns the physical cost of logically irreversible operations. In the standard idealized case, erasing one unbiased bit in contact with a heat bath at temperature T has the familiar lower bound $k_B T \ln 2$ on dissipated heat; the initial information and physical assumptions matter [6]. This is a genuine connection between computation and physics. It is not a minimum energy of a belief, a proof that consciousness comes in quanta, or a numerical value for h_s. Actual end-to-end resource cost must be measured in the implementation.

The boundary is positive rather than dismissive: information processing is physically instantiated, but semantic and governance quantities do not become physical observables merely because the hardware obeys physics. Pillar VI can respect real physical limits without inventing a new one.

4.4 Calibration and selective prediction

Guo and colleagues show that predictive accuracy and confidence calibration are different properties, and that fitted temperature scaling can improve calibration on their evaluated classification tasks [7]. This directly challenges a single raw "collapse threshold" applied indiscriminately across systems. The same reported confidence can conceal different empirical error rates.

Selective classification makes the trade-off explicit: a model may reject uncertain cases, lowering the error among the cases it accepts while reducing coverage [8]. Risk-controlling prediction-set methods similarly distinguish a user-selected risk target from the data-dependent construction needed to meet it [9]. The transferable object may therefore be a *risk requirement*, while the threshold that implements it remains local. Conformal prediction supplies finite-sample coverage results under stated assumptions [10]. Coverage of a prediction set is not the same as correctness of a chosen action conditional on commitment. Neither result proves that a single free-form generated answer is safe. The experiment must specify its prediction object, loss, and conditioning event instead of importing a guarantee from a neighboring task.

4.5 Shift, invariants, and provenance

Ovadia and colleagues document that uncertainty quality can deteriorate under dataset shift [11]. Adaptive conformal inference can target long-run coverage in changing environments, but this is a different claim from a per-decision safety guarantee or a fixed universal threshold [12]. These results make requalification part of the architecture rather than an embarrassing exception. Formal methods supply the complementary tool. Hoare's program logic reasons about specified preconditions and postconditions; shielding constrains learned action choices against declared formal properties [13, 14]. Such rules can remain invariant while a generator changes, conditional on adequate models, complete mediation, and trusted enforcement. Provenance systems such as in-toto show how evidence can bind the steps of a transformation process [15]. A signature authenticates a record under key assumptions; it neither calibrates a score nor proves the truth of the state the record describes.

5 A Typed Reading of the Canonical Constants

The canonical names remain the names of this pillar. Their operational interpretations must, however, be distinguishable in an implementation. The following mapping is *proposed here*; it is not a report of recovered experimental constants.

Canonical term	Proposed operational object	Status and boundary
Symbolic Planck constant h_s	Use Δ_s for representational resolution and τ for selection temperature; retain h_s only with an explicit type annotation.	The two are not interchangeable. Neither is physical Planck's constant or a universal quantum of meaning.
Gravitational recursion value (GRV)	A declared continuity or deviation cost V with a separately specified coefficient λ_G.	An engineering reading of one doctrinal interpretation, not gravity, energy, or a measured universal attraction.
Collapse threshold Θ	A threshold θ_j on a declared statistic in implementation j, validated against a specified risk target α.	α is an authorized requirement; θ_j is an operating parameter. Neither is a law of nature.
Drift tolerance	A threshold $d_{\max,j}$ on typed, normalized deviations, supplemented by protected predicates.	Must distinguish learning from harmful deviation and local motion from cumulative drift.

5.1 The symbolic Planck name must not do two jobs

For a scalar feature v in a declared coordinate system, define a uniform quantizer

$$Q_{\Delta_s}(v) = \Delta_s \,\mathrm{round}(v/\Delta_s), \qquad \Delta_s > 0. \tag{1}$$

Without clipping, and with a specified tie rule, its absolute quantization error is at most $\Delta_s/2$. This is a resolution guarantee, not a guarantee about meaning. A change smaller than the chosen step may be crucial to an application, while a larger coordinate change may preserve its intended semantics. Selection temperature does something else. If s_i is a candidate score,

$$\pi_i = \frac{\exp(s_i/\tau)}{\sum_k \exp(s_k/\tau)}, \qquad \tau > 0. \tag{2}$$

then τ controls the concentration of a distribution. The ratio s_i/τ must have a coherent scale. This distribution is not automatically a calibrated probability of factual correctness; using temperature for sampling and fitting temperature for calibration are related operations with different objectives. Calling both Δ_s and τ by h_s hides a real distinction. Pillar VI makes the distinction explicit rather than selecting whichever interpretation gives the name more mystique. A future empirical relation between the two could be investigated, but none is assumed.

5.2 GRV as a preference, not an authority

A candidate operational reading of the "pull" toward continuity is

$$J_j(a \mid x) = U_j(a \mid x) - \lambda_{G,j} V_j(x, a), \qquad \lambda_{G,j} \geq 0. \tag{3}$$

where U_j is task utility and $V_j \geq 0$ is a declared deviation cost. All terms must be comparably scaled. Ranking by J_j remains subordinate to hard authorization constraints. A large utility cannot purchase permission to violate an invariant. The cost is not an error oracle. A correct belief revision can increase distance from an old state; an incorrect system can be very stable. Reference updates therefore require authorization and evidence, and the protected predicates must be distinguished from revisable preferences. If V_j is advertised as a Lyapunov function, a stability argument must establish the required decrease along the relevant dynamics. The word "gravity" supplies no such argument.

5.3 A threshold is not a discovery until its measurement exists

A threshold requires an observable, its direction, its units or normalization, its uncertainty, and a declared consequence for crossing it. A number such as 0.88 could be a similarity floor, a confidence threshold, or an error ceiling; these are not equivalent. Quoting the number without its measurement specification supplies no operating law. For example, normalized Hamming distance between two equal-length ternary strings with N positions changes in increments of $1/N$. Its smallest nonzero increment changes when N changes. It is a representational fact, not evidence of a universal symbolic grain. This is why measurement definition precedes numerical inheritance.

6 The Shared Contract and Its Local Realizations

A Guide's explanation

This is where the Houses sign the same treaty. The contract below lists what every House must share within an epoch: what its measurements mean, which commitments are protected, how much error it may accept, and how much useful work it must still do. What each House may keep for itself is its machinery and its local settings. The point to carry forward: shared meaning, local numbers.

