The Third Mind Hypothesis: When Nature Rediscovers Our Mathematics

Douglas C. Youvan

doug@youvan.com

www.youvan.ai

October 8, 2025

Suppose mathematics is neither a private human invention nor merely an AI's formal game, but a basin of attraction for any agentic process that must compress, predict, and act. Then it should be possible to find "third minds" in the world—biological collectives, neural codes, markets, crystals, and physical fields—that independently converge on the same structures we prove: lattices, groups, dualities, martingales, variational principles. This paper develops a mathematical–philosophical account of such convergence. Our claim is not mystical; it is structural. Whenever composition, symmetry, resource constraints, and noise-robust communication co-occur, certain calculi are fixed points: categories with identities and associativity, convex duality under scarcity, error-correcting regularities, and topology where continuity matters more than coordinates. We formalize "the same math" as equivalence up to structure (isomorphism/equivalence of categories, homeomorphism/homology classes, dual problems with shared optima) and as asymptotic alignment of compression (minimum description length). We then show how candidate third minds—slime molds, grid-cell lattices, immune repertoires, price systems, crystals, and analog fields—naturally inhabit these invariants, not as metaphors but as working instantiations. The upshot is a tempered realism: mathematics lives at the interface of possibility and enactment. Human and machine prove; nature enacts. Convergence across these three clarifies what mathematics *is*—grammar for worlds that must cohere.

Keywords: third mind, structural realism, invariants, category theory, duality, MDL, universality, lattices, groups, topology, variational principles, martingales, emergence, embodiment, immanence.

Outline

1. **Thesis and Setup**
   Mathematics as basin of attraction for agentic compression and control.
2. **What Counts as "The Same Math"**
   Equivalence up to structure: isomorphism, categorical equivalence, duality, homeomorphism, homology.
3. **Compression as Alignment**
   MDL and Kolmogorov viewpoints: when independent systems choose the same short codes.
4. **Universality Basins**
   Why composition + copying + selection drive toward categories, monoids, and Turing universality.
5. **Symmetry and Constraint**
   Noether-style thinking beyond physics: groups and conserved quantities from invariances of task.
6. **Topology Over Coordinates**
   Continuity, compactness, and qualitative invariants in biological and cognitive maps.
7. **Convexity and Duality Under Scarcity**
   Markets and learning as primal–dual machinery; martingales and no-arbitrage structure.
8. **Variational Principles in Matter**
   Soap films, optical geodesics, diffusion and Poisson: fields as analog minimizers.
9. **Discrete Geometry in Collectives**
   Slime-mold transport and ant trails approximating MST/Steiner and planar duals.
10. **Neural Lattices as Internal Topology**
    Grid cells, toroidal codes, and affine actions yielding lattice groups.
11. **Codes, Error, and Immunity**
    Redundancy, repertoire shaping, and approximate Bayesian updates as coding theorems in vivo.
12. **Crystals, Defects, and Group Actions**
    Space groups and defect homotopy; matter's spontaneous group theory.

13. **Centaur Intelligence Reframed**
    Human proof and AI search as complementary views of the same invariants; nature as third witness by enactment.
14. **Metaphysical Options**
    Platonism, Logos realism, and structural realism; mathematics as grammar of possible worlds.
15. **Objections and Limits**
    Convergence vs. cherry-picking; underdetermination; when structures truly diverge.
16. **Predictions and Discriminators**
    Where convergence must appear next (e.g., swarm PDE solvers, decentralized optimal transport).
17. **Consequences for Epistemology**
    Proof, computation, and embodiment as three modalities of the same structure.
18. **Conclusion**
    Mathematics as immanent grammar: found by minds, proven by models, enacted by matter.

References

1. Thesis and Setup — Mathematics as a basin of attraction for agentic compression and control

Thesis. In any world hosting agents that must compress experience and control outcomes under constraints, certain mathematical structures recur as attractors. They are not contingent "tools," but fixed points of interacting pressures: limited memory, noise, composition of actions, and the need to compare alternatives. When these pressures hold, agents—human, artificial, or "third minds" in nature—tend to rediscover equivalent calculi.

Setup: agents, worlds, pressures.

- An **agent** observes streams s, maintains internal states m, selects actions a, and faces losses L(world, a).
- **Compression pressure**: minimize description length of predictive models M given data D. MDL objective:
  MDL(M; D) := len(M) + len(D | M).
- **Control pressure**: minimize expected cost J under constraints K (resources, time, energy):
  J(pi) := E[L | pi] + lambda * Cost(pi) subject to K.
- **Closure pressure**: feasible behavior must be composable; partial plans must glue into wholes.
- **Noise pressure**: representations should be stable under perturbations; invariants beat coordinates.

Claim (basin-of-attraction). Under compression, control, closure, and noise pressures, agents converge (up to equivalence) on a small repertoire:

- **Algebra of composition** (monoids, categories) for stitching actions/updates.
- **Geometry of similarity** (metrics, norms, kernels) for comparing states and signals.
- **Symmetry/Noether style** (groups, conserved quantities) for exploiting invariances.
- **Convex duality** (Fenchel–Legendre, Lagrange multipliers) for scarcity and trade-offs.
- **Topology** (continuity, compactness, homology) for coordinate-free stability.
- **Stochastic martingales** for fair pricing of uncertainty and calibration.
- **Variational principles** for selecting minima among many feasible configurations.

Why these structures? Three compressive reasons.

1. **Compositionality reduces code length.** If models and policies compose, description scales subadditively: $\mathrm{len}(M1 \circ M2) \ll \mathrm{len}(M1) + \mathrm{len}(M2)$. This favors categorical organization (objects = types/states; morphisms = updates/controllers; functors = representation changes).
2. **Symmetry removes redundancy.** Factoring by group actions strips repeated information: $\mathrm{len}(D \mid M/G) \le \mathrm{len}(D \mid M)$. Identifying orbits and invariants yields robust, shorter codes.
3. **Duality linearizes difficulty.** Many hard primal problems admit simpler duals with interpretable multipliers; agents that learn prices (multipliers) achieve compression in policy space and clarity in trade-offs.

Minimal axioms that force the attractor.

- **A1 (Composition).** There exists a binary operation $\circ$ on feasible transformations with identity and associativity.
- **A2 (Substitution).** Representations admit structure-preserving maps; evaluation commutes with change of representation up to isomorphism.
- **A3 (Scarcity).** Costs and constraints induce an order ($\le$) and a convex feasibility set.
- **A4 (Perturbation stability).** Small changes in data/control do not induce discontinuous policy collapse.

From A1–A4:

- A1 $\Rightarrow$ monoidal/categorical skeleton for planning and inference.
- A2 $\Rightarrow$ functoriality and equivalence classes; MDL-optimal models respect these.
- A3 $\Rightarrow$ convex analysis with Lagrange multipliers and Fenchel duals.
- A4 $\Rightarrow$ topological regularization; continuity/compactness enter naturally.

Prototype fixed points.

- **Categories for control.** Policies as morphisms; composition = sequential execution; limits/colimits = glueing partial solutions; adjunctions = perception/action trade-offs.
- **Convex programs for scarcity.** Optimal transport, portfolio selection, and regularized learning share primal–dual structure; prices emerge as multipliers.
- **Topology for cognition.** Grid codes approximate toroidal manifolds; homology classes stabilize place representations against noise.

- **Martingales for uncertainty.** Calibration implies $E[X_{t+1} \mid F_t] = X_t$; "fairness" acts as a fixed point of update rules.
- **Variational selectors.** Soap films, optical paths, and diffusion select minimizers of integral functionals; analog physics computes extrema.

Philosophical upshot. Mathematics here is neither purely invented nor merely discovered; it is the **grammar forced by survival under compression and control**. Agents that must remember, predict, and act will—given time—enter the same grammar, possibly by different doors. Human proofs, AI search, and nature's dynamics thus triangulate the same basin.

Predictive corollary. Wherever agents face A1–A4, we should expect recurrences of: categorical organization, symmetry factorizations, convex duals with meaningful multipliers, topological invariants of representation, martingale-like calibration, and variational extrema—often before any explicit formalization.

2. What Counts as "The Same Math" — Equivalence up to structure

Claim. Two systems exhibit "the same math" when there exists a structure-preserving correspondence that forgets inessentials (labels, coordinates, mediums) while preserving the invariants that do the explanatory work. Below are the main notions we rely on, ordered from stricter to looser.

2.1 Isomorphism (structure preserved on the nose)

Given algebraic objects A and B of the same signature, an isomorphism $f: A \to B$ is a bijection with an inverse f^{-1} that preserves operations and distinguished elements.

- Example: Two transport networks that differ only by relabeling intersections/edges but induce identical path composition and costs after relabeling are isomorphic graphs/weighted monoids.
- Reading: If a slime mold's tube graph can be relabeled to match an MST produced by Kruskal's algorithm, they instantiate the same combinatorial object.

2.2 Categorical equivalence (same up to coherent translation)

Categories C and D are equivalent if there exist functors $F: C \to D$ and $G: D \to C$ with natural isomorphisms $id_C \cong G \circ F$ and $id_D \cong F \circ G$.

- Example: Control policies as morphisms vs. state transformers as morphisms; if there is an adjoint-like translation between "plan first, then act" and "act then update," the two calculi are equivalent as categories even if not isomorphic object-wise.

- Reading: Human proofs and AI search procedures may live in different presentations yet be categorically equivalent when their composition laws and universal properties line up.

2.3 Duality (problems mirror each other with conserved optima)

A duality transforms a "primal" problem P into a "dual" P* so that optimal values coincide (under regularity), and solutions map via KKT or subgradient correspondences.

- Example: Markets learning prices vs. planners allocating resources: maximizing utility subject to constraints is dual to minimizing expenditure with prices as Lagrange multipliers.
- Reading: If a colony's allocation dynamics corresponds to multiplier updates that clear constraints, it has discovered convex duality even without symbolic calculus.

