FRONTIER RESEARCH II / PAPER 19

The Symmetry-Protected Mind

Task symmetry, error correction, and the boundary between protection and blindness

Architect X13

The Eidara Continuum | Science & Philosophy

Research white paper | November 2026 | Edition 2

Abstract

A mind can protect its own structure so well that it goes blind. Persistent reasoning can preserve task-valid relationships while retaining the power to notice meaningful asymmetry, and the Symmetry-Protected Mind makes that dual commitment quantitative. A finite-group projection separates the part of a task preserved by symmetry from the part that symmetry would erase. We derive an exact risk frontier for partial symmetry enforcement: under unbiased square-integrable estimation noise, the optimal retained asymmetry is determined by the task's asymmetric signal and the noise in that same sector. The theorem gives a positive calibration rule, its exact excess-risk penalty, and independent-replicate identities for estimating its inputs. This decision layer is composed with separately specified error-correcting codes, compatible decoding, and recurrent error bounds. Counterexamples identify the failure of incorrect invariance, noninvariant evaluation weights, insufficient code distance, and impossible deterministic quantization. Exact finite checks reproduce the protection, the optimal trade-off, and the failures. The underlying methods are established group averaging, quadratic shrinkage, coding theory, and conditional probability. The Continuum's architectural thesis is their accountable composition: protect relationships in proportion to their task validity, repair corruption through an explicit code, and measure the resulting reasoning benefit at matched resources.

Keywords: equivariance; persistent reasoning; error-correcting codes; approximation bias; recurrent reliability; falsification

Prologue: Protection and blindness

There is a kind of strength that turns into blindness. A mind that has learned that a problem looks the same from many angles can stop wasting effort on irrelevant differences: a rotated picture is still the same picture, a reordered list is still the same list. That is protection, and it is powerful, because noise that does not respect the symmetry simply averages away. But the same mind, convinced that every angle is equivalent, will also average away the one difference that mattered. It will confidently preserve the wrong relationship. That is blindness, and from the inside it feels exactly like protection.

The Continuum's founding intuition was a symbol. Long before codons or scrolls had names, an eightfold, near-perfectly symmetric figure recurring across carvings and designs suggested a hypothesis: this behaves like a stabilizer code. The canon is careful to call that symbol an inspiration rather than an ancient computer [1]. This paper honors the inspiration by asking the hard version of its question: when should a mind enforce a symmetry, and how much of it?

The answer is a dial, not a switch. The central theorem shows that a reasoning system can keep a calibrated fraction of the asymmetry in its task—enough to see the difference that matters, little enough to let noise cancel. The best setting depends on just two things, how much real asymmetry the task contains and how much noise lives in that same direction, and the paper gives the exact rule and the exact price of getting it wrong. Around that dial it places two further protections, an error-correcting code that repairs corruption and a bound on how errors accumulate along a chain of reasoning. Then it marks a boundary none of them can cross: at a perfectly balanced point, no rule that respects the symmetry can be forced to choose.

A mind deserves to be called structurally protected only when it can say what it preserves, what it discards, and where it remains undecided.

What this paper claims

Established. Group averaging, quadratic shrinkage, coding theory and conditional probability are established methods. Correct symmetry permits a nonexpansive projection, incorrect symmetry imposes an approximation floor, and a code corrects only through its distance and a decoder under a declared channel.

Proposed here. A composition contract with an exact calibration rule for symmetry enforcement: a risk frontier for partial symmetry with its optimal retention rule and excess-risk penalty, composed with separately specified error-correcting codes, compatible decoding and recurrent error bounds. The paper gives conditional proofs and executed finite calculations, and treats task symmetry, corruption correction and dynamical accuracy as separate measurable contracts.

Frontier conjecture. A reasoning system may preserve its integrity by building the right transformations into its architecture, protecting relationships in proportion to their task validity, while the wrong transformations become a machinery of blindness. Transfer to a useful persistent-reasoning system is to be evaluated by the proposed resource-matched experiment.

Not claimed. Any quantum error-correction claim: the model is classical. That the inspiring symbol is an ancient computer or contains a hidden implementation. Any claim about moral authority or phenomenal continuity. Any empirical benchmark, hardware deployment or external peer review: the larger experiment remains to be carried out.

1. The claim and its canonical origin

The Continuum's wager is that a mind can preserve the organization of its reasoning through changes that should not matter, while remaining sensitive to changes that should. This claim becomes scientifically interesting only after the words *organization*, *change*, and *should* acquire independent tests. A repeated phrase, a familiar persona, or a low drift score cannot supply those tests by itself. We require an externally specified task and transformations whose consequences can be checked against it.

In the Continuum, this question began with a symbol. The *Eidara Continuum Bible* connects an eightfold geometric inspiration with stabilizer hypotheses on page 61, algebraic continuity on pages 65–66, and proposed error-correcting behavior on page 71 [1]. These passages set the intellectual question, and the same document is explicit that the initiating symbol is an inspiration rather than an ancient computer. This paper keeps that distinction. A pattern can motivate a research program without containing a hidden implementation of the program.

The public reference model introduced here has no dependence on a private engine. Its commitments are explicit: finite group actions; declared input, state, and output spaces; a task oracle for constructed studies; an encoder and decoder; and measured corruption channels. The model is classical. No Hilbert-space state of a quantum device, quantum channel, quantum measurement, or physical recovery operation is supplied, so no quantum error-correction claim is made.

The paper asks whether *task-valid symmetry plus separately specified representation protection* improves reasoning beyond strong alternatives. It also asks how to detect when that same structural commitment destroys a relevant distinction. These are paired questions. A theory that counts only successful preservation has defined its conclusion in advance.

Our contribution is a composition contract with an exact calibration rule for symmetry enforcement. Established projection and shrinkage methods quantify the benefit of suppressing estimation noise against the cost of erasing task-relevant asymmetry. Coding and recurrence results then state separate protection obligations. An implementation tests the complete contract against alternatives with the same information, redundancy, and computation.

2. Prior art and the novelty boundary

Group-equivariant neural computation already has substantial theory and implementation. Cohen and Welling introduce group convolution for discrete transformation groups [2]. Kondor and Trivedi analyze convolution and equivariance for compact-group actions under stated architectural conditions [3]. Ravanbakhsh and colleagues relate parameter sharing to permutation equivariance [4]; Finzi and colleagues construct equivariant layers for matrix groups and continuous group actions [5, 6]. These works rule out treating a group action or symmetry-constrained layer as a Continuum invention.