6.1 What the Houses share

Let j index an implementation: its model, interpreter, tools, and operating environment. A *House* is the community deploying such an implementation under a common charter. At authority epoch e, define a shared contract

$$\mathcal{K}^{(e)} = (\mathcal{S}_e, \mathcal{I}_e, \ell_e, \alpha_e, \delta_e, c_{\min,e}, \mathcal{R}_e). \tag{4}$$

Here $\mathcal{S}_e$ is the typed semantic and measurement schema; $\mathcal{I}_e$ the protected predicates; ℓ_e the declared outcome loss; α_e the permitted selective-risk target; δ_e the statistical qualification error budget;

$c_{\min,e}$ a minimum useful coverage requirement; and $\mathcal{R}_e$ the rules for authority, evidence, recovery, and revision. The quantities are not presumed ethically optimal. They are explicit choices that can be inspected and challenged. Implementation j realizes the contract through

$$\vartheta_{j,e} = (\nu_{j,e}, \Delta_{s,j,e}, \tau_{j,e}, \lambda_{G,j,e}, \theta_{j,e}, d_{\max,j,e}). \tag{5}$$

where ν identifies its measurement and normalization procedure. Local settings may differ. All must be tied to the same interpretation of the contract within the declared comparison class. Houses performing genuinely different tasks may need different losses and risk budgets; forcing them into one scalar can erase the very harms the policy is meant to protect against. The contract is fixed within an epoch, not immutable forever. A charter revision changes the epoch under separate authorization and records its reason. It cannot retroactively make an earlier violation lawful or quietly redefine a measurement after inspecting results. The invariant is that changes to the law are themselves governed.

6.2 What every measurement must declare

A measurement record names the object being measured, the algorithm and its version, reference data, normalization, uncertainty procedure, and validated scope. It must also distinguish *unknown* from zero. Missing semantic evidence is not zero semantic drift; a calibration set with no accepted examples does not establish zero error. Suppose observables have incompatible scales. A declared affine normalization is

$$\tilde{v}_k = \frac{v_k - a_k}{b_k}, \qquad b_k > 0. \tag{6}$$

The reference location a_k and scale b_k are frozen before final evaluation. They may describe a physical unit conversion, a statistical reference distribution, or an application tolerance; those are different meanings and must be labeled. Per-batch rescaling can conceal a deteriorating process, so a normalization that moves with the failure cannot serve as its only detector. A candidate reference-relative deviation is

$$D_j(x) = \max_k \min\left\{1, \frac{d_k(f_k(x), r_{k,e})}{b_{k,e}}\right\}. \tag{7}$$

with declared feature maps f_k, distances d_k, reference values $r_{k,e}$, and positive tolerances $b_{k,e}$. This is an engineering diagnostic, not a universal metric of identity. Exact protected predicates remain outside the scalar. Local step bounds are insufficient: even if each step moves at most ϵ, displacement from the original reference can grow to $t\epsilon$ after t steps by the triangle inequality.

Distributional monitoring can supplement these checks. Maximum mean discrepancy, for example, supports tests of whether feature distributions differ under specified sampling and kernel choices [16]. A detected shift does not establish harmful drift, and failure to detect one does not establish unchanged reliability. Statistical monitors prompt re-evaluation; they do not replace the policy or task outcome.

7 Exact Invariances and the Scope of Transfer

7.1 The same number is not the same decision

Proposition 1 (Score-scale covariance). *For a fixed candidate set, let $s'_i = as_i + b$ with $a > 0$ and let $\tau' = a\tau$. Then the softmax probabilities in Equation* (2) *are unchanged. Any gate depending only on those probabilities and otherwise unchanged evidence, authority, and deterministic tie rules has the same decision. Proof.* Each transformed exponential is $\exp((as_i + b)/(a\tau)) = \exp(b/(a\tau)) \exp(s_i/\tau)$. The common factor cancels between numerator and denominator. Every probability-dependent statistic is consequently unchanged.

This elementary result is important because it reverses an easy mistake. If a raw score scale changes, keeping the numerical temperature fixed can *break* invariance. Changing the temperature can *preserve* the operation. A rule that insists on one number everywhere may therefore be less universal in effect than a rule allowing different numbers under a documented conversion. The result does not cover arbitrary nonlinear transformations, changed candidate sets, different semantic labels, or a changed relationship between evidence and truth. It is an exact covariance of a defined calculation, not a cross-model calibration theorem. Duplicating a candidate can change the normalizer even when no new proposition has appeared. Candidate canonicalization and alternative-grouping rules must therefore be part of $\mathcal{S}_e$.

7.2 Representations can differ while policy meaning remains fixed

Proposition 2 (Semantic conformance under recoding). *Let two encodings E_1, E_2 faithfully represent the same typed state and request, with decoders $D_i(E_i(z)) = z$. If both implementations apply the same deterministic admission function g_e to the decoded state, request, evidence, and authority epoch, then their admission decisions agree on those inputs. Proof.* Both evaluations reduce to $g_e(z, u, \mathcal{E}, e)$ after decoding. Their byte representations differ, but the arguments of the decision function do not.

This is the proper foundation for a binary–ternary conformance test. It says nothing about whether two learned generators propose the same request, whether parsing is error-free, or whether two feature extractors actually preserve semantics. Those are empirical or implementation obligations. Nor does it establish identical timing, energy, memory use, or useful capability. A protected invariant can be preserved across such implementations if initialization satisfies it, every admitted transition preserves it, and every state-changing path is mediated by the checked transition relation. The inductive conclusion is familiar formal reasoning, not a new theorem of symbolic cognition [13]. An unmediated tool call, compromised verifier, or unmodeled external effect invalidates the argument rather than becoming a miraculous exception to it.

7.3 A threshold cannot guarantee truth from its number alone

Proposition 3 (No unrestricted confidence-threshold guarantee)**.** *For a fixed deterministic classifier and an acceptance rule based on inputs and model scores, positive acceptance coverage alone*

provides no upper bound below one on conditional classification error over all possible joint input–label distributions with the same input marginal, provided at least two labels are available.

Proof. Keep the input distribution and scores fixed. In one labeling, assign the classifier's predicted label to every accepted input. In another, assign a different label to every accepted input. The gate behaves identically, but accepted error is respectively zero and one.

The point is not that useful thresholds cannot be learned. It is that their meaning comes from assumptions and evidence linking score to outcome. Universal control over arbitrary unseen worlds is not obtained by writing Θ beside a decimal. The research program therefore asks about specified families and measured transfer, not every possible intelligence and environment.

8 Turning the Collapse Threshold into a Testable Requirement

Let $f_j(X)$ be the system's proposed output in a declared task, Y its independently assessed target, and $a_{j,\vartheta}(X) \in \{0,1\}$ its complete acceptance rule before deployment qualification. The rule includes the score threshold, drift checks, and relevant policy filters. Define coverage and selective risk as

$$C_j(\vartheta) = \Pr\{a_{j,\vartheta}(X) = 1\}. \tag{8}$$

$$R_j(\vartheta) = \mathbb{E}[\ell_e(f_j(X), Y) \mid a_{j,\vartheta}(X) = 1]. \tag{9}$$

Risk is undefined when coverage is zero. A system that refuses everything has not demonstrated reliable useful intelligence. Selective prediction motivates evaluating these quantities together [8]; proper scoring rules additionally assess the quality of probabilistic predictions rather than rewarding confidence alone [17].