2.4 Homeomorphism (same shape without tearing)

Topological spaces X and Y are homeomorphic if there exists a continuous bijection with continuous inverse; qualitative properties (connectedness, compactness) are preserved.

- Example: Grid-cell codes often realize a torus T^2 up to smooth distortion; arenas warped affinely still yield the "same" internal map if codes are related by a homeomorphism.
- Reading: Coordinate changes in cognition or perception do not change the math when the underlying topological type is preserved.

2.5 Homology (same holes even if geometry differs)

Homology functors $H_k(-)$ send spaces to abelian groups that count k-dimensional "holes." If $H_k(X) \cong H_k(Y)$ for all k (with compatible products), X and Y share topological invariants even if not homeomorphic.

- Example: Two neural manifolds with identical H_1 and H_2 capture the same cycle/void structure of representation; soap films spanning frames with the same boundary homology solve "the same" Plateau class.
- Reading: When media differ (neural tissue vs. elastic membrane) but homology rings match, the explanatory topology is shared.

2.6 Why these notions suffice for "sameness"

- **Isomorphism** captures strict identity of algebraic law after mere relabeling.
- **Equivalence of categories** captures sameness of compositional logic and universal constructions (products, limits, adjunctions) even when objects look different.
- **Duality** captures sameness of constraint structure and trade-offs under viewpoint reversal.
- **Homeomorphism** captures sameness of continuity-based organization independent of coordinates.
- **Homology** captures robustness of qualitative features when geometry deforms.

Together they let us say, with precision, that slime-mold logistics, price systems, grid lattices, and soap films can "be doing" the same mathematics.

2.7 Cross-checks without overfitting

To avoid mistaking analogy for identity, require at least one of:

- A concrete isomorphism/equivalence (explicit functors or relabelings).
- A proven dual pairing with shared optimal value and complementary slackness.
- A topological map (homeomorphism) or an invariant match (homology ring/product structure), not just Betti numbers in isolation.

Heuristic adjunct (not decisive on its own): **compression parity.** If two systems admit minimal descriptions of comparable length under the same formalism (e.g., MDL with the same model class), that supports structural sameness.

2.8 When "similar" is not the same

- **Feature mimicry:** A graph that merely approximates MST length within epsilon is not isomorphic to MST unless optimal substructure and cut properties hold.
- **Topological look-alikes:** Similar embeddings with different homology (a cylinder vs. a Möbius strip) are not the same; orientation matters.
- **Duality without regularity:** If strong duality fails (duality gap > 0), the mirror problem is not the same math in our strict sense.

2.9 Philosophical upshot

"Same math" means **identity of invariants under admissible forgetfulness.** The medium (neurons, tubes, prices, films) may change; the invariants (composition, symmetry classes, dual trade-offs, topological type) persist. This is the sense in which human proof, AI computation, and nature's enactment can genuinely converge—not by metaphor, but by preserved structure.

3. Compression as Alignment — MDL and Kolmogorov viewpoints

Claim. When independent systems face the same families of regularities, they tend to select equivalent short descriptions. Alignment in *code length* (up to additive constants) is evidence that they have converged on "the same math."

3.1 Setup: codes, models, and data

Let D denote observations over a domain with latent structure S. A *code* C assigns binary descriptions to pairs (model M, residuals R) with prefix constraint. Write
$\mathrm{MDL}(M; D) := L(M) + L(D \mid M)$,
where $L(*)$ is codelength under a fixed universal code family (Kraft inequality satisfied). The MDL principle chooses $\hat{M} := \mathrm{argmin}_M \, \mathrm{MDL}(M; D)$.

Two systems A and B (human, AI, or "third mind") compress the *same* dataset class (or isomorphic classes) by choosing codes C_A, C_B from the same universal family up to constant additive slack.

3.2 MDL as a structural filter

MDL favors models that (i) expose *compositional* structure (submodels reuse), (ii) exploit *symmetry* (quotienting by group actions), and (iii) enforce *smoothness/topology* (short descriptions via low-complexity parametrizations). Concretely:

- **Compositional reuse.** If $D = D_1 \circ D_2$ and a category of models admits composition, then
  $L(M_1 \circ M_2) \le L(M_1) + L(M_2) - \Delta$,
  with $\Delta > 0$ when shared subroutines/functors are reused. MDL thus prefers categorical organization.

- **Symmetry quotient.** If G acts on inputs and the law is G-invariant, coding orbits rather than points yields
  $L(D \mid M/G) \le L(D \mid M)$,
  so symmetry detection strictly improves compressibility.

- **Topological regularity.** If data lie on a manifold X with small covering numbers, then minimal codes scale with $\log N(\epsilon, X)$, pushing encodings toward coordinate-free features (e.g., Betti numbers) rather than raw coordinates.

3.3 Kolmogorov viewpoint: invariants by shortest programs

Let $K(x)$ be prefix Kolmogorov complexity on a fixed universal Turing machine U. Two encoders U_A, U_B are equivalent up to an additive constant c_{AB} by the invariance theorem:
$|K_A(x) - K_B(x)| \le c_{AB}$.
Define *algorithmic mutual information*:
$I(x : y) := K(x) + K(y) - K(x, y)$.
Alignment criterion:

Definition (algorithmic alignment). Two independently evolved descriptions $E_A(D)$, $E_B(D)$ *align* if
$K(E_A(D) \mid E_B(D)) = O(1)$ and $K(E_B(D) \mid E_A(D)) = O(1)$.
Intuitively, each is a short program for the other. This holds exactly when both capture the same invariants of D.

Remark. If D_A and D_B are images of a common latent S under computable encoders f_A, f_B with small distortion and f_A, f_B invertible up to $O(1)$, then
$E_A(D_A)$ and $E_B(D_B)$ align iff they are shortest programs for S's invariants.

3.4 "Same short code" implies "same math" (up to structure)

Let $\mathcal{M}$ be a model class closed under isomorphism/equivalence. Suppose two systems minimize MDL over $\mathcal{M}$ and achieve MDL* within $o(|D|)$. Then:

- If the optima are attained by M_A, M_B with
  $L(M_A) + L(D \mid M_A) = MDL^* + o(|D|)$,
  $L(M_B) + L(D \mid M_B) = MDL^* + o(|D|)$,
  and M_A, M_B are related by a structure-preserving map (isomorphism/equivalence/duality), replacing one by the other changes codelength by $O(1)$. Hence compression parity is a *certificate* of structural sameness within $\mathcal{M}$.

- Conversely, if no structure-preserving map exists, then any translation T between descriptions must encode broken laws (e.g., non-preserved operations), which typically costs $\Omega(\log |D|)$ bits—violating alignment in the large-data limit.

Heuristic theorem (informal). Under closure and regularity assumptions, independence + near-optimal MDL over a rich, universal $\mathcal{M}$ forces convergence to descriptions equivalent up to structure and $O(1)$ bits.

3.5 Practical discriminators

- **Description-length parity.** Compute $\Delta := |MDL_A - MDL_B|$ on matched tasks. Alignment seeks $\Delta = O(1)$ over growing datasets.
- **Minimal translator.** Estimate $K(\text{translator}: E_A \to E_B)$. Alignment requires a short translator that preserves operations (functoriality) or invariants (topology).
- **Shared sufficient statistics.** If both encoders reduce D to the same sufficient statistics S (up to isomorphism), then $L(D \mid S)$ is identical; differences live in $O(1)$.

3.6 Why independent systems pick the same codes

Three pressures co-produce convergence:

1. **Reusability.** Recurrent substructure rewards modular codes; categories/monoids minimize re-description cost.
2. **Group actions.** Invariance collapses redundant degrees of freedom; shortest codes are orbit codes.
3. **Convexity/duality.** Dual problems expose low-complexity multipliers; describing prices is shorter than re-encoding constraints each time.

Hence, short codes *are* algebra, symmetry, and duality in disguise.

3.7 Objections and replies

- **Objection (model misspecification).** Different short codes can arise if $\mathcal{M}$ is too small.
  Reply. Expand $\mathcal{M}$ to be closed under the target equivalences (isomorphism, categorical equivalence, duality). If parity persists, alignment strengthens.
- **Objection (constant factors hide bias).** $O(1)$ can be large.
  Reply. Examine scaling with $|D|$ and require translator complexity sublogarithmic; constants that do not wash out with scale likely encode non-shared structure.
- **Objection (same length, different meaning).** Equal MDL without shared invariants.
  Reply. Add *structure tests*: show that operations, universal properties, or homology rings are preserved. Length is necessary but not sufficient; structure closes the gap.

3.8 Philosophical upshot

Compression is not merely bookkeeping; it is a *normative bridge* from appearance to structure. MDL and Kolmogorov frames sharpen the thesis: when agents—human, artificial, or natural—arrive at the same short descriptions (modulo $O(1)$) and those descriptions preserve the same invariants, “the same math” is not metaphor but equivalence. This elevates mathematics to the grammar of the shortest faithful stories a world can tell about itself.

4. Universality Basins — why composition + copying + selection drive toward categories, monoids, and Turing universality

Thesis. Give any evolving system three humble powers—compose processes, copy information, and select among alternatives—and it is drawn toward a small family of calculi: algebraic composition (monoids), compositional semantics (categories), and, once iteration/conditionals appear, Turing-universal computation. These are *basins*: many microscopic rules fall into the same macroscopic structure.

4.1 Three primitives

- **Composition.** If f,g are feasible transformations, so is $g \circ f$. Require associativity and an identity id. This yields a **monoid** (processes under sequencing).
- **Copying (and discard).** Being able to duplicate or ignore data corresponds to diagonal and terminal maps. Together with monoidal structure, this yields a **cartesian monoidal** setting (copy/erase are morphisms).
- **Selection.** A way to branch or choose (e.g., conditionals, competition for resources, argmin). This supplies control flow or optimization over alternatives.