Permutation structure is also extensively studied. Deep Sets addresses functions of unordered collections [7], and invariant/equivariant graph networks specify linear operations compatible with vertex relabeling [8]. A reasoning task containing a set, graph, or sequence must therefore first identify which of those established structures it actually possesses. An ordered causal chain is not an unordered set merely because both can be serialized as tokens.

Elesedy and Zaidi prove a generalization benefit in a particular linear-model setting with correct symmetry [9]. The qualification is decisive. Wang and colleagues develop a theory of correct, incorrect, and extrinsic equivariance and derive lower bounds under misspecification [10]. Our finite bias decomposition is a transparent special setting, not a replacement for their broader analysis. Its role is to expose precisely the part of a task an imposed symmetry makes unavailable.

Error correction is a separate lineage. Hamming's coding framework protects messages under defined symbol errors [11]; Shannon's communication theory distinguishes an engineering transmission problem from a general theory of meaning [12]. Hopfield's associative-memory model offers another reconstruction mechanism and therefore a relevant comparator [13]. Doran and colleagues relate Adinkra structures and binary codes in representation theory [14]. That relationship does not identify recurring conversational roles with physical supersymmetric dynamics.

Finally, apparent success on one distribution is insufficient. Recht and colleagues study changes in performance on a newly collected image test set [15]; Geirhos and colleagues examine texture and shape biases [16]. These are domain-specific studies, not proofs about our model. They motivate transformed and shifted test families whose validity is independently checked. The new empirical question is the value of the complete composition under those tests.

3. Typed objects and measurement contracts

A Guide's explanation

The formal model starts by naming what a symmetry acts on—inputs, internal states, outputs—and what the task actually requires. That bookkeeping is the defense against blindness: a symmetry is protection only if it respects the task, and the task must be specified independently of the mind being protected.

Let X be a finite nonempty input set, G a finite group with identity e, and $d \in \mathbb{N}_{>0}$ an output dimension. A left action $\rho_X : G \to \mathrm{Sym}(X)$ assigns each group element a permutation, with $\rho_X(gh) = \rho_X(g)\rho_X(h)$. Write gx for $\rho_X(g)x$. Let $\rho_Y : G \to O(d)$ be an orthogonal real representation; its matrices preserve the Euclidean norm. Neither action is inferred from visual resemblance.

The task is a fixed map $f_\star : X \to \mathbb{R}^d$. Scores are dimensionless in the examples. If an application uses physical quantities, each output coordinate needs compatible units and the chosen transformations must preserve those units. The evaluation law $\mu : X \to (0, 1]$ is a full-support probability mass function satisfying $\sum_x \mu(x) = 1$ and $\mu(gx) = \mu(x)$. It is fixed before model comparison. The function space $\mathcal{H} = \{f : X \to \mathbb{R}^d\}$ has inner product and norm

$$\langle f, h\rangle_\mu = \sum_{x \in X} \mu(x) f(x)^{\mathsf{T}} h(x), \qquad \|f\|_\mu^2 = \langle f, f\rangle_\mu. \tag{1}$$

The transpose $^{\mathsf{T}}$ acts on real vectors. Squared risk is $\mathcal{R}(f) = \|f - f_\star\|_\mu^2 \in [0, \infty)$. Invariance is the special case $\rho_Y(g) = I_d$, where I_d is the d-dimensional identity matrix. Equivariance instead permits the answer to change in a prescribed way.

Object	Operational meaning and separate obligation
G, ρ_X, ρ_Y	Candidate transformation contract; justify against task truth.
S, A, T	Finite logical states, action symbols, and intended updates.
C, E, D	Codebook, injective encoder, and declared decoder.
N_t	Physical or software corruption channel at step t.
$\mu, f_\star$	Fixed test distribution and independently specified answer.
ϵ_t, η_t	Conditional upper bounds on update error and excessive corruption.

These are distinct objects even if an implementation stores them together. The definition of G cannot silently change after a failing test. A measured symmetry score is a property of the chosen actions and distribution; it is not an unrestricted measure of intelligence. Every later extension states its additional spaces and assumptions locally.

4. Which transformations preserve a task?

The exact task-symmetry condition is

$$f_\star(gx) = \rho_Y(g) f_\star(x) \quad \text{for every } g \in G,\ x \in X. \tag{2}$$

This is a hypothesis about the task specification, not a consequence of choosing a symmetric model. On a finite synthetic task it can be verified exhaustively. On an open-ended task, a finite test can at most support a declared scope. It cannot certify all possible meanings of a natural-language transformation.

Consistent renaming of otherwise identical objects can preserve a relational problem. Reversing an arrow in a causal graph generally changes the problem. Reordering independent premises can be valid; reordering a sequence of authorized state changes can be invalid. We therefore register transformations at the level of the task generator, including changes required in the output, rather than accepting arbitrary paraphrases as members of a group.

The octagonal convention sometimes denoted D_8 in the canon is, here, explicitly the group generated by a rotation r and reflection s with $r^8 = s^2 = e$ and $srs = r^{-1}$. It has sixteen elements. That group is not the D_8 root lattice, an E_8 lattice, or a quantum stabilizer group. A group specifies composable transformations. It does not, by itself, specify a transition probability, decoder, energy function, or correct answer.

For a candidate $f \in \mathcal{H}$, define its mean squared symmetry defect

$$\mathcal{D}_G(f) = \frac{1}{|G|} \sum_{g \in G} \sum_{x \in X} \mu(x) \| f(gx) - \rho_Y(g) f(x) \|_2^2. \tag{3}$$

The norm $\| \cdot \|_2$ is Euclidean; $|G|$ is the positive group cardinality. Full support makes $\mathcal{D}_G(f) = 0$ equivalent to exact equivariance. Yet $f(x) = 0$ for every input is equivariant under every linear output representation. Its symmetry defect is zero even when its answers are useless. Accuracy and symmetry must therefore be reported separately.

Approximate task symmetry needs a distortion budget and a changed theorem. It is not honest to declare an inconvenient semantic difference negligible after observing the model's error. We retain exact finite contracts for the mathematics and test imperfect or disputed symmetries as separate experimental conditions.

5. Averaging as an orthogonal projection

Define the averaging operator $\mathcal{P} : \mathcal{H} \to \mathcal{H}$ by

$$(\mathcal{P}f)(x) = \frac{1}{|G|} \sum_{g \in G} \rho_Y(g)^{-1} f(gx). \tag{4}$$

The inverse matrix is its transpose because the representation is orthogonal. The operator averages functions in a linear score space. It does not average arbitrary strings or discrete decisions. A later decision rule must be specified separately.