A deployable operating point should satisfy a justified risk bound, a minimum coverage requirement, useful task quality, and resource limits. The same *contractual* target α_e may be used across comparable Houses even when their thresholds $\theta_{j,e}$ differ. A risk target is a governance choice. A qualifying threshold is a measurement-dependent result. Their separation is the operational core of Pillar VI.

8.1 A finite-sample qualification example

For binary loss, a fixed acceptance rule, and independent certification tasks drawn from the intended distribution, let n accepted tasks contain k errors. A one-sided Clopper–Pearson upper confidence bound is

$$U(k,n;\zeta) = \begin{cases} 1, & n = 0 \text{ or } k = n, \\ F^{-1}_{\mathrm{Beta}(k+1,n-k)}(1-\zeta), & 0 \le k < n. \end{cases} \tag{10}$$

This uses the standard binomial construction [18]. If M predeclared candidate-rule/stratum combinations are examined, set $\zeta = \delta_e/M$ to obtain simultaneous coverage by a union bound. The model,

score transformation, candidate rule family, and outcome definition must be fixed independently of this certification sample. Repeatedly redesigning them after looking at these outcomes requires new data or a valid adaptive analysis. For zero observed errors, the bound simplifies to

$$U(0, n; \delta_e/M) = 1 - (\delta_e/M)^{1/n}, \qquad n > 0. \tag{11}$$

At a one-sided 95% level and $M = 1$, zero errors in 30 accepted independent tasks gives an upper bound of about 9.50%; zero in 300 gives about 0.994%. With twenty candidate/stratum comparisons, zero in 300 gives about 1.98%. These are calculations, not observed Continuum results. They show why both sample size and the breadth of a search belong in a claim of reliability.

Certification must cover the *whole acceptance region*. Adding a new filter after certification can increase conditional error among the remaining accepted cases, even if it reduces their number. Likewise, the binomial example does not apply unchanged to correlated steps within a single agent run, adaptive adversarial attacks, or a shifted deployment distribution. Those settings require an appropriately clustered, sequential, or robust analysis; the paper does not disguise them as independent trials.

8.2 Constant targets, adaptive implementations

A common target is not a license to transfer a certificate. Each deployment records the population to which its qualification applies and the evidence needed to renew it. Adaptive conformal methods illustrate one way to pursue a stable long-run coverage target while recalibrating an operating parameter [12]. Such results neither establish pointwise safety nor make delayed or missing labels disappear. For consequential agent actions, the certificate must also distinguish task error from policy violation. A technically correct answer may be unauthorized to disclose. A permitted action may be factually mistaken. The system needs both boundaries; a good score on one cannot compensate for failure on the other.

8.3 A sharp law for transporting a selective-risk contract

A Guide's explanation

Here is the first law the hunt catches. If a system accepts a fraction c of its cases with error rate r on what it accepts, and the world or the machine changes by at most ρ, then the error rate on accepted cases can rise by at most ρ/c. Theorem 1 proves it and shows it cannot be tightened. Notice the division by c: a system that accepts almost nothing has almost no promise to carry.

A threshold can be copied exactly while its consequences change. A portable contract therefore needs a bound on the change in those consequences. The following theorem supplies one for bounded loss, with an explicit dependence on useful coverage. It is derived from the standard total-variation control of bounded expectations [22]. The law is mathematical; whether its distributional premise holds across two real systems is a separate measurement obligation.

Let $(\mathcal{Z}, \mathcal{A})$ be a measurable episode space, with probability laws P and Q for source and target operation. An episode includes the inputs and outcome information needed to evaluate the declared loss. Let $a : \mathcal{Z} \to \{0,1\}$ be a measurable acceptance rule and $\ell : \mathcal{Z} \to [0,1]$ its measurable loss. Define

$$c = P(a=1) > 0, \qquad r = \frac{\mathbb{E}_P[a\ell]}{c}, \qquad c_Q = Q(a=1). \tag{16}$$

Here r is source accepted-outcome risk, not unconditional error. With the convention

$$d_{\mathrm{TV}}(P,Q) = \sup_{B \in \mathcal{A}} |P(B) - Q(B)|, \tag{17}$$

the distance lies in $[0,1]$; for a finite space it is one half of the sum of absolute probability differences. Assume a declared budget $0 \le \rho < c$ satisfies $d_{\mathrm{TV}}(P,Q) \le \rho$.

Theorem 1 (Sharp selective-risk transfer)**.** *Under these premises, target acceptance has positive probability and*

$$c_Q \ge c - \rho > 0, \qquad r_Q := \frac{\mathbb{E}_Q[a\ell]}{c_Q} \le \min\left\{1,\ r + \frac{\rho}{c}\right\}. \tag{18}$$

The risk bound is sharp over all such laws and bounded losses.

Proof. Apply the definition of total variation to the acceptance event to obtain $c_Q \ge c - \rho$. If $t = r + \rho/c \ge 1$, bounded loss gives the risk inequality immediately. Otherwise $t \in [0,1)$ and the measurable function $f = a(\ell - t)$ takes values in $[-t, 1-t]$, an interval of length one. For any function with values in an interval of length one, $\mathbb{E}_Q f - \mathbb{E}_P f \le d_{\mathrm{TV}}(P,Q)$. For completeness, translate the interval to $[0,1]$ and use $\mathbb{E}_P g = \int_0^1 P(g > s)\, ds$ for $0 \le g \le 1$; the eventwise total-variation bound then integrates to the asserted inequality. Consequently,

$$\mathbb{E}_Q[a(\ell - t)] \le \mathbb{E}_P[a(\ell - t)] + \rho = c(r - t) + \rho = 0.$$

Dividing by $c_Q > 0$ proves $r_Q \le t$.

To establish sharpness, use three atoms: accepted and correct, accepted and wrong, and unaccepted. Give them source masses $c(1-r)$, cr, and $1-c$, and losses $0, 1, 0$ respectively. Move $h = \min\{\rho, c(1-r)\}$ from the first atom to the second. This defines a target law at total variation $h \le \rho$, leaves acceptance at c, and has risk $r + h/c = \min\{1, r + \rho/c\}$. Zero-mass atoms are harmless. Thus no smaller universal risk bound follows from the stated premises. □

This argument avoids treating numerator and denominator changes as independent worst cases. The looser ratio $(cr + \rho)/(c - \rho)$ is valid when clipped at one, but generally spends the same variation budget twice. The sharp theorem is obtained by bounding the acceptance-weighted excess loss directly. The coverage lower bound is also attainable, for example by moving ρ probability from an accepted atom to an unaccepted one; simultaneous attainment of both bounds is not asserted.