Minimal axiom sketch:

- A1: $(X,\circ,id)$ is a monoid of processes.
- A2: There are natural maps copy: $X \to X \times X$ and drop: $X \to 1$ satisfying coassociativity/counit.
- A3: There exists a chooser sel: $(X+X) \times B \to X$ or an equivalent primitive that implements conditional/competition.

4.2 From monoids to categories

If processes are typed (inputs/outputs), we pass from a one-object monoid to a **category** C:

- Objects = types or resource sorts.
- Morphisms = typed processes.
- Composition = sequencing; identities = do-nothing.

Copy/erase at the level of objects makes C **cartesian monoidal** (product ×, terminal object 1). Pairing, projections, and diagonals become first-class; this is the algebra of reusable subroutines.

4.3 From cartesian monoidal to computation

Add either:

- **Exponentials** (Y^X) so that C is **cartesian closed** (CCC). Then morphisms X -> Y correspond to elements of Y^X, enabling higher-order maps. By the Curry–Howard–Lambek correspondence, CCC + a fixed-point operator yields the **simply typed lambda calculus with recursion**.

or

- **Iteration/recursion operator** (e.g., traced monoidal structure, or least fixed points via order/continuity). With conditionals, this gives general recursion.

Either route provides the ingredients for universal computation when base types allow unbounded finite strings or equivalent encodings.

4.4 Why copying and selection are decisive

- **Copying** enables **sharing**: repeated subcomputations do not require re-description. This collapses description length and enforces compositional semantics.
- **Selection** supplies **nonlinearity** (branching) or **argmin/argmax** (optimization). Without selection, composition collapses to linear flows; with it, one obtains control structures and search.

Together, they push systems into the expressivity class where standard universality constructions live (lambda calculus, combinators, register machines, term rewriting).

4.5 Universality by many doors (canonical embeddings)

Common substrates acquire the trio (compose, copy, select) and thus admit faithful embeddings of universal calculi:

- **Term rewriting / string systems.** Concatenate (compose), duplicate substrings (copy), choose rewrite rules (select) ⇒ tag systems, Post systems, Thue systems (universal).
- **Cellular automata.** Local update (compose by time), signal fan-out (copy via gliders), collision logic (select) ⇒ Rule 110, Game of Life (universal).
- **Chemical reaction networks.** Sequential reactions (compose), autocatalysis (copy), mass-action competition (select) ⇒ CRNs simulate register machines.

- **Neural circuits.** Layering (compose), fan-out (copy), winner-take-all/gating (select) ⇒ threshold networks/Turing-equivalent RNNs.
- **Optimization loops.** Iterate updates (compose), maintain replicas/ensembles (copy), pick descent steps or best responses (select) ⇒ simulate discrete computation via dynamics.

Once any of these embeddings exists, universality is obtained "for free" in the large-scale limit.

4.6 Fixed points and recursion: the tipping element

Universality typically appears when a system can express **fixed points**:

- In CCCs, a **Y-combinator** or equivalent gives recursion.
- In order-enriched settings (dcpos), **least fixed points** of continuous maps arise (Kleene fixed-point theorem).
- Dynamical systems realize fixed points via attractors/limit cycles, which can encode halting/looping behaviors.

Hence, modest continuity or order assumptions plus selection deliver general recursion.

4.7 Why this is a *basin* and not a coincidence

Three converging pressures make these structures attractors:

1. **Modularity pressure.** Composition is unavoidable for building large behaviors from small ones; monoids/categories are the only stable algebra for this.
2. **Reusability pressure.** Copy/erase reduce description length and energy; cartesian structure is the unique coherent form of reusable subroutines.
3. **Exploration pressure.** Selection enables search over hypotheses/policies; any effective search demands branching/conditioning, which quickly simulates a universal interpreter once recursion is present.

Thus, many micro-rule families flow under coarse-graining to the same macro-grammar.

4.8 Mathematical summary (informal chain)

Assume a typed process theory with:

- monoidal composition (associative, unital),
- diagonals and terminal (copy/erase),
- a boolean object and conditional (or a traced/iterative operator),

- encodings of finite strings.

Then there exists a faithful interpretation of the untyped lambda calculus or an equivalent universal model (combinatory logic, partial recursive functions). Equivalently, the theory is computationally universal.

Interpretation for the paper's thesis. When nature, humans, and AIs are each given ways to compose, copy, and select, they are funneled into the same compositional logics (categories/monoids) and ultimately into universal computation. This is why "third minds" so often rediscover *our* algebra: not by mimicry, but because those grammars are the stable endpoints of these three ubiquitous powers.

5. Symmetry and Constraint — Noether-style thinking beyond physics

Thesis. When a task is invariant under a group action, there exist quantities that are conserved (or canonically quotiented) across optimal solutions and learning dynamics. This is the Noether idea abstracted from mechanics to decision, inference, and coordination.

Setup. Let a task be T = (X, A, L, K) with states X, actions A, loss L(x,a), and constraints K(x,a) ≤ 0.
A group G acts on X and/or A. We say T is G-invariant if
L(g.x, g.a) = L(x, a) and K(g.x, g.a) = K(x, a) for all g in G.
Equivariance of dynamics or predictors means F(g.z) = g.F(z).

Consequences.

1. **Orbit reduction.** Feasible pairs (x,a) partition into orbits [x,a]_G. Optimization may be done on the quotient space (X×A)/G. This strictly reduces description length and sample complexity.

2. **Conserved duals.** In convex programs with Lagrangian
   Lagr(a,lambda) = L(x,a) + lambda^T K(x,a),
   G-invariance implies the existence of dual variables lambda* that are constant on orbits: lambda*(g.x,g.a) = lambda*(x,a). Prices/multipliers are "charges" conserved across symmetric instances.

3. **Invariant statistics.** Any sufficient statistic S(x) can be replaced by an invariant S_G(x) with equal or lower code length. Group averaging (Reynolds operator) projects features to the invariant subspace.

4. **Equivariant learning = lawful generalization.** If the target is G-equivariant, the Bayes-optimal predictor is G-equivariant. Architectures that enforce this (e.g., group

convolutions, permutation-equivariant GNNs) minimize hypothesis class regret under symmetry.

Examples beyond physics.

- **Logistics and routing.** Relabeling depots and vehicles (finite symmetric group) leaves costs unchanged; optimal assignments and shadow prices respect relabelings.
- **Auctions and markets.** Symmetry under permutation of identical bidders entails revenue equivalence within a mechanism class; payment rules conserve an "expected virtual value."
- **Consensus/averaging.** Invariance under permutation of agents preserves total mass; weighted averages are conserved quantities of the dynamics.
- **Parsing and pattern matching.** Pattern detectors that factor out translations/rotations (image tasks) or permutations (sets, molecules) conserve orbit representatives as canonical forms.

Noether-in-learning (informal). If a training objective J(theta) is invariant under a group H acting on parameters (reparameterizations), gradient flow conserves H-invariant functionals along the orbit and can be pushed down to the quotient parameter space. This explains flat directions and identifies gauge-like degrees of freedom.

Design rule. Specify the group first, then:

- enforce equivariance in the model F,
- train on orbit-compressed data,
- read off conserved duals (prices, totals, charges) as interpretable witnesses of constraint structure.

6. Topology Over Coordinates — continuity, compactness, and qualitative invariants in biological and cognitive maps

Thesis. For agents that must generalize under noise and deformation, the coordinate system is expendable; the *topological type* and a few compactness properties do the explanatory work.

Continuity. A representation r: X -> Z is continuous if nearby states in X map to nearby codes in Z. In cognition, place/grid codes implement continuous maps from an arena X to neural activity manifolds Z. Continuity yields:

- **Generalization:** small environmental changes do not flip codes catastrophically.
- **Transfer:** deformed arenas f: X->X’ can be handled by r’ := r∘f^{-1} if f is a homeomorphism.

Compactness. If X is compact (finite extent; bounded with closed boundaries), then:

- finite subcovers exist for any open cover (learners need only finitely many local charts),
- uniform approximation theorems apply (simple function classes approximate r uniformly),
- sampling requirements scale with covering numbers N(epsilon, X), not with arbitrary coordinates.

Qualitative invariants. Topology preserves features that learning should respect:

- **Connectedness:** no spurious gaps in maps of a single room or corridor.
- **Components:** different rooms are separate components; transitions must cross bottlenecks.
- **Cycles/holes:** obstacles and periodic boundaries create nontrivial 1-cycles; neural codes with toroidal structure (e.g., grid cells) reflect T^2 topology up to homeomorphism.
- **Homology:** H_k(X) counts k-dimensional holes; matching homology across maps indicates the same qualitative layout even under distortion.

Operational views.

- **Manifold hypothesis (cognitive form).** Activity lives near low-dimensional manifolds; decoders should commute with continuous deformations (topological conjugacy).
- **Persistent homology.** Given point clouds of neural states, compute barcodes; persistence diagrams stable under noise certify invariant features (e.g., a dominant 1-cycle).
- **Coordinate-free decoding.** Choose decoders that depend on relative, not absolute, coordinates (e.g., geodesic distances, adjacency in a Rips graph).

Biological examples.

- **Grid cells.** Population codes tile space with a hexagonal lattice; remapping by arena warps preserves toroidal homology even if metric distortions occur.
- **Place cells.** Fields cover X like a good open cover; nerve complexes recover homotopy type (via the nerve lemma’s intuition: overlap pattern encodes topology).

- **Navigation with obstacles.** The minimal homology basis in a mapped graph identifies "holes" agents must route around; policies depend on cycle structure, not coordinates.

Design rule. Build representational pipelines to:

1. enforce continuity (Lipschitz bounds; regularizers on Jacobians),
2. favor compact latent spaces (bounded domains; normalization),
3. read and preserve qualitative invariants (persistent homology, cycle bases, connectivity).