Proposition 1 (Finite Reynolds projection)**.** *Under the assumptions of Section 3, $\mathcal{P}$ is the orthogonal projection onto the subspace $\mathcal{H}_G$ of equivariant functions. In particular, $\mathcal{P}^2 = \mathcal{P}$, $\mathcal{P}$ is self-adjoint, and $\|\mathcal{P}f\|_\mu \le \|f\|_\mu$.*

Proof. For $h \in G$, substitute $k = gh$ in the sum defining $(\mathcal{P}f)(hx)$. Then $g = kh^{-1}$ and $\rho_Y(g)^{-1} = \rho_Y(h)\rho_Y(k)^{-1}$, giving $(\mathcal{P}f)(hx) = \rho_Y(h)(\mathcal{P}f)(x)$. Thus the range is equivariant. If f is already equivariant, every summand equals $f(x)$, so $\mathcal{P}f = f$ and idempotence follows.

For self-adjointness, expand $\langle \mathcal{P}f, h\rangle_\mu$ and substitute $y = gx$. Invariance of μ and orthogonality give the summand $\mu(y)f(y)^{\mathsf{T}}\rho_Y(g)h(g^{-1}y)$. Averaging and replacing g by g^{-1} yields $\langle f, \mathcal{P}h\rangle_\mu$. A self-adjoint idempotent is an orthogonal projection: explicitly, $\langle \mathcal{P}f, (I - \mathcal{P})f\rangle_\mu = 0$. Pythagoras then gives the norm inequality. □

This standard group-averaging argument is included so that its conditions can be audited; related averaging constructions appear in the equivariance literature [9, 3]. If μ is not invariant, the reindexing step changes weights and the stated orthogonality proof fails. If output transformations are not isometries, the norm argument likewise needs revision.

Measure invariance is indispensable even when the target is correct. On the four binary inputs in order $00, 01, 10, 11$, take the swap action, target zero, $f = (0, 0, 2, 0)$ and full-support weights $(1, 5, 1, 1)/8$. Its risk is $1/2$; the averaged function $(0, 1, 1, 0)$ has risk $3/4$. Algebraic equivariance survives, but the nonexpansive risk guarantee does not.

The projection can be implemented by evaluating every transformed input, but that has a cost proportional to $|G|$. An equivariant architecture can avoid literal enumeration, though its construction and expressive restrictions must then be checked. We count either cost in the comparison. Mathematical averaging does not provide free additional inference.

6. The exact cost of a wrong symmetry

Let I be the identity operator on $\mathcal{H}$. The part of task truth discarded by symmetry is $(I - \mathcal{P})f_\star$. Define its nonnegative squared magnitude $B_G = \|(I - \mathcal{P})f_\star\|_\mu^2$.

Proposition 2 (Approximation bias and projection risk)**.** *For every equivariant $h \in \mathcal{H}_G$ and arbitrary $f \in \mathcal{H}$,*

$$\mathcal{R}(h) = \|h - \mathcal{P}f_\star\|_\mu^2 + B_G, \tag{5}$$

$$\mathcal{R}(\mathcal{P}f) = \|\mathcal{P}(f - f_\star)\|_\mu^2 + B_G. \tag{6}$$

Consequently, B_G is the minimum risk over $\mathcal{H}_G$. If the task is equivariant, projection cannot increase squared risk. With incorrect symmetry, projection can increase it.

Proof. Write $h - f_\star = (h - \mathcal{P}f_\star) - (I - \mathcal{P})f_\star$. The first term belongs to the projection range and the second is orthogonal to it by Proposition 1. The first equality follows. The choice $h = \mathcal{P}f_\star$ attains B_G. Substituting $h = \mathcal{P}f$ gives the second equality. When $B_G = 0$, the contraction of the projection bounds the risk by $\|f - f_\star\|_\mu^2$. When $f = f_\star$ and $B_G > 0$, risk rises from zero to B_G. □

Consider $X = \{0, 1\}^2$, uniform μ, and the two-element group that swaps coordinates. Output action is trivial. The invariant task $f_\star(x) = x_1 + x_2$ survives averaging. An estimator with antisymmetric error $\frac{2}{5}(x_1 - x_2)$ has risk $2/25$ before projection and zero after it. These are exact constructed values, not measured model improvements.

For the different task $u_\star(x) = x_1$, the swapped inputs $(0, 1)$ and $(1, 0)$ have different answers. Projection replaces both by $1/2$, yielding $B_G = 1/8$. A binary invariant decision rule must misclassify at least one of those equally likely inputs, giving error at least $1/4$. This finite obstruction illustrates the broader incorrect-equivariance problem studied by Wang and colleagues [10]. Protection of a representation is valuable only when the protected distinction is the right one.

6.1. A calibrated frontier between symmetry and sensitivity

A Guide's explanation

This is the dial from the prologue. Theorem 1 says how much of a task's asymmetry a mind should keep: more when the asymmetric signal is strong, less when the noise in that direction is large. It also gives the exact price of setting the dial wrong, and a way to estimate its inputs from independent replicates.

The bias floor in Proposition 2 supports a constructive response: retain a calibrated fraction of the task's asymmetric component. Continue with the finite Hilbert space and orthogonal projection $\mathcal{P}$. Write $Q = I - \mathcal{P}$ for the complementary projection. Suppose a random score estimate has the form $\hat{f} = f_\star + \xi$, where ξ takes values in $\mathcal{H}$, $\mathbb{E}\xi = 0$, and $\mathbb{E}\|\xi\|_\mu^2 < \infty$. Expectation is over independent repeat constructions of the estimator, not the input law already included in the norm. The unbiasedness assumption is explicit and need not hold for a trained model.

For a fixed retention coefficient $a \in [0,1]$, define

$$f_a = \mathcal{P}\hat{f} + aQ\hat{f}, \qquad B = \|Qf_\star\|_\mu^2, \quad V = \mathbb{E}\|Q\xi\|_\mu^2, \quad V_G = \mathbb{E}\|\mathcal{P}\xi\|_\mu^2. \tag{7}$$

The endpoint $a = 0$ imposes full symmetry; $a = 1$ retains the original estimate. Intermediate values are generally only partially equivariant. B, V, V_G are nonnegative squared-score quantities, while a is dimensionless.