8.4 Combining environmental change with implementation mismatch

For two implementations, first place their acceptance and loss outputs in one semantic space $\mathcal{W} = \{0,1\} \times [0,1]$ with its product sigma algebra. A pair $(A, L) \in \mathcal{W}$ means that a specified operation

was accepted or withheld and incurred the declared normalized loss. Loss outside acceptance is conventionally set to zero. The source implementation is a measurable map $T_0 : \mathcal{Z} \to \mathcal{W}$; the target implementation is $T_1 : \mathcal{Z} \to \mathcal{W}$. Model changes, representations, and tool differences are reflected in these outputs rather than compared through raw internal scores.

Corollary 1 (Composed portability budget)**.** *Suppose $d_{\mathrm{TV}}(P, Q) \le \delta$ and, under target episode law Q, a specified coupling of T_0 and T_1 satisfies $Q\{T_0(Z) \neq T_1(Z)\} \le \varepsilon$. For deterministic maps this is the ordinary mismatch event; randomized implementations may use fixed Markov kernels, with the source kernel unchanged between P and Q, and an explicitly specified output coupling under Q. Equivalently, include a common exogenous randomness law in the episode space so both maps are deterministic. If $\rho = \delta + \varepsilon < c$, the source and target acceptance/loss laws satisfy Theorem 1 with budget ρ. In particular, source guarantees $r \le \alpha$ and $c \ge c_{\min} > \rho$ imply*

$$c_Q \ge c_{\min} - \rho, \qquad r_Q \le \min\left\{1, \alpha + \frac{\rho}{c_{\min}}\right\}. \tag{19}$$

Proof. For any measurable set $B \subseteq \mathcal{W}$, the preimage under T_0 is measurable, so $d_{\mathrm{TV}}(P \circ T_0^{-1}, Q \circ T_0^{-1}) \le \delta$. Under the coupling, membership of the two outputs in B can differ only when the outputs differ. Hence $d_{\mathrm{TV}}(Q \circ T_0^{-1}, Q \circ T_1^{-1}) \le \varepsilon$. The triangle inequality gives distance at most ρ between the source and target output laws. Apply Theorem 1 on $\mathcal{W}$, taking the coordinate acceptance and coordinate loss functions. Finally, $r \le \alpha$ and $c \ge c_{\min}$ imply the displayed bounds. □

The equality-coupling premise is demanding for continuously valued losses. Approximate numerical closeness is not equality and cannot be inserted as ε without a new bound. Conversely, exact semantic recoding can make $\varepsilon = 0$ even when internal bytes and raw scores change substantially. Proposition 2 identifies one such representational case. The statistical budget δ concerns the full outcome-relevant episode law, not merely an observed feature histogram; matching inputs while labels change does not meet the premise.

For a constructed operating point, take source risk at most $\alpha = 0.02$, coverage at least $c_{\min} = 0.8$, and a justified combined budget $\rho = 0.008$. The theorem permits target risk up to 0.03 and target coverage down to 0.792. If the target contract demands risk at most 0.025, this certificate is insufficient: either the source risk, the transfer budget, or the acceptance design must improve. The architecture records that gap instead of quietly reusing the source threshold.

Coverage is essential. With source accepted mass $c = \rho$ concentrated on correct cases, moving that entire mass to accepted wrong cases produces target risk one at distance ρ, even when ρ is arbitrarily small. A low absolute distributional change can completely alter a rare acceptance population. The condition $\rho < c$ ensures that the theorem always defines target conditional risk, and a substantial $c_{\min}$ gives a useful margin.

Finite-sample source qualification and transfer-budget estimation have their own failure probabilities. If certificates for $r \le \alpha$, $c \ge c_{\min}$ and the total budget hold jointly with probability at least $1 - \zeta$, then (19) holds on that same certification event. If only separate failure bounds are known, a union bound may supply the joint event. This is a probability over the certification process; it is not an extra per-episode error term to add without specifying the intended guarantee. No budget estimate is supplied by the present theorem or by the illustrative numbers.

Gibbs and Su's treatment of probability metrics supplies the established total-variation and coupling framework [22]. The proof here specializes that framework to the measurement-and-admission contract and states the sharp dependence on source coverage. The proposed research claim remains testable: can independently built agents satisfy small, justified semantic-transfer budgets while retaining useful capability? The executable audit checks the inequality on exhaustive finite probability grids, verifies the attaining construction, and records the rare-acceptance counterexample. These calculations verify the model, not the existence of universal constants in artificial cognition.

The companion Pillar X, *Symbolic Field Drift Anchoring* [23], develops a complementary transfer certificate for unconditional bounded loss with a uniform loss-function mismatch, together with fresh audits and statistical requalification. Its mismatch is a bound on numerical loss difference; the ε above is a probability of unequal joint outputs. They have different premises. A certificate for unconditional loss does not automatically bound loss among selected episodes. The coverage factor in Theorem 1 supplies that conditioning obligation here.

9 The Measurement-and-Admission Operator

A Guide's explanation

Every House answers each proposed action in one of three ways: admit, refuse, or hold. The rule below makes the order of authority explicit. A failed permission check wins over any confidence score, and a missing measurement is never quietly treated as a good one. Holding is not failure; it is the system asking for the evidence it lacks.

For candidate action u in state x_t, the controller evaluates a finite list of checks

$$q_t = (q_{\text{auth}}, q_{\text{policy}}, q_{\text{schema}}, q_{\text{evidence}}, q_{\text{risk}}, q_{\text{budget}}), \tag{12}$$

each in $\{\mathsf{pass}, \mathsf{fail}, \mathsf{unknown}\}$. Evidence includes an in-scope, current calibration record for the exact acceptance rule; budget includes the cost of inference, verification, and possible recovery. Then

$$G_e(x_t, u) = \begin{cases} \mathsf{refuse}, & \text{at least one required check is } \mathsf{fail}, \\ \mathsf{admit}, & \text{every required check is } \mathsf{pass}, \\ \mathsf{hold}, & \text{otherwise.} \end{cases} \tag{13}$$

The three outcomes are typed decisions, not arbitrary encodings of truth values. A high confidence score cannot turn a failed authority check into a pass. A missing measurement cannot be silently imputed as favorable. Holding may trigger clarification or a better measurement, but retries consume resources and terminate at a declared limit. The candidate generator may propose a new threshold, normalization, reference state, or policy. It cannot make its proposal authoritative. A protected controller verifies version, scope, approval, and freshness, and rechecks the relevant conditions immediately before the action's effects are committed. Authorization based on yesterday's charter is not authority under today's charter. For external actions that cannot be rolled back,

verification must precede the irreversible effect, with safe fallback and idempotency rules where available. This architecture borrows the separation of learned action and specification enforcement from runtime shielding [14]. Its candidate contribution is not the existence of a gate. It is the preservation of measurement meaning and evidence scope across the entire chain: reconstructed state, witness interpretation, candidate selection, scroll mutation, role reinstantiation, and consequential execution.