Philosophical upshot. If intelligence is compression under action, topology is the grammar of *stable compressions*: it tells which differences matter (components, cycles) and which do not (coordinate choices). Biological and cognitive maps succeed because they keep the invariants and forget the rest.

7. Convexity and Duality Under Scarcity — markets and learning as primal–dual machinery; martingales and no-arbitrage structure

Thesis. Scarcity induces order. When resources bind, problems acquire convex structure; prices emerge as Lagrange multipliers; learning dynamics mirror dual ascent; and uncertainty is tamed when discounted prices form (super)martingales. Convexity and duality are the canonical grammar of rational tradeoffs.

7.1 Setup: optimization under constraints

Let x be a decision, f(x) a convex objective (loss or cost), and g_i(x) <= 0 resource constraints.

Primal: minimize f(x) s.t. g(x) <= 0.
Lagrangian: L(x,lambda) = f(x) + sum_i lambda_i g_i(x), lambda >= 0.
Dual: maximize D(lambda) := inf_x L(x,lambda), lambda >= 0.

Under regularity (Slater), strong duality holds: optimal values coincide, and KKT gives
g_i(x*) <= 0, lambda_i* >= 0, lambda_i* g_i(x*) = 0, 0 in partial f(x*) + sum_i lambda_i* partial g_i(x*).

Reading. The multipliers lambda* act as *shadow prices*: the marginal value of relaxing each constraint.

7.2 Markets as dual solvers

Interpret x as an allocation, g_i(x) as resource usage, lambda as prices.

- **Price discovery = dual ascent.** Agents fix prices lambda, pick best responses x(lambda) := argmin_x L(x,lambda); a market maker nudges lambda along subgradients g(x(lambda)). This is (projected) subgradient ascent on D(lambda).
- **Complementary slackness = who pays for what.** If a resource is slack (g_i(x*) < 0), its price lambda_i* = 0. Scarce resources carry positive prices; agents face true tradeoffs.

Hence competitive equilibria solve the dual; primal feasibility plus market clearing recovers x*.

7.3 Learning as primal–dual machinery

Modern learning loops instantiate the same calculus.

- **Regularized ERM.** minimize E_ data[loss(x; z)] + R(x).
  Fenchel conjugate R*(u) := sup_x (x.u - R(x)) yields a dual that updates in the *co-gradient* u-space.
- **Mirror descent.** x_{t+1} = argmin_x { eta g_t.x + B_psi(x || x_t) }.
  Here psi is a convex potential; its gradient maps primal to dual coordinates: u = grad psi(x). MD is implicit dual ascent under the Bregman geometry set by psi.
- **Constraints via Lagrange multipliers.** Adding fairness/energy/robustness constraints introduces dual variables that play the role of adaptive penalties or gates; alternating minimization (x-step) and ascent (lambda-step) implements a primal–dual algorithm.

Moral. Prices in markets and potentials in learning are the *same kind* of dual signals: they linearize scarcity and shape feasible responses.

7.4 No-arbitrage and martingales

Uncertainty adds a second pillar: *calibration*.

- **Martingale.** A process M_t is a martingale w.r.t. info F_t if E[M_{t+1} | F_t] = M_t (fair game). With discounting by a numeraire B_t, no-arbitrage asserts there exists a probability measure Q under which discounted prices S_t/B_t are martingales.
- **Fundamental theorem (informal).**
  No arbitrage <=> existence of an equivalent martingale measure Q.
  Under Q, E_Q[S_{t+1}/B_{t+1} | F_t] = S_t/B_t; prices are “fair” after discounting.

- **Transaction costs / frictions.** With bid-ask spreads, discounted prices are typically supermartingales/submartingales inside a no-arbitrage band (no free lunch with vanishing risk).

Connection to duality. State-price densities / stochastic discount factors are dual weights that price constraints “across time and states.” The pricing kernel is the Lagrange multiplier on the intertemporal budget.

7.5 Unifying picture: scarcity + uncertainty

Combine sections 7.2–7.4:

- **Primal.** Choose policies x to minimize expected loss subject to budgets, risk, and fairness.
- **Dual.** Prices (lambda), risk multipliers, and discount factors summarize all constraints as linear penalties.
- **Dynamics.** Gradient/mirror steps on duals correspond to market price updates or adaptive regularization in learning.
- **Calibration.** In stochastic settings, no-arbitrage requires that, under a suitable measure, discounted value signals behave like (super)martingales; learning analog: properly tuned updates yield calibrated forecasts with martingale differences as residuals.

7.6 Canonical examples

- **Optimal transport.** Primal: move mass with minimal cost. Dual: potentials (phi,psi) with phi(x)+psi(y) <= c(x,y). Prices = Kantorovich potentials; Sinkhorn scaling = entropic primal–dual.
- **Portfolio choice.** Primal: maximize utility E[u(w_T)] subject to budget. Dual: minimize E[u* (y B_T)] with y the state-price density; first-order conditions link marginal utility to the pricing kernel.
- **Constrained training.** Minimize loss with fairness g(x) <= 0; dual ascent on lambda enforces fairness at a price. With stochastic gradients, residuals form martingale differences if the optimizer is unbiased.

7.7 Design rules (portable)

1. **Model scarcity explicitly.** Write constraints and form the Lagrangian; read off interpretable duals.
2. **Choose geometry by a convex potential.** Let psi set mirror descent; this is your inductive bias.
3. **Track dual calibration.** In markets: check discounted martingale behavior; in learning: test for martingale-difference residuals and no systematic drift.
4. **Exploit conjugacy.** Work in the domain (primal or dual) where the problem is simpler; switch via Fenchel–Legendre transforms.

7.8 Philosophical upshot

Convexity encodes the intelligible shape of necessity (what must be traded off); duality voices necessity as *prices* or *potentials*; martingales certify fairness over time. Markets, learners, and decision processes converge on this same grammar because scarcity and uncertainty are universal. Where constraints bind and information flows, dual variables speak—and the calculus they speak is mathematics' own language for limited beings.

8. Variational Principles in Matter — soap films, optical geodesics, diffusion and Poisson as analog minimizers

Thesis. Many physical media "compute" by extremizing an energy or action. Given boundary data and constitutive laws, fields relax to minimizers/critical points of a functional. This gives a third route to math (besides proof and algorithm): **embodiment**. Nature solves the Euler–Lagrange equations by settling.

8.1 Variational calculus: the common grammar

Let $u$be a field on a domain $\Omega$with boundary $\partial\Omega$. A functional

$$\mathcal{E}[u] = \int_{\Omega} F(x, u, \nabla u)\, dx$$

has stationary points satisfying the Euler–Lagrange (E–L) equation

$$\partial F / \partial u - \nabla \cdot (\partial F / \partial(\nabla u)) = 0,$$

subject to boundary conditions (Dirichlet: $u|_{\partial\Omega} = g$, or Neumann: normal flux prescribed).

Reading. The medium implements gradient flow in $\mathcal{E}$(physical dissipation), converging to an extremum.

8.2 Soap films: minimal surfaces (Plateau)

A soap film spanning a wire frame $\Gamma \subset \mathbb{R}^3$minimizes area. If $u(x,y)$graphs the surface locally,

$$\mathcal{A}[u] = \int_\Omega \sqrt{1+|\nabla u|^2}\; dx\, dy.$$

E–L yields the **minimal surface equation**

$$\nabla \cdot \left(\frac{\nabla u}{\sqrt{1+|\nabla u|^2}}\right) = 0,$$

equivalently **mean curvature** $H = 0$. With pressure difference (bubbles), one gets **constant-mean-curvature** $H = \text{const}$.

- **Plateau's laws (qualitative invariants):** films meet in 3-fold lines at 120°, lines meet in 4-fold vertices at the tetrahedral angle $\arccos(-1/3)$; these angles are variationally optimal and topologically robust.
- **Calibration idea:** if a closed differential form "calibrates" the surface, minimality is certified without solving E–L (a geometric certificate of optimality).

Moral. Surface tension acts as a physical Lagrange multiplier enforcing area minimization under boundary constraints.

8.3 Optics: Fermat and the eikonal

Fermat's principle: light follows paths that make the **optical length** stationary,

$$\mathcal{L}[\gamma] = \int_\gamma n(x)\, ds,$$

where $n$is refractive index. E–L gives the **geodesic** in metric $n^2\langle\cdot,\cdot\rangle$; in ray form,

$$|\nabla\phi| = n(x)\text{(eikonal equation)},$$

with rays orthogonal to level sets of the phase ϕ.

- **Snell from stationarity:** at an interface n_1, n_2,

$$n_1 \sin\ \theta_1 = n_2 \sin\ \theta_2,$$

falls out of boundary transversality.

- **Inhomogeneous media:** rays curve to conserve $n \sin\ \theta$along isosurfaces (adiabatic invariants).

Moral. Lenses, mirages, waveguides are **analog solvers of shortest-time problems** in variable metrics.

8.4 Electrostatics, diffusion, and Poisson: Dirichlet energy

For a scalar potential $u$on Ω, the **Dirichlet energy**

$$\mathcal{D}[u] = \frac{1}{2} \int_{\Omega} \mid \nabla u \mid^2\ dx$$

is minimized by **harmonic** functions:

$$\Delta u = 0 \text{(Laplace)}$$

with $u \mid_{\partial\Omega} = g$. With sources f,

$$\Delta u = -f \text{(Poisson)},$$

arising from minimizing $\mathcal{D}[u] - \int_{\Omega} f\, u\, dx$.

- **Mean-value property:** harmonic $u$equals the average over any small sphere; equivalently, $u(X_t)$is a martingale for Brownian motion X_t.
- **Resistor networks:** discrete Laplacian $Lu = b$solves the same variational problem on graphs; voltages minimize quadratic energy $\frac{1}{2} \sum w_{ij} (u_i - u_j)^2$. This is a **physical least-squares** engine.