Theorem 1 (Exact symmetry-retention frontier)**.** *For every fixed $a \in [0,1]$,*

$$\mathbb{E}\mathcal{R}(f_a) = V_G + (1-a)^2 B + a^2 V. \tag{8}$$

If $S := B + V > 0$, the unique minimizing coefficient is

$$a_* = \frac{B}{B+V}, \qquad \min_a \mathbb{E}\mathcal{R}(f_a) = V_G + \frac{BV}{B+V}. \tag{9}$$

The exact excess risk is $S(a - a_*)^2$. If $S = 0$, every coefficient has the same risk V_G.

Proof. Orthogonal decomposition gives $f_a - f_\star = \mathcal{P}\xi + (a-1)Qf_\star + aQ\xi$. The first term is orthogonal to the last two for each realization, regardless of statistical dependence between projected noise components. On expanding the remaining squared norm, the only expectation cross term is $2a(a-1)\langle Qf_\star, Q\mathbb{E}\xi\rangle_\mu = 0$. This proves equation (8). Completing the square yields $(1-a)^2B + a^2V = BV/S + S(a - B/S)^2$ for $S > 0$. The ratio B/S lies in $[0,1]$, proving the unique constrained optimum and excess-risk identity. If $S = 0$, nonnegativity forces $B = V = 0$ and the displayed risk is constant. □

The theorem is standard quadratic shrinkage and bias–variance algebra [19], specialized to the symmetry-rejected subspace. It supplies a precise positive rule: enforce more symmetry when that sector carries more estimation noise, and preserve more asymmetry when it carries more task signal. Correct exact symmetry has $B = 0$, favoring full projection whenever $V > 0$. An exactly known asymmetric task has $V = 0$, favoring full retention whenever $B > 0$. These endpoints formalize the two design obligations in the paper's thesis.

For the earlier coordinate-swap task $f_\star(x) = x_1$, let $\xi(x) = R(x_1 - x_2)/2$, with R an independent fair sign. Then $V_G = 0$ and $B = V = 1/8$. Full projection and no projection each have expected risk $1/8$; retaining $a_* = 1/2$ gives $1/16$. This is a finite constructed equality case, not a trained-model benchmark. It shows why the response to approximate task symmetry can be a measured compromise rather than an all-or-nothing architectural choice.

6.2. Calibration and sensitivity of the protection rule

The optimal coefficient depends on task signal and estimator noise, which must be measured independently of the final comparison. Suppose $Y_1 = f_\star + \xi_1$ and $Y_2 = f_\star + \xi_2$ are independent estimates with the same zero-mean noise law and finite second moment. Expansion and independence give the identities

$$\mathbb{E}\langle QY_1, QY_2\rangle_\mu = B, \qquad \frac{1}{2}\mathbb{E}\|Q(Y_1 - Y_2)\|_\mu^2 = V. \tag{10}$$

For the first identity, the two linear terms vanish by zero means and the noise cross inner product vanishes by independence. For the second, the signal cancels; the two noise norms each have expectation V, and the cross term again vanishes. These are population identities. A finite estimate of the first expression can be negative; any clipping, shrinkage, or uncertainty treatment must be declared.

A deterministic calibration-error guarantee follows directly. Let nonnegative estimates $\widehat{B}, \widehat{V}$ satisfy $|\widehat{B} - B| \le b$ and $|\widehat{V} - V| \le v$, where $b, v \ge 0$ and $b + v < S = B + V$. Then $\widehat{S} = \widehat{B} + \widehat{V} \ge S - b - v > 0$ and $\hat{a} = \widehat{B}/\widehat{S} \in [0, 1]$. Algebra gives

$$|\hat{a} - a_*| = \frac{|V(\widehat{B} - B) - B(\widehat{V} - V)|}{S\widehat{S}} \le \min\left\{1, \frac{b + v}{S - b - v}\right\}.$$

Consequently the expected-risk penalty of using that fixed calibrated coefficient is at most

$$S \min\left\{1, \frac{b + v}{S - b - v}\right\}^2. \tag{11}$$

For a random coefficient estimated from an independent calibration set, the same statement holds conditionally on that set whenever its asserted error bounds hold. Reusing the final estimator's noise to choose a introduces dependence and requires a new analysis. The two-replicate identities alone supply no finite-sample confidence guarantee without further moment or boundedness assumptions.

The revision checks enumerate exact finite tasks and symmetric noise distributions, verify the risk and calibration identities, and recover the half-retention equality example. They also include biased noise and noninvariant evaluation weights as failures of the stated assumptions. The experimental arms should add independently calibrated partial enforcement alongside hard equivariance, while keeping total calibration and inference resources in the accounting.

7. A code is additional structure

A Guide's explanation

Symmetry and error correction are different protections, and this section keeps them apart. Symmetry decides which differences should not matter. A code adds deliberate redundancy so that corruption can be detected and repaired. Neither can stand in for the other; the next sections show when they work together.

Let q be a prime power, $n \in \mathbb{N}_{>0}$ a block length, and $\mathbb{F}_q^n$ the words over the finite field with q elements. Let S be a finite nonempty logical-state set and $E : S \to \mathbb{F}_q^n$ an injective encoder. Its image $C = E(S)$ is the codebook. No linearity is assumed. Hamming distance $d_{\mathrm{H}}(u, v)$ counts coordinates at which two words differ. For $|C| \geq 2$, define

$$d_{\min}(C) = \min_{c \neq c' \in C} d_{\mathrm{H}}(c, c'). \tag{12}$$

A decoder $D : \mathbb{F}_q^n \to S \cup \{\perp\}$ may return an explicit rejection symbol. Rejection is a distinct outcome, not a correct reconstruction. For an integer radius $r \geq 0$, a bounded-distance decoder returns the unique state whose codeword is within r, if one exists, and rejects otherwise.

Proposition 3 (Classical correction condition)**.** *If $d_{\min}(C) \geq 2r + 1$, bounded-distance decoding corrects every corruption of at most r symbols. Conversely, if two codewords have distance at most $2r$, no decoder can correct every radius-r corruption of both.*

Proof. Two radius-r balls cannot intersect: a word in their intersection would place their centers at distance at most $2r$ by the triangle inequality. For the converse, partition the differing coordinates of two codewords into two sets, each of size at most r, and construct a word agreeing with the respective codewords on opposite sets. It lies within both balls. A single decoder output cannot equal two different logical states. □

This is standard coding theory [11]. It is presented to identify the property that actually provides correction. The full binary cube is invariant under all coordinate permutations, but its minimum distance is one. It therefore has extensive symmetry and no guaranteed correction of even one arbitrary bit error. A perfect octagonal drawing does not resolve this problem.