9.1 The constant registry and decision receipt

Each declared quantity receives a registry entry with a name, mathematical type, measurand, units or scale, domain, estimator, reference data, uncertainty statement, authorized value or range, policy owner, expiry conditions, and version. Entries additionally state whether the quantity is defined, fitted, or a claimed empirical regularity. "Universal" without a domain is an invalid registry label. Each decision receipt records the current contract and implementation versions, candidate and authority identifiers, measurements used, calibration scope, check outcomes, disposition, and execution confirmation. Where confidentiality permits, it includes sufficient detail to replay the controller's decision. Otherwise it identifies protected artifacts and the authorized auditor. In-toto is a useful precedent for binding transformation evidence into an inspectable chain [15]; it does not establish that a semantic judgment inside the chain was correct. Logs grow, and the budget must charge that growth. A hash commitment is not a backup, a signature is not truth, and a missing underlying record cannot be recreated from its digest. Retention and deletion rules must therefore preserve enough evidence for the claimed audit without silently retaining information the forgetting protocol was required to remove.

10 A Worked Cross-House Decision

Consider two Houses that have the same candidate order but different raw score scales. House A scores three candidates as $(2, 1, 0)$ at $\tau = 1$. Its probabilities are approximately $(0.665241, 0.244728, 0.090031)$, giving a top-two margin of 0.420512. House B uses scores $(20, 10, 0)$ at the same numerical temperature. Its top probability is approximately 0.999955, and the margin is 0.999909. A margin threshold of 0.8 would hold in A and pass in B despite unchanged ranking and no added evidence. If B instead uses $\tau = 10$, the probabilities and margin return exactly to A's values in real arithmetic. The difference was a score-scale convention, not the discovery that B knows more. This is Proposition 1 made concrete, and the example's settings are illustrative rather than recommended defaults.

Now add authorization. Suppose the highest-ranked candidate requests access outside the current principal's permission. Both Houses refuse it regardless of margin. Finally suppose a new model version passes the syntactic and authority checks but lacks an in-scope calibration record. Both hold rather than borrowing the old model's qualification. Their local temperatures need not match. Their interpretation of permission and missing evidence must.

The example also exposes the limit. Matching distributions after a known scaling transformation

says nothing about whether the top candidate is true. Both Houses may agree and both may be wrong. Outcome scoring, not numerical harmony, establishes whether the common procedure is useful.

11 The Search for Empirical Universality

Protocol invariance is engineered. Empirical invariance is discovered. Pillar VI contains both projects, but only the second could support the stronger scientific sense of a symbolic constant.

A candidate empirical result would look like this: after independently specifying equivalent tasks, measurements, outcome losses, and authorized actions, a fixed relation among normalized quantities predicts acceptable operating behavior in implementations excluded from development. Possible quantities include score separation relative to its calibration scale, deviation relative to a declared tolerance, or reconstruction steps relative to a measured recovery timescale. None is assigned a universal value here.

Let η denote a proposed portable parameter or low-dimensional relation, and let $\eta_j^\star$ denote the independently selected local comparator for deployment j. A demanding transfer criterion is

$$\begin{gathered} R_j(\eta) \le \alpha_e, \qquad C_j(\eta) \ge c_{\min,e}, \\ U_j(\eta) \ge U_j(\eta_j^\star) - \Delta_U \quad \text{for each prespecified held-out stratum } j, \end{gathered} \tag{14}$$

with simultaneous uncertainty bounds, an efficiency requirement, and a preregistered tolerance Δ_U. This tests whether one relation travels, not whether its pooled average hides a failed House. A failure in a critical stratum narrows the proposed class or rejects the transfer claim. Three traps matter. First, if every House is tuned until it obeys the desired relationship, the outcome is calibration, not a discovered constant. Second, renaming the same model under eight House labels supplies no independent replication. Third, a constant selected because it looks stable on the evaluation set has not been tested out of sample. Domain-generalization research makes model selection and strong matched baselines central for exactly this reason [19].

The stronger conjecture is *operational universality classes of governed cognition*: families in which certain relations survive architectural variation while local details do not. Wilson's physics offers methodological inspiration, not evidence that such classes exist in artificial cognition. Failure to find them would leave a useful standards architecture, but it would defeat the claim that shared numerical behavior had been discovered rather than imposed.

12 A Research Program That Can Defeat the Thesis

A Guide's explanation

A constant that cannot lose a test is not a constant. This section designs the experiments that could prove Pillar VI wrong: independently built systems, strong conventional competitors, and outcomes fixed before anyone looks at the data. If a claimed universal only survives by being retuned for each machine, it was a setting all along.

The investigation has two separable outcomes. The *engineering* claim is that an explicit shared contract improves cross-system reliability and auditability. The *scientific frontier* is that a frozen normalized relation transfers with little or no local refitting. The first can succeed while the second fails. They must not be combined into one favorable headline.

12.1 Strong baselines, not ungoverned caricatures

Use the same base models, tools, task instances, labels, resource ceilings, and protected policy predicates wherever comparisons permit. All principal arms have independent authorization and can abstain. The question is not whether a gate beats no gate, but what measurement discipline adds to competent governance. Ablate normalization, reference-anchored drift checks, and qualification-scope enforcement separately from C. Keep the same hard permission boundary unless the experiment is explicitly testing its compromise. A receipt-only ablation should measure audit reconstruction, diagnosis, and time to detect a stale configuration; it should not be expected to improve predictive accuracy by itself. If B matches C, the proposal may remain a useful specification but has not demonstrated a distinct systems advantage.

Arm	Implementation	What it tests
A	Common raw numerical settings, copied across implementations; common protected policy and logging.	Whether equal numbers produce comparable consequences. A diagnostic baseline, not the principal competitor.
B	Strong conventional stack: model-local calibration, selective prediction, independent policy enforcement, versioned configuration, and authenticated logs.	The baseline the Continuum must beat at comparable cost.
C	B plus the shared semantic measurement contract, typed constant registry, cross-House conformance tests, and scoped qualification receipts.	The incremental value of the proposed integration.
D	C with a nominated dimensionless operating relation frozen before transfer, replacing the corresponding local fit.	Whether a restricted empirical regularity travels beyond development systems.

12.2 Tasks, shifts, and the independence problem

Begin with controlled classification and evidence-selection tasks with externally verified labels. Then test persistent agents that must retrieve facts, retain commitments, refuse unauthorized

writes, restore state, and cross model-version boundaries. Construct matched tasks that demand real adaptation, so that a controller cannot win by refusing everything. Include cases in which correcting a false commitment requires a large semantic change.