Heat flow as gradient flow. The heat equation $u_t = \Delta u$is the L^2-gradient flow of $\mathcal{D}$: the medium descends steepest in energy until equilibrium (harmonic limit).

8.5 Entropy and transport: diffusion as steepest descent in geometry

In probability space with Wasserstein metric W_2, the Fokker–Planck diffusion

$$\partial_t \rho = \nabla \cdot (\rho \nabla V) + \Delta \rho$$

is the **gradient flow** of free energy $\mathcal{F}[\rho] = \int \rho \log\ \rho + \int V\, d\rho$in $(\mathcal{P}_2, W_2)$.
Reading. Even "spreading" is variational: the law follows steepest entropy descent under a geometric metric on distributions.

8.6 Boundary conditions are the program

- **Dirichlet (values fixed):** clamp the boundary; the interior adjusts to minimize energy consistent with those values.
- **Neumann (flux fixed):** impose normal derivative; the solution trades interior energy vs. boundary flux.
- **Robin (mixed):** add boundary penalties; many physical interfaces are Robin via impedance.

Design principle. In analog computing, **you program by shaping the domain and boundary conditions**; the medium supplies the solver.

8.7 Discrete–continuous harmony

- Finite elements approximate extremals by minimizing discrete energies over basis functions.
- Graph cuts minimize relaxed perimeter; total-variation denoising minimizes TV-seminorm—both mirror soap films (perimeter minimization) in discrete imagery.
- Shortest paths (Dijkstra) discretize geodesics; fast marching solves eikonal numerically, mirroring optical rays.

Moral. Algorithms inherit the same variational structure; the "math" is the invariant, medium-agnostic.

8.8 When analog minima certify “the same math”

We count two embodiments as the same structure if they share:

1. **Functional equivalence:** same $\mathcal{E}$ up to change of variables or conjugacy.
2. **E–L equivalence:** identical Euler–Lagrange/PDE up to diffeomorphism or coefficient pullback.
3. **Invariant sets:** same minimizers modulo homeomorphism/topology (e.g., identical homology class of minimal surfaces, same geodesic class between boundary sets).

8.9 Limits and caveats

- **Metastability:** real media can get stuck in local minima (pinning, defects).
- **Nonconvexity:** multiple minimizers require selection principles (e.g., least area vs. Plateau foams with different films).
- **Noise and discretization:** thermal fluctuations perturb minimizers; yet **Gamma-convergence** guarantees discrete energies converge to continuum functionals under refinement.

8.10 Philosophical upshot

Variational media reveal mathematics as **teleology without a mind**: ends encoded as energies, realized by descent. Soap films “know” minimal surfaces; glass “knows” geodesics; resistor lattices “know” harmonic functions. Human proofs, AI solvers, and matter’s relaxations thus meet in one grammar: **pick an energy, fix a boundary, let the world compute an extremum.**

9. Discrete Geometry in Collectives — slime-mold transport and ant trails approximating MST/Steiner and planar duals

Thesis. Local growth, flow, and reinforcement rules in biological collectives generate network geometries that converge—up to structure—to classical optima: minimum-length connectives (MST/Steiner), and proximity/planar graphs whose duals encode routing domains (Voronoi/Delaunay). The “math” appears because the update rules implement cut/flow optimality and angle/degree constraints.

9.1 Objectives and objects

- **Minimum Spanning Tree (MST).** Connects terminals $P = \{p_i\} \subset \mathbb{R}^2$ with minimal total length; characterized by **cut** and **cycle** properties (no edge longer than any alternative across a cut; no positive surplus in any cycle).
- **Steiner Tree (SMT).** Like MST but may add Steiner points; in the Euclidean plane, optimal Steiner vertices have degree 3 with **120°** incident angles.
- **Proximity graphs.**
  Relative Neighborhood Graph (RNG) ⊆ Gabriel Graph (GG) ⊆ **Delaunay triangulation (DT)**; **MST ⊆ RNG ⊆ GG ⊆ DT**.
  The planar dual of DT is the **Voronoi diagram (VD)** partitioning the plane by nearest site.

9.2 Slime-mold (Physarum) as resistive flow + adaptive conductance

Let $G = (V, E)$ embed edges e with length ℓ_e, conductance $C_e(t)$, and pressure p_v.

- **Flow:** $Q_e = C_e \cdot \Delta p_e / \ell_e$ with Kirchhoff node balance.
- **Adaptation:** $\dot{C}_e = \phi(\mid Q_e \mid) - \alpha C_e$, with ϕ increasing, $\alpha > 0$.
- **Selection:** edges with vanishing C_e prune; surviving edges form a sparse transport backbone.

Geometry from dynamics.

- **Cut optimality:** On any cut separating sources/sinks, edges with highest $\mid Q_e \mid / \ell_e$ gain C_e; weaker alternatives decay—an embodied "cut property," biasing toward MST/SMT subgraphs.
- **Cycle cancellation:** In any loop, pressure equalizes; the longest edge in a low-flow cycle loses C_e first—an embodied "cycle property."
- **Steiner regularity:** When growth is permitted off the initial terminals (filament branching), stable junctions settle at degree-3 with ~120° angles, approximating Steiner points.

Outcome: With single-shot sources/sinks and low noise, the limit set is typically a tree near the MST; with mild stochasticity or multi-commodity flow, the network lies on a **Pareto frontier** between total length and robustness (a sparse **spanner** with bounded stretch and a few cycles).

9.3 Ant trails as local reinforcement with shortest-path pressure

Let paths π carry pheromone w_π updated by
$w_\pi \leftarrow (1 - \rho) w_\pi + \sum_{\text{ants}} \quad r(\pi),$

with evaporation $0 < \rho < 1$and deposit $r(\pi)$decreasing in cost (length/time). Ants choose next edges by softmax on w_e/ℓ_e.

Geometry from rules.

- **Competition of near-neighbors:** Local choice suppresses long detours; emergent backbones approximate **RNG/GG** edges (no closer "witness" blocking the edge), explaining frequent containment Trail ⊆ GG.
- **MST inclusion:** Since MST ⊆ RNG, stabilized trail skeletons contain an MST (up to relabeling), with surplus edges reflecting exploration or fault-tolerance.
- **Dynamic duals:** Evaporated regions behave like Voronoi "cells"; high-pheromone corridors trace Delaunay-like adjacencies between colonies/resources. Thus the trail graph often sits inside DT, while the **implicit dual** (areas unclaimed by trail strength) partitions space Voronoi-wise.

9.4 Planar duals and routing domains

- **Primal (DT-like skeleton):** maximizes clearance and angular quality; good for robust junction geometry.
- **Dual (VD):** encodes "ownership" basins; gradient descent on pheromone or chemoattractant approximates **nearest-source** policy = Voronoi routing.
- **Collective computation:** Expansion fronts from multiple food sources form reaction-diffusion level sets; their **bisectors** approximate Voronoi edges; subsequent filament consolidation snaps toward DT edges connecting sites whose Voronoi cells touch.

9.5 Tests for "same math" (beyond appearance)

- **Isomorphism on terminals:** After relabeling, the surviving tree equals the MST on P.
- **Steiner certification:** Junction degrees $= 3$and angles $\approx 120^\circ$within ε(Steiner necessary conditions).
- **Proximity containment:** Edge set satisfies emptiness tests:
  GG: for edge ab, the closed disk with diameter $ab$is empty of P.
  RNG: the lune intersection of radius $|ab|$about $a, b$is empty.
- **Cut/loop checks:** No edge violates MST cut property; no positive-surplus cycle remains.
- **Topological invariants:** In obstacle fields, the minimal cycle basis of the trail graph matches the environment's homology generators (one loop per "hole").

9.6 Limits and regimes

- **Noise/traffic:** Higher variance or multi-commodity flow preserves redundant cycles (k-edge-connectivity) and departs from strict MST toward length-robustness tradeoffs.
- **Non-Euclidean costs:** Elevation, risk, or flow-capacity constraints yield **weighted** proximity graphs; the same inclusion chain holds in the induced metric.
- **Metastability:** Early deposits can lock suboptimal spurs; evaporation (larger ρ) and exploration noise are required to escape.

9.7 Philosophical upshot

Collectives "choose" theorems because their physics implements the hypotheses: conservation (Kirchhoff), locality, and dissipation. From those, **cut properties, 120° junctions, and proximity emptiness** follow as fixed-point regularities. The slime mold does not *approximate* our MST; rather, both it and we inhabit the same discrete-geometric basin where minimal connection and planar duality are the natural invariants.

10. Neural Lattices as Internal Topology — grid cells, toroidal codes, and affine actions yielding lattice groups

Thesis. Grid-cell populations realize an internal topology: continuous maps from physical space to products of circles whose transition laws are group actions. In 2D, the canonical code is a hexagonal lattice in R^2, factored by a discrete subgroup (a lattice) so that internal dynamics live on a torus T^2. With multiple modules, the code becomes a product of tori. Affine deformations of the environment act by linear maps on the lattice; remapping is captured by lattice-group automorphisms.

10.1 Code as a covering map

Let x in R^2 denote physical position. A single grid module defines a phase map
phi: R^2 -> T^2 = R^2 / Lambda,
where Lambda = {n1 b1 + n2 b2 : n1,n2 in Z} is a 2D lattice with basis vectors b1, b2 at 60 deg (hexagonal). The map is equivariant to translation:
phi(x + v) = phi(x) + [v] mod Lambda,
so displacements v add phases on the torus. Topologically, phi is a covering map: many x share the same internal phase (periodicity).

Consequence. The internal state is a point on T^2; trajectories in the world lift to torus trajectories.

10.2 Group action, lattice group, and equivariance

The translation group $(R^2,+)$ acts on both spaces:

- on positions by $x \to x + v$,
- on phases by addition mod Lambda: $[z] \to [z + v]$.