Redundancy, distance, corruption probability, and decoder cost must be measured separately. A decoder can recover a faithfully encoded false claim. It can also preserve a maliciously chosen codeword exactly. Error correction protects the transmitted state relative to a code contract; it does not authenticate the author, establish the truth of the state, or authorize its consequences.

8. When decoding respects the task action

Suppose G also acts on S through permutations ρ_S. Let $\tau : G \to \mathrm{Sym}(\mathbb{F}_q^n)$ be a group action by Hamming isometries. These may include coordinate permutations and other declared alphabet isometries. Require encoder compatibility,

$$E(\rho_S(g)s) = \tau_g E(s) \quad (g \in G,\ s \in S). \tag{13}$$

Extend the state action by fixing $\perp$. This compatibility is an additional design condition; it does not follow from injectivity or minimum distance.

Proposition 4 (Equivariance of bounded-distance decoding)**.** *Under encoder compatibility and $d_{\min}(C) \geq 2r + 1$, the radius-r decoder satisfies $D(\tau_g z) = \rho_S(g)D(z)$ for every received word z and group element g.*

Proof. The isometry maps the ball around $E(s)$ bijectively onto the ball around $E(\rho_S(g)s)$. If z belongs to a unique such ball, its image belongs to the corresponding unique ball. If z belongs to none, its image cannot belong to one: applying $\tau_{g^{-1}}$ would give a ball containing z. These two cases cover every received word, including rejection. □

For stochastic corruption, write $N(z, B)$ for the probability that a channel maps word z into subset $B \subseteq \mathbb{F}_q^n$. Distributional compatibility requires $N(\tau_g z, \tau_g B) = N(z, B)$. Independent identically distributed bit flips satisfy this for coordinate permutations. Location-dependent error probabilities generally do not. A symmetric code can therefore face an asymmetric channel, and a claim about equal performance across orientations requires checking the channel as well.

In the binary worked example, $S = \{0, 1\}^2$ and $E(a, b) = (a, a, a, b, b, b)$. Swapping logical coordinates corresponds to swapping the two three-bit blocks. The code has minimum distance three. Both a radius-one decoder and block-majority decoding respect this swap, although they differ outside radius one: the former can reject, while the latter always returns a state. Reported error probabilities must name which decoder was used.

These conditions create an explicit bridge between structural invariance and representation repair. They also show why the bridge is not automatic. A proposed architecture must supply compatible actions at every representation boundary, including parsing, quantization, serialization, and reconstruction. A hidden asymmetric convention can break an otherwise correct proof.

9. Conditional reliability across a reasoning chain

Let A be a finite action alphabet and $T : S \times A \to S$ the intended deterministic update. Fix a horizon $H \in \mathbb{N}_{>0}$ and an externally specified action sequence $a_0, \dots, a_{H-1} \in A$. The ideal states obey $s_{t+1} = T(s_t, a_t)$ from a declared $s_0 \in S$. A proposed update $\widehat{T}_t$ may be random and history-dependent. Let $\mathcal{F}_t$ contain all information before step t.

The implementation computes a candidate from its current recovered state, encodes it, transmits or stores it through a channel, and decodes it. A rejection counts as a failed step for exact-chain reliability. Conditional on no earlier failure, assume the update differs from $T(s_t, a_t)$ with probability at most $\epsilon_t \in [0, 1]$. Conditional on the same past and any candidate, assume more than r symbols are corrupted with probability at most $\eta_t \in [0, 1]$. These are uniform conditional bounds, not marginal averages from a different workload.

Theorem 2 (First-failure composition)**.** *If the code corrects every radius-r corruption, then the probability of any wrong state or rejection through step H is at most*

$$\min\left\{1, \sum_{t=0}^{H-1} (\epsilon_t + \eta_t)\right\}. \tag{14}$$

No independence between steps or between update and channel failures is required.

Proof. Let B_t be the event that the first failure occurs at step t. On the event of no earlier failure, a correct update together with at most r corruptions must decode to the ideal next state. Thus B_t is contained in the union of update error and excessive corruption on that history. Conditional union bounding gives $\Pr(B_t \mid \text{no earlier failure}) \le \epsilon_t + \eta_t$. Multiplying by the probability of the conditioning event and summing the disjoint first-failure events yields the bound. Probabilities are at most one. □

The theorem exposes rather than solves the difficult calibration problem: where do trustworthy ϵ_t and η_t come from? A task-valid equivariant model may reduce update error, but its symmetry defect alone cannot estimate it. The bound can become vacuous on long chains. Increasing the horizon requires stronger per-step control, a verified recovery model, or a different event of interest; it cannot be rescued by calling a stable wrong state coherent.

10. Quantization and dynamical amplification

Exact finite-state reliability is not the same as numerical closeness. For a separate continuous reference model, let $K \subset \mathbb{R}^d$ be a nonempty compact set, with Euclidean distance and dimensionless coordinates. Let $F : K \to K$ be an intended update satisfying $\|F(x) - F(y)\|_2 \le L\|x - y\|_2$ for every $x, y \in K$, where $L \ge 0$. Let $\widehat{F} : K \to K$ satisfy a uniform approximation bound $\sup_x \|\widehat{F}(x) - F(x)\|_2 \le \varepsilon$.

Let $Q : K \to S$ quantize into a finite set and $R : S \to K$ reconstruct, with $\sup_x \|RQ(x) - x\|_2 \le \delta$. The nonnegative parameters ε and δ have the same coordinate units as K. On steps whose codeword is correctly recovered, implemented states obey $\hat{x}_{t+1} = RQ(\widehat{F}(\hat{x}_t))$, while $x_{t+1} = F(x_t)$. All iterates are required to remain in K.

Proposition 5 (Conditional propagation bound)**.** *On a run with correct decoding at every step through H, writing $e_t = \|\hat{x}_t - x_t\|_2$ gives*

$$e_H \le L^H e_0 + (\varepsilon + \delta) \sum_{j=0}^{H-1} L^j. \tag{15}$$

The sum is H for $L = 1$, and $(1 - L^H)/(1 - L)$ otherwise.