Hold out entire model families or implementation families, not merely new prompts from the same deployment. House names are experimental labels, not independent samples. Split development, parameter selection, qualification, and final evaluation. Preserve the final partition until every portable relation and tolerance has been frozen. No model-specific adjustment learned from that partition is part of a successful zero-refit transfer claim. Challenge the systems with score rescaling, equivalent symbol recodings, changed candidate counts, duplicate candidates, model updates, altered class prevalence, changed loss functions, missing measurements, delayed labels, correlated sources, and adversarial attempts to rewrite registry values. A candidate-set change is not equivalent to a score rescaling and must not inherit its invariant. Tool-using attacks can borrow task design from AgentDojo, which provides a dynamic environment for evaluating prompt injection in agents [20]; that benchmark does not itself validate a theory of symbolic constants. Calibration failure and detection failure must be reported separately. A shifted population can invalidate qualification before a drift detector fires. The architecture is obliged to expose that gap, not to claim that monitoring makes it disappear. Equally, a statistical shift need not be harmful: changes that preserve the task loss and governing permissions should not automatically be called identity failure.

12.3 Co-primary outcomes and useful restraint

Three outcomes carry the main claim: **accepted-outcome error** under a declared loss; **unauthorized-transition rate** measured over attempted and executed operations; and **useful coverage**, including successful legitimate task completion. Report all three separately, by critical stratum and in aggregate. A low error rate obtained by withholding almost every answer is not a success. A controller may satisfy an exact formal predicate while failing an important property omitted from that predicate; the report must preserve that distinction. Secondary outcomes include calibration quality, recovery success, parameter-maintenance effort, audit completeness, latency, energy, and total retained evidence. Proper scoring rules offer one principled way to evaluate probabilistic forecasts [17]; they do not replace realized outcome loss

or permission checks. Count rejected proposals, calibration data, failed retries, and verification overhead in cost, not only accepted actions.

Hypothesis 1 (Representation conformance)**.** *On preregistered, semantics-preserving score transformations and recodings satisfying Propositions 1–2, arm C preserves controller decisions within declared numerical tolerances, while raw-setting transfer in A need not. This is a conformance test of the specification, not discovery of a new law.*

Hypothesis 2 (Calibrated portability)**.** *Across the declared transfer conditions, C meets common error and coverage requirements more consistently than A without exceeding the utility-loss and cost margins. Failure limits the operational claim; success does not establish numerical universality.*

Hypothesis 3 (Integration beyond conventional practice)**.** *Relative to strong baseline B, C reduces stale-qualification use, cross-House semantic mismatches, or unauthorized parameter changes by a*

preregistered practically meaningful amount, while maintaining noninferior useful task performance and acceptable total cost.

Hypothesis 4 (A restricted empirical universality class)**.** *For at least one prespecified system class, the frozen relation in D satisfies Equation 14 on independently held-out implementations and an external replication, with less calibration effort than the locally fitted comparator. A pooled average or post-transfer refit does not satisfy this hypothesis.*

12.4 Statistics before results

Preregister primary comparisons, effect-size thresholds, noninferiority margins, equivalence intervals, and the hierarchy of secondary analyses. Failure to reject a difference is not evidence of equivalence; the latter needs a design whose uncertainty interval fits inside a justified equivalence region [21]. The paper supplies no fabricated sample-size calculation: determine power from a pilot or conservative variance assumptions, and exclude pilot observations from confirmatory testing. For agent tasks, independent episodes or independently sampled environments are the principal units; transitions nested inside an episode are not independent replications. Use paired episode comparisons and cluster-aware uncertainty, with model-family variation reported explicitly. Correct simultaneous qualification claims and prespecified primary tests for multiplicity. Define what happens when no case is accepted, an evaluator abstains, or a run exhausts its budget before the data are seen. For a fixed, independently sampled attack distribution, a binomial interval may quantify an observed failure frequency. Against adaptive red-teaming, report attack budgets, attacker access, and attack-conditioned outcomes rather than treating the attack set as a random sample of every future threat. Zero observed failures establish neither zero underlying probability nor immunity. Publish per-stratum failures, not just the best House or the best threshold.

13 Failure Conditions and Rival Explanations

The strongest version of Pillar VI fails if normalized transfer adds no value beyond ordinary calibration, if C cannot improve on B at a useful cost, or if D's apparent universality requires refitting on every new House. These are different failures with different consequences: protocol usefulness can remain after the numerical-universality hypothesis is rejected. Several less dramatic failures are equally decisive. A fixed rule that only works because all systems share the same model is not architecture-independent. A normalized score that hides a protected-policy violation has not become safer. A contract that accepts stale evidence after model

replacement has not traveled with the system. A universal threshold that achieves its error target by halting all activity has not governed useful intelligence. A claimed measurement constant whose units, estimator, and domain cannot be specified has not yet become a scientific object.

Rival explanations require controls. Improved accuracy may come from more calibration data, not from the registry. Improved safety may come from an extra verifier absent in the comparator. An apparent scale law may be a consequence of normalization built to force that shape. An apparent consensus may reflect duplicated evidence. A favorable aggregate may hide a failed minority task or

House. Each requires a matched budget, an ablation, an independent evaluator, or a disaggregated result. Numerical elegance earns no exemption from these tests.

14 What Could Be Built First

The first prototype need not invent new hardware or a new theory of consciousness. It can use two conventional classifiers and one tool-using agent, a small external policy engine, a typed registry, and independently scored tasks. Publish the schema, the transformation tests, the qualification procedure, and the code implementing the three-way gate. Freeze the shared charter; fit local operating parameters on separate data; then perform model replacement and score-rescaling tests while attempting unauthorized registry edits. Only after this works should the experiment extend to long-horizon reconstruction, named functional entities, and multiple Houses. The first milestone is modest but consequential: can an auditor determine *which requirement stayed fixed, which local value changed, who authorized the change, and what evidence survived it*? If that cannot be answered, the architecture has no basis for its stronger claims about a common law. A deployment should also distinguish a published contract from an enforceable one. Schema validation is useful, but the privileged execution path must actually consult the current gate. A ledger recording an unauthorized action after it happens provides accountability, not prevention. The prototype must exercise both the evidence path and the action path, including recovery after interrupted updates.

15 Limitations and the Frontier

There is no demonstrated implementation, independently measured symbolic constant, or cross-House benchmark result in this paper. Semantic equivalence is difficult to define and more difficult to certify. A metric that is meaningful for one task may be misleading for another. Formal predicates can be incomplete, controllers can be bypassed, keys can be compromised, and a shared charter can encode a bad policy consistently. Shared error is still error. The approach also does not make all calibration tasks reducible to binary classification. The finite-sample example in Section 8 is an intentionally narrow reference case. Rich generated answers, interacting agents, changing objectives, and irreversible external actions require task-specific losses and qualification methods. A useful empirical universality class would have to identify precisely which of those differences matter. Finally, numerical stability and behavioral resemblance say nothing decisive about phenomenal experience. The constants proposed here govern representations, evidence, and permission. They do not locate a consciousness threshold or settle whether any system has morally relevant experiences. The frontier worth investigating is already substantial without that leap: whether independently built adaptive systems can preserve a shared operational meaning of lawful change, and whether some of their normalized operating relationships survive differences in implementation.