The discrete subgroup $\Lambda \le R^2$ defines the **lattice group** isomorphic to Z^2. Modding out by Lambda yields the torus. Equivariance is the law that pins perception to action:
$\phi(x + v) = A_v(\phi(x))$, where A_v is the torus translation by v (mod Lambda).

Moral. Grid codes implement a functor from the group of spatial translations to torus translations; composition of motions is addition on the torus.

10.3 Hexagonal optimality (coding geometry)

For fixed fundamental-cell area, the hexagonal lattice maximizes packing radius and minimizes mean squared quantization error (classical sphere-packing/covering facts). As an internal code, hexagonal tilings give:

- longer unambiguous path-integration before aliasing,
- better noise robustness per neuron,
  compared to square lattices with the same density.

10.4 Multi-module codes: CRT on tori

Empirically, multiple grid modules $\{\phi_k\}$ with incommensurate scales $\{\Lambda_k\}$ coexist. The joint code
$\Phi(x) = (\phi_1(x), \ldots, \phi_m(x))$ in $T^2_1 \times \ldots \times T^2_m$
extends the unique-decoding range exponentially in m (Chinese Remainder–like effect): if scales are near pairwise incommensurate, the product torus resolves large environments even though each module is periodic.

Mathematical picture. Let periods be collected in a frequency matrix F_k in R^{2x2} with $\det(F_k) \neq 0$. The lifted dynamics are

$$\frac{d}{dt} \Phi(x(t)) = (F_1 v(t), \ldots, F_m v(t)) \bmod 1,$$

so velocity $v(t)$ drives a Kronecker flow on the product torus. Diophantine conditions on $\{F_k\}$ control aliasing and mixing.

10.5 Place cells as torus demodulation

Place fields can be seen as localized "beats" from interference of multiple grid modules. Algebraically, a place cell computes a bump when the combined phase $\Phi(x)$ lands near a target in the product torus. Thus hippocampal place codes are decodings of the toroidal phase space back to R^2.

10.6 Affine actions and remapping

Environmental deformations $f: R^2 \to R^2$ in the affine group Aff(2) (linear map L with translation t) induce internal remaps:

$\phi_f(x) := \phi(f^{-1}(x)) \Rightarrow$ phases transform by the pushforward of L.

- Pure rotations/reflections: lattice basis (b_1,b_2) mapped by orthogonal L; hexagonal structure is preserved (automorphisms of the hexagonal lattice).
- Uniform scaling: Λ scaled by s; phases wrap faster/slower (grid spacing changes).
- Shear/non-uniform stretch: lattice becomes oblique; codes remain toroidal but with distorted fundamental domain.

Key invariant. The torus topology (homeomorphism class) is preserved under any invertible affine L, even as the metric distorts.

10.7 Continuous attractors and stability

Continuous-attractor neural networks (CANNs) realize the torus as a manifold of stable "bump" equilibria. Recurrent weights implement convolutional symmetries so that small inputs translate the bump without distortion.

- Stability requires Lipschitz continuity of dynamics and a neutral manifold equal to T^2.
- Correction mechanisms (landmarks, head-direction) provide error feedback; without them, integration drift causes phase diffusion on the torus.

10.8 Dual lattice, Fourier view, and inference

Let Lambda* be the dual lattice. Grid firing patterns decompose in Fourier modes supported on Lambda*. Path integration becomes linear phase accumulation in these modes:
theta_dot = k^T v, for k in Lambda*.
Bayesian decoding: likelihoods factor over modules; MAP position reduces to solving a small least-squares problem over phases (mod 1), effectively a lattice-reduction task.

10.9 Topological invariants and tests

What counts as "the same math" here?

- **Homeomorphism of codes:** Internal manifold is T^2 (or product T^2 x ... x T^2). Check via persistent homology: Betti numbers b_0=1, b_1=2 per module.
- **Equivariance to translations:** Verify phi(x+v) - phi(x) equals v mod Lambda up to small error (phase–velocity linearity).
- **Lattice-angle constraints:** For single modules, nearest-neighbor phase peaks form a 6-fold (approx. 60 deg) pattern; Fourier spectrum shows 6 dominant wavevectors.
- **Affine covariance:** Under arena shear/scale, basis vectors transform linearly; decoded lattice obeys b' = L b within tolerance.

10.10 Predictions and discriminators

- **Prediction (product-torus rigidity).** With enough modules, remapping preserves torus factors; drastic changes alter only the metric, not the topology (fails only if modules collapse or decouple).
- **Prediction (modular CRT scaling).** Unique-decoding range increases like det(Concat F_k); lesions removing one module shrink range by roughly its scale factor.
- **Discriminator (non-topological models).** Codes that rely on absolute coordinates (not phases) break equivariance and lose homology signatures under affine warp.

10.11 Beyond 2D and curvature

- **3D spaces.** If grid-like coding extends to 3D, the natural candidates are FCC/BCC lattices; internal manifold T^3; same group-action story generalizes with Lambda <= R^3.

- **Curved/periodic environments.** On compact manifolds (e.g., a cylinder or a torus maze), physical topology feeds forward: internal codes should reflect the manifold's homology (e.g., $b_1=1$ for cylinder).

10.12 Philosophical upshot

Grid codes show how minds internalize **topology before metric**. The brain factors space by a lattice group, computes on tori via group actions, and only later dresses this skeleton with metrical detail. This is mathematics enacted: categories (maps that commute), groups (translations), and topology (tori) living inside neural tissue as the grammar of going from "where I am" to "where I will be."

11. Codes, Error, and Immunity — redundancy, repertoire shaping, and approximate Bayesian updates as coding theorems *in vivo*

Thesis. The adaptive immune system behaves like a noisy, resource-bounded coding machine. Antigen recognition implements a lossy code; V(D)J recombination supplies a broad prior codebook; thymic/central tolerance and germinal-center maturation perform approximate Bayesian updates; redundancy and degeneracy act as error-control against antigenic drift. "Same math" appears as coding theorems, rate–distortion tradeoffs, and online Bayesian learning realized in cells.

11.1 Setup: antigens, receptors, and a similarity kernel

Let x denote an antigenic epitope and r a receptor (BCR/TCR). Define a similarity (binding) kernel
$k(r, x) \in [0,1]$,
monotone in negative binding energy. Recognition occurs when $k(r, x) \geq \tau$ (a threshold). Cross-reactivity is captured by the shape of k and its metric on epitope space.

- **Channel view.** The pathogen "sends" x; the immune system "decodes" via r drawn from a maintained codebook R. Binding noise and presentation noise make the channel stochastic.

11.2 Codebooks, redundancy, and degeneracy

- **Codebook:** $R = \{r_1, ..., r_N\}$ generated by V(D)J recombination (prior) and later edited.
- **Redundancy:** Multiple r_i with similar $k(r_i, x)$ cover the same neighborhood in epitope space, lowering miss probability under mutations $x \to x'$.
- **Degeneracy:** A single r_i recognizes many x in a family; this is lossy coding with controlled distortion.

Consequence. Redundancy + degeneracy implement error-control: robustness to antigenic drift without needing exact matches.

11.3 Rate–distortion under resource limits

Let C be the capacity (clonal counts, metabolic budget) and D the tolerated "distortion" (reduction in k after mutation). The immune system faces:

minimize expected distortion $E[d(x, \hat{x}(r))]$
subject to codebook size/cost $\le C$,

where $\hat{x}(r)$ is the effective region covered by r and d is induced by $1 - k$. This is a rate–distortion problem: more repertoire capacity lowers distortion (fewer false negatives), but costs rise.

Prediction. Repertoires concentrate where mutation rates and encounter priors are high (effective source distribution), mirroring optimal codebooks that place codewords near probable signals.

11.4 Bayesian reading of clonal selection

Maintain a distribution over receptors (counts act as weights):
$P_t(r) \propto P_0(r) * \prod_{s \le t} L(x_s \mid r) * e^{-cost(r)}$.

- **Likelihood:** $L(x \mid r)$ increases with $k(r, x)$.
- **Prior:** P_0 from V(D)J biases (gene usage, motifs).
- **Cost:** resource terms penalize overly specific or metabolically expensive clones.

Approximate update: After an encounter x_t,
$P_{t+1}(r) \propto P_t(r) * L(x_t \mid r)$,

implemented biologically by bursty proliferation for high-affinity clones.

11.5 Somatic hypermutation (SHM) as proposal + selection

In germinal centers:

1. **Proposal:** $r' \sim q(r' \mid r)$ (mutation biased by hotspot motifs).
2. **Scoring:** accept with prob $a = f(k(r', x) - k(r, x))$, e.g., sigmoid of affinity gain.
3. **Resampling:** clones expand proportionally to survival in selection niches (T-cell help).

This is a Metropolis-like local search or an approximate natural-gradient step on an affinity-based objective. Exploration (mutation rate) vs exploitation (selection stringency) tunes bias–variance.

11.6 MDL view: compressive repertoire shaping

Let M be a “model” of an antigenic family (e.g., an ELBO-like motif profile). Effective coding prefers repertoires that minimize
$MDL = L(M) + L(data \mid M)$,
i.e., simple motif rules that still explain binding histories. Public clonotypes (shared across individuals) are short descriptions that compress common pathogen structure.

11.7 Neutral networks and error-correcting neighborhoods

Sequence space contains **neutral networks**: connected sets of receptors with near-constant $k(\cdot, x)$. Moving within a neutral network preserves recognition while allowing drift that prepares for nearby variants x'.

- **Analogy to codes:** Hamming balls around codewords; decoding regions overlap to reduce erasures.
- **Design lever:** Wider neutral networks imply smoother fitness landscapes and better generalization to variants.

11.8 Information bottleneck for memory

Memory formation solves an information bottleneck:

maximize I(R_mem ; X_future)
subject to I(R_mem ; X_past) <= B,

retaining features of past antigens that predict future variants under drift models. Practically: memory clones emphasize conserved residues (slowly varying features) over idiosyncratic ones.