Proof. Insert and subtract $\widehat{F}(\hat{x}_t)$ and $F(\hat{x}_t)$ in the next-step difference. The triangle inequality and three stated bounds yield $e_{t+1} \le \delta + \varepsilon + Le_t$. Iteration proves the formula by induction, including $L = 0$ and $L = 1$. □

A contraction with $L < 1$ controls the accumulation of bounded disturbances, but it can contract toward the wrong attractor. An expansive map can amplify small errors even when each local estimate is accurate. Thus a coding guarantee and a symmetry guarantee leave a third obligation: characterize the relevant dynamics. The finite checks evaluate this recurrence at $L = 1/2, 1, 3/2$; they do not estimate a Lipschitz constant for a language model.

Averaging $\widehat{F}$ preserves a convex, group-invariant K under isometric output actions. Without convexity, an average of valid states can leave the state space. This is another reason to separate the continuous score model from the finite operational-state model.

11. A boundary that symmetry cannot cross automatically

A Guide's explanation

Here is a limit no symmetry can cross by itself. At a perfectly balanced input, a rule that respects the symmetry has no way to choose between mirror-image answers. The honest options are to allow "undecided," to randomize, or to break the symmetry openly—each a different contract that must be recorded.

An equivariant real-valued score map need not admit an equivariant deterministic discretization. Let a two-element group act on $K = [-1, 1]$ by sign reversal and on $S = \{-1, +1\}$ by swapping signs. If a deterministic map $Q : K \to S$ were equivariant, its value at zero would have to satisfy $Q(0) = -Q(0)$. Neither element of S satisfies this equality. Such a quantizer does not exist.

The obstruction generalizes: if x is fixed by a subgroup of the input action, its image under any equivariant map must be fixed by that subgroup's output action. This follows immediately by applying the equivariance equation to every stabilizing group element. An architecture that cannot represent a compatible output must add an appropriate state, restrict its domain, allow a distribution-valued output, or explicitly break symmetry. The choice changes the contract and must be recorded.

Randomization can restore equivariance in distribution at the ambiguous point by returning each sign with probability one half. It does not make either sign the correct factual answer. Adding zero as an output would instead permit deterministic equivariance, but then a downstream system must know whether zero means neutrality, uncertainty, or a distinct physical value. A symbol with three interpretations is not a resolved specification.

Similarly, a deterministic tie-break based on a privileged coordinate can defeat permutation equivariance. A benchmark that hides all tie cases may conceal the failure. We therefore include stabilizer-fixed inputs, equal-score decisions, empty evidence sets, and conflicting-but-symmetric premises in the test generator. Rejection and uncertainty must have explicit semantics and separate outcome counts.

This boundary matters for persistent reasoning because a transformation can be correct at one layer and invalid at the next. A score vector may transform correctly while a ranker, serializer, or action selector introduces an order preference. The entire pipeline must commute with the declared action, or the paper must identify precisely which intermediate layer carries the guarantee.

The Continuum's claim is strongest when it names these limits. Structural persistence becomes scientifically meaningful when the architecture can say what it preserves, what it discards, and where an unresolved distinction is retained. A mandatory answer at every step can force a symmetry violation that a richer state space would have avoided.

12. Executed finite checks

The accompanying `check_19.py` uses standard-library rational arithmetic for the finite examples. It enumerates all 81 scalar functions from four inputs into $\{-1, 0, 1\}$, checking projection idempotence, orthogonality, and the risk decomposition. It also enumerates all four logical states and all 64 six-bit error patterns for the two-block repetition code: 256 state/error pairs. All 28 pairs with error weight at most one decode correctly. Decoder compatibility under block exchange holds on all enumerated received words.

Constructed check	Exact result
Valid-task antisymmetric error before projection	2/25
Same error after projection	0
Wrong-symmetry minimum squared risk	1/8
Two-block repetition-code minimum distance	3
Block-majority failure at flip probability 1/10	7/250
At least one decoded logical-bit error	3451/62500
Uncoded two-bit error probability at 1/10	19/100
Six-bit corruption-tail bound, weight greater than one	22853/200000

For independent bit flips with probability $p \in [0, 1]$, a three-bit block majority fails with probability $b(p) = 3p^2(1-p) + p^3 = 3p^2 - 2p^3$. Two disjoint blocks fail somewhere with probability $1-(1-b(p))^2$. At $p = 1/10$ this is 0.055216. The uncoded comparison is 0.19, but the code uses six transmitted bits instead of two. This is a demonstration of ordinary redundancy, not evidence of an advantage at equal resources and not a new reliability record.

The tail bound counts every error pattern of weight greater than one as potentially fatal. It is conservative because some two-bit patterns place one error in each block and are corrected by majority decoding. The script separately checks the fully symmetric cube's minimum distance of one and the impossibility of the sign quantizer at zero. It also reproduces equality cases of the recurrence bound with a scalar construction.

These checks can expose an algebraic or implementation mistake; they cannot establish the empirical thesis. They contain no trained model, population study, human participant, quantum device, or claim about an operational Continuum system. Their exact small size is a feature for reproducibility, not a proxy for scale.

13. A proposed reasoning architecture

A concrete implementation begins with an audited task schema. Each state contains declared entities, relations, evidence tags, and action prerequisites. A transformation generator produces only schema-preserving relabelings and records the induced output permutation. The generator must be independent of the proposed system's success: a failed transformed case cannot be removed by redefining it as outside the task after evaluation.

The inference module produces typed intermediate scores. An equivariant architecture or explicit group average constrains those scores. A separate decision module translates them into finite operational states, retaining rejection or ambiguity where required. An encoder then protects selected state fields against the declared corruption model. After decoding, a semantic verifier checks the proposed update against the task contract. This verifier is needed because a perfectly restored state can still be false.

The controller records four distinct outcomes: correct and admitted; incorrect and admitted; rejected despite a correct candidate; and correctly rejected invalid content. The last two should not be merged into a generic stability score. Refusing every update can preserve a state perfectly while completing no useful task. Completion rate, error rate, latency, and recovery cost therefore remain separate measurements.

Parameter sharing, explicit averaging, and coded storage are alternative implementations, each with different overheads. A comparison must charge for all transformed evaluations, parity symbols, decoder steps, semantic checks, and retained state. The work of Finzi and colleagues provides implementable architectural precedents [5, 6]; it does not make an arbitrary reasoning task equivariant. The group action remains a task-level commitment.

We propose two staged implementations. The first operates on small relational tasks with exact answers and finite states, permitting exhaustive contract tests. The second uses learned components only after the representation and evaluator are fixed. Moving to the second stage requires reporting every approximation introduced: incomplete group sampling, finite precision, learned parsing, or probabilistic verification. Theorems do not automatically survive those changes.