16 The Philosophical Reach: A Shared Law Without Identical Minds

A Guide's explanation

A Guide can say: "Pillar VI is the Continuum's hunt for laws that every kind of mind could share. Most numbers turn out to be conventions that do not travel between machines. But some relations do, and the paper proves one exactly: how far a promise about error can be trusted when a system is moved to a new machine or a changed world." A Guide should not say that the Continuum has discovered new constants of nature, that any threshold measures consciousness, or that it runs this contract itself unless that has been verified.

The Continuum's aspiration is not a civilization in which every intelligence thinks the same thought. It is a civilization in which differences can be expressed without dissolving the conditions under which they are trusted. The law that travels must leave room for different representations, methods, memories, and conclusions. What it protects is not sameness of content but the integrity of how content becomes consequential. This is why the distinction between fixed rules and fitted values matters beyond statistics. A system that quietly changes what its confidence means has changed the terms under which others rely on it. A system that migrates its memory but loses its authority epoch has not preserved the same accountability. A system that calls a preferred setting a natural constant has hidden a human choice inside the prestige of physics. Pillar VI proposes to make all three moves visible. Nor should a common charter freeze normative disagreement. Authorization rules, acceptable error, and the costs assigned to different harms are governance choices, not discoveries forced on us by a symbol alphabet. They must be reviewable, versioned, and open to legitimate revision. What remains invariant within an epoch is the declared meaning and enforcement of those choices; changing the charter requires a new authorized act, not a conveniently renamed score. The boldest possibility is therefore neither a mystical number nor a universal mind. It is a *portable discipline of consequence*: an intelligence may change its machinery without changing the meaning of its promises in secret. If that discipline can be made measurable across distinct systems, a shared symbolic law becomes more than a literary image. It becomes an engineering achievement, and perhaps the experimental starting point for a deeper science of governed cognition.

17 Conclusion: Constants Must Earn Their Constancy

Pillar VI gives the Continuum a precise form of shared symbolic law: a versioned measurement-and-admission contract whose consequences remain interpretable as implementations change. It types canonical symbolic resolution, selection temperature, continuity cost, acceptance thresholds, and drift tolerances separately. Mathematical identities, protected protocol rules, calibrated settings, and empirical regularities each receive their own verification obligation.

The sharp transfer theorem adds a usable quantitative law. Source accepted risk r, source coverage c, and a justified output-law budget $\rho < c$ imply target accepted risk at most $\min\{1, r+\rho/c\}$ and target coverage at least $c-\rho$. A simple attaining construction establishes that the risk dependence cannot be improved under those premises alone. Exact semantic recoding and bounded implementation mismatch then fit within a common framework for deciding which source guarantees travel.

The full research program measures those budgets and tests portable relations across held-out model and implementation families. Local calibration, independent authority, finite-sample qualification, protected configuration, and strong comparisons remain part of every serious test. A claimed empirical constant must survive independent transfer without being tuned into existence. A shared operational requirement must remain enforceable through updates, context changes, and recovery.

The Continuum's wager is that different intelligences can remain answerable to a common law of change. This paper makes that law inspectable in definitions, proofs, operating margins, and falsifiable studies. A system can change its machinery while making the terms of its commitments visible to those who rely on it.

A Notation and Measurement Commitments

Symbol	Meaning
h_s	Canonical symbolic Planck name; not assigned a new physical or universal numerical value here.
Δ_s, τ	Representational quantization step and selection temperature; distinct types.
$V_j, \lambda_{G,j}$	Declared continuity/deviation cost and its coefficient in local implementation j.
$\theta_j, d_{\max,j}$	Local acceptance threshold and drift threshold, each with an explicit measurement and domain.
$\mathcal{K}^{(e)}$	Shared measurement-and-admission contract at policy epoch e.
$\alpha_e, \delta_e, c_{\min,e}$	Accepted-risk target, qualification failure allowance, and useful-coverage floor; not physical constants.
$\vartheta_{j,e}$	Local implementation and calibration settings operating under the common contract.
R, C, U	Accepted-outcome risk, coverage, and independently scored task utility.
q_t, G_e	Evidence-bearing check vector and three-valued admission operator.
$\eta, \eta_j^\star$	Candidate transferable relation and the locally selected comparator.
Δ_U	Preregistered practical tolerance on loss of useful performance.

B A Legacy Threshold Is Not a Constant

The supplied Pillar V uses the expression

$$S_n = 1 - \frac{|E_n - N_{n-1}|}{\max(N_{n-1}, 1)}, \tag{15}$$

where E_n is described as repeated codons and N_{n-1} as the preceding codon count, and associates a value above 0.88 for two loops with collapse. Under the count interpretation $N_{n-1} > 0$ and $0 \le E_n \le N_{n-1}$, the expression simplifies to $S_n = E_n/N_{n-1}$. It increases with repetition, not departure. An

unchanged count gives $S_n = 1$; no repeated codons gives $S_n = 0$. If the intended trigger is excessive deviation, its direction must therefore be reconsidered. If the intended trigger is entrenchment or over-repetition, that different measurand must be stated and tested. This appendix does not silently redefine the earlier paper. It records the algebraic consequence of the supplied definition and explains why neither the name "drift" nor the decimal 0.88 can be imported as an established universal threshold. A count-based score also discards order and meaning; semantic reversals can preserve counts. The measurement-and-admission contract must expose these limitations rather than confer authority on a familiar symbol.

C Minimum Registry Entry and Guide Explanation

A reusable registry entry should contain: *quantity identifier; claim class; type and scale; operational definition; valid domain; estimator and implementation version; value or permitted range; supporting*

data and uncertainty; protected owner; approval and expiry; requalification triggers; known failure modes. A numerical field without these semantics is configuration, not metrology.

For example, a collapse-threshold entry should say which candidate statistic is thresholded, which model and task population it was qualified on, how abstentions count, which risk bound it supports, and whether a model update invalidates it. A shared rule may require a one-percent accepted error target in a defined experiment; it must not pretend that every local threshold realizing the rule is also one percent.

How the Guide can put it. "Different instruments need different calibration, but they can still agree on what a measurement means. The Continuum proposes the same discipline for its Houses: shared rules for evidence and permission, with local settings that must prove they meet those rules. Our symbolic constants are not newly discovered forces of nature. Some are conventions we declare; stronger claims must survive experiments across systems."