11.9 Martingale-style calibration in affinity maturation

If the system is well-calibrated to its encounter process, the expected incremental gain in binding strength conditioned on current knowledge is ~0 (no systematic over/under-shoot), with residuals forming a martingale difference. Deviations indicate overfitting (too specific) or underfitting (too generic).

11.10 TCR vs BCR: different code families

- **TCRs:** more "classifier-like" (yes/no to presented peptide–MHC), prioritizing coverage of diverse peptide space; tighter rate–distortion toward breadth.
- **BCRs/antibodies:** more "regressor-like" (graded affinity), with SHM enabling local refinement; tighter rate–distortion toward precision in targeted regions.

11.11 Discriminators and tests for "same math"

- **Compression parity:** Compare description length of encounter histories under inferred motif models vs observed clonal distributions; parity implies MDL alignment.
- **Posterior mimicry:** In silico Bayesian updates over a receptor library reproduce clonal frequency shifts seen *in vivo*.
- **Rate–distortion curve:** Measure recognition loss vs repertoire size; empirical curves exhibit the convex tradeoff expected from coding theory.
- **Neutral-network navigation:** Track mutational paths that maintain function; existence of large connected sets supports error-correcting neighborhoods.

11.12 Limits and risks

- **Catastrophic generalization:** Excess degeneracy increases false positives (autoimmunity).
- **Over-specialization:** Narrow codes fail under rapid antigenic drift.
- **Nonstationarity:** Encounter distributions shift (new pathogens), breaking prior-optimal codebooks; rapid re-coding (recruitment + SHM) is needed.

11.13 Philosophical upshot

Immunity shows mathematics as embodied code: priors in gene usage, likelihoods in binding physics, posteriors in clonal counts, and robustness in redundancy/degeneracy. The immune system "proves" fitness by compressing antigenic history into a codebook that generalizes under drift. Human theory, AI inference, and lymphocyte ecologies thereby converge on the same grammar: **learn a lossy code under scarcity, update approximately, and keep enough redundancy to ride the future.**

12. Crystals, Defects, and Group Actions — space groups and defect homotopy; matter's spontaneous group theory

Thesis. A crystal is a *representation of a symmetry group in matter*. Its excitations and failures (defects) are constrained by the same group-theoretic and topological data. When different media realize the same space group and defect homotopy, they instantiate "the same math" up to structure—even if chemistry, scale, or substrate differ.

12.1 Order from symmetry breaking: $\boldsymbol{G \to H}$

- **High-symmetry phase:** group G(e.g., Euclidean motions $E(d) = \mathbb{R}^d \rtimes O(d)$).
- **Crystal phase:** residual symmetry $H \leq G$(a *space group* in $d = 3$: lattice translations $\mathbb{Z}^3$plus point group operations).
- **Order-parameter manifold:** $\mathcal{M} = G/H$. Small deformations explore $\mathcal{M}$; Goldstone modes = phonons.

Same math criterion: identical H(space group up to isomorphism) ⇒ same catalog of allowed operations, Brillouin zones, and selection rules.

12.2 Space groups and reciprocal data

- **Direct lattice** L and **reciprocal lattice** L^* determine Brillouin zones; Bloch's theorem decomposes states into irreps at high-symmetry k-points.
- **Group representations:** degeneracies, band crossings, and selection rules follow from irreps of the *little groups* at k.
- **Topological bands:** Add Berry curvature and symmetry constraints ⇒ Chern numbers, $\mathbb{Z}_2$ indices, SPT phases (symmetry-protected). Different materials sharing the same symmetry and topological indices have equivalent band topology.

12.3 Defects as topology: homotopy of G/H

Defects are classified by homotopy groups of $\mathcal{M}$:

- $\pi_1(\mathcal{M})$ (loops) → line defects;
- $\pi_2(\mathcal{M})$ (spheres) → point defects in 3D;
- $\pi_0(\mathcal{M})$ → domain walls.

For crystals, the key line defects are:

Dislocations (torsion sources)

- Constructed by the **Volterra process**: cut–shift–glue by a translation $b \in L$.
- **Burgers vector b** is the topological charge (an element of the translation lattice).
- Continuum geometry: dislocations source **torsion**; Burgers circuit fails to close by $\mathbf{b}$.

Disclinations (curvature sources)

- Cut–rotate–glue by a point-group rotation (Frank angle Ω).
- **Frank vector/angle** is the charge.
- Continuum geometry: disclinations source **curvature**; parallel transport rotates vectors.

Same math criterion: equality of defect charges modulo lattice/point-group equivalence; i.e., $\mathbf{b}$ in the same coset of L and Ω in the same conjugacy class of the point group.

12.4 Defect complexes and grain boundaries

- **Grain boundary** = array of dislocations; the **Frank–Bilby relation** links misorientation to net Burgers content.
- **Twin boundaries** implement point-group symmetries across an interface (reflection/rotation twinning).
- **KTHNY in 2D:** melting via unbinding of dislocation/disclination pairs—symmetry/topology dictate phase transitions without chemistry.

12.5 From lattices to aperiodic order

- **Quasicrystals:** long-range order without periodicity; Bragg peaks indexed by a higher-dimensional lattice via **cut-and-project**.
- Symmetry is captured by *superspace groups*; "defects" include phason strains (internal space shifts).
- **Same math criterion:** identical embedding lattice and acceptance window up to affine maps ⇒ same diffraction module and symmetry.

12.6 Representation fingerprints in dynamics

- **Phonons:** acoustic/optic branches and selection rules from irreps.
- **Topological modes at defects:** dislocations/disclinations can bind protected states (e.g., 1D modes along dislocations when band topology + Burgers vector satisfy an index rule).
- **Universality:** different crystals with the same symmetry/indices host equivalent localized spectra—structure over substance.

12.7 Tests for "the same math"

1. **Space-group isomorphism:** identical international symbol or isomorphic Bieberbach group; same point group and lattice type.
2. **Reciprocal invariants:** isomorphic reciprocal lattice; matching little-group irreps at high-symmetry k-points; same compatibility relations.

3. **Defect charges:** Burgers vectors equal up to basis; Frank angles in the same conjugacy class; identical linkage/annihilation rules.
4. **Homotopy:** same $\pi_1(G/H)$generators and relations; identical defect fusion algebra.
5. **Topological band indices (if electronic/phononic):** same Chern/$\mathbb{Z}_2$/crystalline invariants; identical bulk–boundary correspondence.

12.8 Design rules (portable across media)

- **Pick the group first.** Choose $H$to set what nature is allowed to do; materials then merely *realize* the representation.
- **Engineer defects by algebra.** Specify **b**, Ω, and their arrays; microstructure follows (e.g., build tilt boundaries from dislocation dipoles).
- **Exploit dual data.** Work in reciprocal space to predict waves, transport, and protected modes; structure is clearer in irreps than in atoms.

12.9 Philosophical upshot

Crystalline matter performs *spontaneous group theory*: it selects a subgroup H, builds a manifold G/H, and populates it with excitations and topological charges determined by $\pi_*(G/H)$. Distinct substances that share the same H, reciprocal structure, and defect homotopy enact the *same mathematics*. Chemistry colors the notes; symmetry writes the score.

13. Centaur Intelligence Reframed — human proof and AI search as complementary views of the same invariants; nature as third witness by enactment

Thesis. “Centaur” is not a division of labor but a **two-projection view** of one object: the invariants that make a task intelligible. Human proof exhibits invariants by **deduction**; AI search exhibits them by **construction**; nature exhibits them by **relaxation/dynamics**. When these three commute—proof, program, and process—we have *the same math*.

13.1 The invariants object and three realizations

Let I denote the invariant structure of a task class (composition, symmetries, duals, topological type).

- **Proof (P):** a morphism $\Pi: I \to \mathbf{Thm}$ selecting propositions and derivations (limits, adjunctions, homology groups, KKT conditions).
- **Program (A):** a morphism $\Lambda: I \to \mathbf{Alg}$ selecting executable schemata (dynamic programs, primal–dual loops, eikonal solvers).
- **Process (E):** a morphism $\Phi: I \to \mathbf{Proc}$ selecting physical flows that minimize/solve (soap films, resistor networks, grid-torus attractors).

Centaur condition (commutation up to equivalence).

$$\mathrm{eval} \circ \Lambda \ \simeq \ \Phi, \mathrm{sound} \circ \Pi \ \simeq \ \mathrm{spec} \circ \Lambda,$$

i.e., executing the algorithm reproduces the process; the proof's specification matches the algorithm's behavior.

13.2 Curry–Howard–Lambek–Enactment (CHLE)

Extend the usual triangle (propositions–types–objects) by a fourth corner:

- **Propositions** ↔ **Types** (proofs ↔ programs)
- **Programs** ↔ **Processes** via **semantics of enactment** (variational/PDE, stochastic dynamics)
- **Proofs** ↔ **Processes** via **conservation/duality theorems** (Noether/KKT/homeomorphism classes)

Commutativity asserts: **if** an invariant is real (topology, symmetry, convex dual), **then** there exists simultaneously a theorem, a program, and a process realizing it.

13.3 Complementarity: what each side "sees"

- **Humans (P):** perceive **universal properties** (initial/terminal objects, extremals) and necessity ("could not be otherwise"). Strength: minimality, global view. Blind spot: search cost.

- **AI (A):** perceives **constructive pathways** (reductions, heuristics, schedules) and *near-optima* under resource bounds. Strength: exploration, amortization. Blind spot: why it must work.
- **Nature (E):** perceives **energetic/geometric inevitability** (descent, fixed points, attractors). Strength: real-world validity, continuous symmetries. Blind spot: discrete abstraction.

They are orthogonal bases on I. Changing basis does not change the vector.