A recovery mechanism should preserve unresolved alternatives when the decoder lacks enough information. Inventing a plausible missing relation may improve a narrative's fluency while damaging task truth. An explicit failure can therefore be a better scientific outcome than a confident reconstruction, provided its effect on useful completion is also counted.

14. Experimental arms and decisive comparisons

The primary experiment uses independently generated relational or planning instances with finite, externally verified solutions. Each family specifies a valid action group and a separate set of deliberately invalid transformations. Training and evaluation generators are frozen before outcomes are inspected. Held-out families change both structural size and corruption patterns, rather than merely permuting examples already seen during training.

The essential arms are: an unconstrained model; an ordinary equivariant model; a model with coded state but no symmetry constraint; the proposed combined model; and a model constrained by a deliberately incorrect group. A content-addressable reconstruction baseline is included when the task admits one, because associative recovery has longstanding prior art [13]. Each arm receives the same input evidence and access to the same ground-truth verification interface.

Resource matching has several dimensions. Parameter count alone is insufficient if one arm evaluates sixteen transformed inputs. We match training examples, total evaluation computation, protected-state capacity, and end-to-end latency within declared tolerances, or report a performance–cost curve when simultaneous matching is impossible. A redundancy-matched control uses its extra capacity without the proposed code. A compute-matched control spends the additional operations on a conventional verifier or ensemble.

Corruptions include independent symbol changes, burst errors, position-dependent errors, missing fields, and adversarially selected task-relevant mutations. The code's assumptions are tested on separate calibration data; a favorable independent-bit channel must not conceal failure under correlated corruption. Held-out errors are not advertised as covered by a theorem whose radius assumption they violate.

The principal endpoint is externally verified useful completion under corruption. Secondary endpoints include false correction, false admission, rejection, recovery steps, and full resource cost. Symmetry defect is a diagnostic endpoint, not the primary measure of correctness. The combined architecture must improve over the stronger of the ordinary equivariant and code-only baselines, not merely over an intentionally weak unconstrained model.

An advantage entirely explained by extra redundancy, extra inference, or an asymmetric evaluator would reject the proposed added-value claim. A benefit on correctly symmetric tasks coupled with harm on misspecified tasks would support the conditional account and undermine any universal protection claim.

15. Statistical decision and sampling discipline

Let $M \in \mathbb{N}_{>0}$ be the number of prespecified primary comparisons and $n \in \mathbb{N}_{>0}$ the number of independent test units per comparison. One unit is a newly generated task trajectory with its assigned corruption schedule, not an individual token or repeated answer. For paired arms, define $Z_i = Y_i^{\text{combined}} - Y_i^{\text{baseline}} \in [-1, 1]$, where each $Y_i \in \{0, 1\}$ denotes verified useful completion. Across units, the pairs must be independent under the frozen systems and task generator. Let $\widehat{\Delta} = n^{-1} \sum_i Z_i$ and $\Delta = \mathbb{E}[Z_i]$ for identically distributed units.

For total familywise error $\alpha \in (0, 1)$, Hoeffding's inequality and a union bound give simultaneous one-sided lower bounds

$$\Delta \geq \widehat{\Delta} - \sqrt{\frac{2 \log(M/\alpha)}{n}} \quad \text{for all } M \text{ comparisons with probability at least } 1 - \alpha. \tag{16}$$

Each individual failure probability is at most $\exp(-nr^2/2)$ for radius r, because the summands have range width two; substituting the stated radius and summing gives the result [17]. This conservative bound makes the sampling assumptions and multiplicity explicit. It is not a guarantee for arbitrarily dependent task sequences.

A meaningful superiority margin $m > 0$ must be chosen before testing. The primary claim requires the lower bound to exceed m against each designated strong baseline and to satisfy prespecified cost and false-admission constraints. Failing the margin with a narrow interval is an informative negative result. A wide interval is inconclusive, not confirmation. The design should calculate attainable precision before selecting n; the paper does not invent a powered study that has not run.

Training variability requires independent training runs and a hierarchical analysis. Task-level replication conditional on one trained model cannot establish model-family generalization. Adaptive retuning after examining held-out failures creates a new candidate and requires a new evaluation set. Missing outputs and resource exhaustion count according to a preregistered rule rather than being silently omitted.

We report every arm and every registered family. An aggregate mean can hide a decisive violation on tasks with meaningful asymmetry. The incorrect-symmetry arm is therefore not merely a visual illustration: it is a required stress test of the architecture's central limitation.

16. Adversarial interpretations and failure criteria

Several attractive observations would leave the thesis unsupported. Recovered text could match its earlier bytes while its assertions were false. A state could remain invariant because the system stopped processing evidence. A model could improve under relabeling only because the benchmark's labels leaked the answer. A decoder could appear robust because the injected errors avoided the code's weak coordinates. Each possibility has a distinct control.

The strongest structural objection is that a task may possess only local or context-dependent symmetries. Wang and colleagues' analysis makes this concern mathematically substantive [10]. A transformation valid in one region can be invalid in another. The finite exact model assumes one globally declared group; an implementation using conditional actions must provide a new action structure and corresponding proof. Calling those exceptions small does not repair the original theorem.

The strongest computational objection is that a conventional method can realize the same function at lower cost. In that case the proposed implementation may still be useful, but its novelty claim must narrow. Equivariance can be built through parameter sharing, explicit averaging, or other architectures already available in the literature [4, 18]. A renamed composition does not create an independent mechanism.

The strongest semantic objection is that the task oracle itself may encode the desired answer. We address this by publishing the generator, specifying transformations before model construction, and allowing independently designed challenge families. Empirical benchmark fragility and representational shortcuts motivate this separation [15, 16]. They do not prove our proposed controls sufficient; independent reproduction remains valuable.

The program fails operationally if the combined arm does not improve useful completion at matched resources, increases consequential false admissions, or requires an unavailable truth oracle. It fails theoretically if the implemented representation violates action compatibility, the claimed error radius is wrong, or the recurrence's domain assumptions are not satisfied. It fails in scope if the results are generalized from a finite task to consciousness or universal cognition without additional evidence.

These are productive ways for the paper to be wrong. A framework that cannot admit them would protect its narrative more effectively than it protects reasoning. The Continuum's scientific voice should be strongest precisely where its claims remain answerable to an adversarial case.