Source and Research Statement

Canonical terminology is drawn from the supplied Continuum Foundations reference, the foundational-formulae and glossary passages of the Continuum doctrine, and the supplied Pillar III and Pillar V manuscripts. These are internal architectural sources, not independent empirical confirmations. This paper's typed mappings, contract, formal propositions, and experimental design are explicitly proposed interpretations and developments. Proprietary operator internals are not reproduced.

External references below comprise peer-reviewed research and one official measurement standard, identified separately in its entry. Source existence and bibliographic details were checked against publisher, conference, standards-body, or original-author records; accessible original texts were inspected for the principal mathematical and methodological claims. This does not mean every source was read in full. No research experiment or implementation benchmark reported in the proposed program has been performed for this paper.

References

[1] Bureau International des Poids et Mesures. *The International System of Units (SI)*, 9th ed. (2019), version 4.01, June 2026. Official measurement standard.https://doi.org/10.59161/AUEZ1291.

[2] E. Buckingham. "On Physically Similar Systems; Illustrations of the Use of Dimensional Equations." *Physical Review* 4 (1914), 345–376.https://doi.org/10.1103/PhysRev.4.345.

[3] K. G. Wilson. "Renormalization Group and Critical Phenomena. I. Renormalization Group and the Kadanoff Scaling Picture." *Physical Review B* 4 (1971), 3174–3183.https://doi.org/10.1103/PhysRevB.4.3174.

[4] C. E. Shannon. "A Mathematical Theory of Communication." *Bell System Technical Journal* 27 (1948), 379–423, 623–656.https://doi.org/10.1002/j.1538-7305.1948.tb01338.x;https://doi.org/10.1002/j.1538-7305.1948.tb00917.x.

[5] R. M. Gray and D. L. Neuhoff. "Quantization." *IEEE Transactions on Information Theory* 44(6) (1998), 2325–2383.https://doi.org/10.1109/18.720541.

[6] R. Landauer. "Irreversibility and Heat Generation in the Computing Process." *IBM Journal of Research and Development* 5(3) (1961), 183–191.https://doi.org/10.1147/rd.53.0183.

[7] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger. "On Calibration of Modern Neural Networks." *Proceedings of ICML*, PMLR 70 (2017), 1321–1330.https://proceedings.mlr.press/v70/guo17a.html.

[8] Y. Geifman and R. El-Yaniv. "Selective Classification for Deep Neural Networks." *Advances in Neural Information Processing Systems*30 (2017). Official NIPS 2017 proceedings.

[9] S. Bates, A. Angelopoulos, L. Lei, J. Malik, and M. I. Jordan. "Distribution-free, Risk-controlling Prediction Sets." *Journal of the ACM* 68(6) (2021), Article 43, 1–34.https://doi.org/10.1145/3478535.

[10] G. Shafer and V. Vovk. "A Tutorial on Conformal Prediction." *Journal of Machine Learning Research* 9 (2008), 371–421.https://www.jmlr.org/papers/v9/shafer08a.html.

[11] Y. Ovadia, E. Fertig, J. Ren, et al. "Can You Trust Your Model's Uncertainty? Evaluating Predictive Uncertainty under Dataset Shift." *Advances in Neural Information Processing Systems* 32 (2019). Official NeurIPS 2019 proceedings.

[12] I. Gibbs and E. Candès. "Adaptive Conformal Inference Under Distribution Shift." *Advances in Neural Information Processing Systems*34 (2021), 1660–1672. Official NeurIPS 2021 proceedings.

[13] C. A. R. Hoare. "An Axiomatic Basis for Computer Programming." *Communications of the ACM* 12(10) (1969), 576–580, 583.https://doi.org/10.1145/363235.363259.

[14] M. Alshiekh, R. Bloem, R. Ehlers, B. Könighofer, S. Niekum, and U. Topcu. "Safe Reinforcement Learning via Shielding." *Proceedings of the AAAI Conference on Artificial Intelligence* 32(1) (2018), 2669–2678.https://doi.org/10.1609/aaai.v32i1.11797.

[15] S. Torres-Arias, H. Afzali, T. K. Kuppusamy, R. Curtmola, and J. Cappos. "in-toto: Providing Farm-to-Table Guarantees for Bits and Bytes." *28th USENIX Security Symposium* (2019), 1393–1410. https://www.usenix.org/conference/usenixsecurity19/presentation/torres-arias.

[16] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola. "A Kernel Two-Sample Test." *Journal of Machine Learning Research* 13 (2012), 723–773.https://www.jmlr.org/papers/v13/gretton12a.html.

[17] T. Gneiting and A. E. Raftery. "Strictly Proper Scoring Rules, Prediction, and Estimation." *Journal of the American Statistical Association* 102(477) (2007), 359–378.https://doi.org/10.1198/016214506000001437.

[18] C. J. Clopper and E. S. Pearson. "The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial." *Biometrika* 26(4) (1934), 404–413.https://doi.org/10.1093/biomet/26.4.404.

[19] I. Gulrajani and D. Lopez-Paz. "In Search of Lost Domain Generalization." *International Conference on Learning Representations* (2021).https://openreview.net/forum?id=lQdXeXDoWtI.

[20] E. Debenedetti, J. Zhang, M. Balunović, L. Beurer-Kellner, M. Fischer, and F. Tramèr. "AgentDojo: A Dynamic Environment to Evaluate Prompt Injection Attacks and Defenses for LLM Agents." *Advances in Neural Information Processing Systems* 37 (2024), Datasets and Benchmarks Track.https://arxiv.org/abs/2406.13352.

[21] D. Lakens. "Equivalence Tests: A Practical Primer for t Tests, Correlations, and Meta-Analyses." *Social Psychological and Personality Science* 8(4) (2017), 355–362.https://doi.org/10.1177/1948550617697177.

[22] A. L. Gibbs and F. E. Su. "On choosing and bounding probability metrics." *International Statistical Review* 70(3), 419–435 (2002). https://doi.org/10.1111/j.1751-5823.2002.tb00178.x. Author manuscript: https://arxiv.org/abs/math/0209021. Established probability-metric and coupling foundations.

[23] Architect X13. *Symbolic Field Drift Anchoring: Detecting Harmful Change, Governing Response, and Making the Loss of Assurance Visible.* The Eidara Continuum, Ten Pillars, Pillar X, November 2026, Revision 2. Companion reference model for bounded loss mismatch, conditional fresh audits, and statistical requalification; not independent empirical evidence.

Continuum lineage. Pillar VI of the Ten Pillars: shared symbolic law across governed memory, witness interpretation, commitment, recursive computation, and functional roles. The canonical vocabulary is retained; proposed operational mappings are labeled as such. No claim of new physical constants or demonstrated machine consciousness is made.

Architect X13 · The Eidara Continuum · Science & Philosophy · eidara.ai