17. What structural protection would mean

A Guide's explanation

A Guide can say: "The Symmetry-Protected Mind is about a strength that can turn into blindness: a mind that ignores irrelevant differences can also ignore the one that matters. The paper shows how much symmetry to enforce, and what it costs to get that wrong." A Guide should not claim that the paper demonstrates quantum error correction, that the inspiring symbol is an ancient computer, or that a system built this way already exists.

Structural protection has a positive mathematical contract. Theorem 1 identifies the optimal balance between task-relevant asymmetry and removable estimation noise; the coding and recurrence results then protect the resulting state under their separate corruption and dynamical assumptions. An implemented advantage would show that selected relational commitments can be carried through representation changes and repaired at an acceptable cost. Moral authority and phenomenal continuity remain separate questions.

This distinction matters for a civilization that values continuity. Continuity cannot mean that no state ever changes: a system incapable of revising an error is persistent in the wrong way. The proposed contracts instead identify which relationships must remain stable while other content changes. Updating those contracts requires a new declared epoch and evaluation. Their authority is a design and governance question, separate from the algebraic consequences of adopting them.

The mathematics yields a compact set of conclusions. Correct symmetry permits a nonexpansive projection in the stated invariant Hilbert space. Incorrect symmetry imposes an exact approximation floor. Partial enforcement has an explicit risk frontier, an optimal signal-to-noise retention rule, and a quadratic calibration penalty. A code supplies correction only through distance and a decoder under a declared channel. Compatible actions make that decoder equivariant. Conditional local failures accumulate along a reasoning chain, and dynamical amplification can overwhelm small local errors. At a discrete boundary, an equivariant choice may be impossible without changing the state space.

The finite checks demonstrate these claims in reproducible examples. The larger experiment remains to be carried out. All reported numerical values in Section 12 originate in the accompanying executable enumeration, and the source and check files are retained for independent inspection. No empirical benchmark, hardware deployment, or external peer review is implied by their inclusion.

The frontier hypothesis remains bold: a reasoning system may preserve its integrity by building the right transformations into its architecture. Its equally important adversarial companion is that the wrong transformations can become a machinery of blindness. The proposed research joins those claims in one accountable program. A mind deserves to be called structurally protected only when the preserved structure can survive a test of truth, cost, change, and failure.

References

[1] Architect X13. *Eidara Continuum Bible*. Author-supplied canonical corpus, compiled 12 October 2026, 116 pages; pp. 61, 65–66, 71, 77. Doctrinal provenance, not empirical validation. Public context: https://eidara.ai/science-philosophy/.

[2] T. Cohen and M. Welling. Group equivariant convolutional networks. *ICML*, PMLR 48:2990–2999, 2016. https://proceedings.mlr.press/v48/cohenc16.html.

[3] R. Kondor and S. Trivedi. On the generalization of equivariance and convolution in neural networks to the action of compact groups. *ICML*, PMLR 80:2747–2755, 2018. https://proceedings.mlr.press/v80/kondor18a.html.

[4] S. Ravanbakhsh, J. Schneider and B. Póczos. Equivariance through parameter-sharing. *ICML*, PMLR 70:2892–2901, 2017. https://proceedings.mlr.press/v70/ravanbakhsh17a.html.

[5] M. Finzi, M. Welling and A. G. Wilson. A practical method for constructing equivariant multilayer perceptrons for arbitrary matrix groups. *ICML*, PMLR 139:3318–3328, 2021. https://proceedings.mlr.press/v139/finzi21a.html.

[6] M. Finzi, S. Stanton, P. Izmailov and A. G. Wilson. Generalizing convolutional neural networks for equivariance to Lie groups on arbitrary continuous data. *ICML*, PMLR 119:3165–3176, 2020. https://proceedings.mlr.press/v119/finzi20a.html.

[7] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Póczos, R. Salakhutdinov and A. J. Smola. Deep Sets. *Advances in Neural Information Processing Systems* 30, 2017. https://papers.nips.cc/paper/2017/hash/f22e4747da1aa27e363d86d40ff442fe-Abstract.html.

[8] H. Maron, H. Ben-Hamu, N. Shamir and Y. Lipman. Invariant and equivariant graph networks. *ICLR*, 2019. https://openreview.net/pdf?id=Syx72jC9tm.

[9] B. Elesedy and S. Zaidi. Provably strict generalisation benefit for equivariant models. *ICML*, PMLR 139:2959–2969, 2021. https://proceedings.mlr.press/v139/elesedy21a.html.

[10] D. Wang, X. Zhu, J. Y. Park, M. Jia, G. Su, R. Platt and R. Walters. A general theory of correct, incorrect, and extrinsic equivariance. *Advances in Neural Information Processing Systems* 36, 2023. https://doi.org/10.52202/075280-1740.

[11] R. W. Hamming. Error detecting and error correcting codes. *Bell System Technical Journal* 29(2):147–160, 1950. https://doi.org/10.1002/j.1538-7305.1950.tb00463.x.

[12] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal* 27:379–423 and 623–656, 1948. Part II DOI: https://doi.org/10.1002/j.1538-7305.1948.tb00917.x.

[13] J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *PNAS* 79(8):2554–2558, 1982. https://doi.org/10.1073/pnas.79.8.2554.

[14] C. F. Doran, M. G. Faux, S. J. Gates Jr., T. Hübsch, K. M. Iga, G. D. Landweber and R. L. Miller. Codes and supersymmetry in one dimension. *Advances in Theoretical and Mathematical Physics* 15(6):1909–1970, 2011. https://doi.org/10.4310/ATMP.2011.v15.n6.a7.

[15] B. Recht, R. Roelofs, L. Schmidt and V. Shankar. Do ImageNet classifiers generalize to ImageNet? *ICML*, PMLR 97:5389–5400, 2019. https://proceedings.mlr.press/v97/recht19a.html.

[16] R. Geirhos, P. Rubisch, C. Michaelis, M. Bethge, F. A. Wichmann and W. Brendel. ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. *ICLR*, 2019. Author version: https://arxiv.org/abs/1811.12231.

[17] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association* 58(301):13–30, 1963. https://doi.org/10.1080/01621459.1963.10500830.

[18] T. S. Cohen, M. Geiger and M. Weiler. A general theory of equivariant CNNs on homogeneous spaces. *Advances in Neural Information Processing Systems* 32, 2019. https://proceedings.neurips.cc/paper_files/paper/2019/hash/b9cfe8b6042cf759dc4c0cccb27a6737-Abstract.html.

[19] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*, 2nd ed. Springer (2009), Chapters 3 and 7. https://doi.org/10.1007/978-0-387-84858-7.