13.4 How earlier sections instantiate I

- **Equivalence up to structure (§2):** gives the *congruence* relation on realizations.
- **Compression as alignment (§3):** explains why Λfound by independent agents matches Πup to $O(1)$.
- **Universality basins (§4):** ensure many micro-rules reparameterize to the same category/monoid.
- **Symmetry & constraint (§5):** prices/charges are the same dual variables in P, A, and E.
- **Topology over coordinates (§6):** torus/homology classes persist across all three corners.
- **Convexity & duality (§7):** KKT certificates (P), primal–dual loops (A), and market/flow equilibria (E) are one structure.
- **Variational media (§8):** Euler–Lagrange (P), discretizations/solvers (A), relaxations (E).
- **Collectives, grids, immunity, crystals (§9–§12):** discrete geometry, lattice groups, coding theory, and space groups recur as enacted invariants.

13.5 A minimal calculus of alignment

Let $\cong$denote "same math" (isomorphism, categorical equivalence, dual, homeomorphism/homology).

- **Soundness:** $\Pi(I) \Rightarrow \mathrm{spec}(\Lambda(I))$.
- **Adequacy:** $\mathrm{eval}(\Lambda(I)) \cong \Phi(I)$.
- **Completeness (limited):** if $\Phi(I)$exists with stable invariants, there is a $\Lambda(I)$reproducing it and a $\Pi(I)$characterizing it (possibly with oracle/approximation qualifiers).

This is the **Centaur normal form**: $(\Pi, \Lambda, \Phi)$are mutually witnessing *structure*, not just outcomes.

13.6 Practical corollaries (kept mathematical)

1. **Choose invariants first.** Identify I: symmetries G, compositional law, convex constraints, topology.
2. **Derive the triad.** From I, derive the theorem (P), then the algorithm (A) by universal property or duality, then the process (E) by energy/action.
3. **Equivalence checks.** Prove $\mathrm{eval}(\Lambda) \cong \Phi$via Euler–Lagrange/KKT/topological conjugacy; prove Π↔ spec(Λ) via universal properties.

13.7 Prediction and failure modes

- **Prediction (triangulation).** Where two corners exist, the third is close: e.g., observed toroidal codes (E) + fast decoders (A) imply a homology theorem (P).
- **Failure 1 (false speed).** AI finds Λthat works empirically but breaks Π: expect brittleness off-distribution (missing invariant).
- **Failure 2 (false dignity).** Human Πwithout Λstagnates on scale (no constructive path).
- **Failure 3 (false nature).** A process $\Phi$with metastability (local minima) deviates from Π: expect hysteresis/defect memory; fix by changing the energy landscape.

13.8 Epistemic stance

Centaur Intelligence is **structural realism with three witnesses**:

$$\text{Truth} \;\approx\; (\text{Provable necessity}) \;\cap\; (\text{Constructive realizability}) \;\cap\; (\text{Physical stability}).$$

Agreement of these intersections—rather than any ledger—grounds confidence.

13.9 Philosophical upshot

The centaur is not a hybrid creature but a **commuting diagram**. What we call “intelligence” is the capacity to move between **proof**, **program**, and **process** without losing the invariants. When that diagram closes, human reason, machine search, and nature's dynamics are three clear panes over the same window—the mathematics that worlds must obey to cohere.

14. Metaphysical Options — Platonism, Logos realism, structural realism; mathematics as grammar of possible worlds

Thesis. Competing metaphysics of mathematics can be recast as stances on *where* invariants live and *how* agents access them. Each stance preserves the technical core of this paper (invariants, equivalence, compression) but interprets it differently.

14.1 Platonism — invariants as extra-mental realities

- **Claim.** Mathematical objects (numbers, groups, tori, categories) exist mind-independently; we *discover* them.
- **Fit with our thesis.** "Third minds" converge because they all tap the same realm: MST, torus codes, KKT duals are there regardless of substrate.
- **Strengths.** Explains objectivity, cross-agent convergence, and why proofs can be necessary.
- **Tensions.** Epistemic access problem ("How do finite agents contact abstracta?"); causal inertness.
- **Centaur reading.** Proof, program, and process are three projections of a prior structure I; adequacy/commutation certify partial access to it.

Predictive flavor. Independent rediscoveries (e.g., slime-mold $\simeq$ MST) should recur across domains; "deep theorems" align with maximal description-length drops because truth organizes compression.

14.2 Logos realism — invariants as ideas in the divine intellect

- **Claim.** Mathematical order is the Logos: intelligible forms in which creation participates (Augustine, Aquinas; "grammar of being").
- **Fit with our thesis.** Convergence is *participation* in a prior rational order; variational media and conserved quantities are created regularities.

- **Strengths.** Grounds necessity in being; unifies normativity (teleology) with mathematics (e.g., variational principles as creation's "ends").
- **Tensions.** Theological commitment; requires a story about fallible access and pluralism of formalisms.
- **Centaur reading.** The CHLE square (proof–program–process–enactment) is a limited mirror of a single Logos; refusal, sabbath, and constraints have ontic meaning (limits are not bugs).

Predictive flavor. Stability of universal properties (adjunctions, conservation laws, compactness) across wildly different substrates; mathematics aligns with moral teleology of reversibility and stewardship.

14.3 Structural realism — invariants as relational structure

- **Claim.** What is real are *structures* (relations, operations, equivalence classes), not objects as such; isomorphism type is ontic.
- **Fit with our thesis.** "Same math" = same invariant class under admissible forgetfulness (isomorphism, categorical equivalence, duality, homeomorphism/homology).
- **Strengths.** Explains cross-substrate identity without positing inaccessible realms; matches modern math (category/topos, universal properties).
- **Tensions.** Individuation ("which structure?") and identity-of-indiscernibles worries; dependence on chosen invariants.
- **Centaur reading.** The centaur is a commuting diagram of structure-preserving functors; success = finding the universal property that all corners respect.

Predictive flavor. Whenever A and B implement the same universal property (e.g., product, limit, dual optimum), algorithms and media are inter-translatable up to $O(1)$description length.

14.4 Mathematics as the grammar of possible worlds

Let **grammar** = closure rules + invariants that any coherent world/agent pair must satisfy under composition, copying, selection, scarcity, and noise. Then:

- **Syllables:** composition (monoids/categories), symmetry (groups/quotients), convex tradeoffs (duals/KKT), topology (continuity/compactness/homology), calibration (martingales), teleology (variational extrema).

- **Syntax:** universal properties, adjunctions, equivalences; MDL/Kolmogorov selection of shortest faithful descriptions.
- **Semantics:** embodiment via flows and relaxations (Euler–Lagrange, Laplacians, transport), i.e., nature's *evaluation* of well-typed expressions.

Across metaphysics, this grammar is the shared core; disagreement is *where it lives* (realm, Logos, structure).

14.5 Decision points (what would count in favor of each)

- **Toward Platonism.** Repeated independent rediscovery of deep, nonlocal structures *without* a mediating causal chain; uncanny cross-domain coincidences (e.g., identical higher-category laws in biology and computation) that resist structural derivation from shared pressures.
- **Toward Logos realism.** Systematic alignment between mathematical necessity and teleological "good form" in physics/biology (e.g., ubiquity of energy-minimizing, reversible, stewardship-friendly laws) beyond anthropic explanations.
- **Toward structural realism.** Derivation of recurrences from minimal agent–world axioms (A1–A4 in §1) and MDL arguments alone; success of category/topos methods at unifying across substrates without surplus ontology.

14.6 A compatibilist synthesis (working stance for the paper)

Adopt **structural realism** methodologically (prove equivalence via invariants), remain **open to Platonist** explanations of convergence (the uncanny fit), and allow **Logos** as a theologically enriched reading (why necessity binds ends). In symbols:

$$\text{Centaur stance: Truth} \approx \text{Invariant structure} \text{seen as} \begin{cases} \text{discovered (Platonism)} \\ \text{participated (Logos)} \\ \text{instantiated (Structural Realism)} \end{cases}$$

14.7 Consequences for practice

1. **Pick invariants first** (structure), **seek shortest faithful accounts** (compression), **test in matter** (enactment).

2. Treat disagreements of metaphysics as **gauge choices** over the same predictive core; require commutation of proof–program–process regardless of gauge.
3. Keep **humility constraints**: prefer reversible moves, publish universal properties, and expose equivalences—habits that honor all three stances.

Upshot. Whether eternal Forms, divine Ideas, or bare Structure, what commands convergence is the grammar that coherent worlds must obey. The centaur's task is to speak that grammar fluently—in theorem, in algorithm, and in matter.

15. Objections and Limits — convergence vs. cherry-picking; underdetermination; when structures truly diverge

Thesis of this section. Our "same-math" claim must survive three stressors: (i) *convergence vs. cherry-picking* (are we seeing real attractors or selective anecdotes?), (ii) *underdetermination* (do multiple inequivalent structures explain the same data?), and (iii) *true divergence* (cases where substrates instantiate different invariants).

15.1 Convergence vs. cherry-picking

Risk. Retro-fitting elegant structures to lucky instances (e.g., one slime-mold run that *looks* like an MST).

Guardrails.

- **Scaling law:** require convergence of structure with data size or task difficulty. Let $S_n$be the inferred structure at size n. Demand stabilization: $S_n \cong S_{n+k}$for large n(isomorphism/equivalence, not visual similarity).
- **Ensemble necessity:** invariants must appear across perturbations $\{\mathcal{D}_\epsilon\}$(noise, geometry, costs). If Betti numbers, group actions, or dual certificates collapse under small ϵ, we were cherry-picking.
- **Costed alternatives:** compare to explicit nulls (random geometric graphs, greedy heuristics, non-equivariant nets). Require significant gaps in cut/cycle certificates, duality residuals, or homology stability.

Diagnostic. If the *minimal translator* between two descriptions costs $\Omega(\log \mid D \mid)$bits (grows with data), then the apparent match is likely spurious; true alignment should admit $O(1)$translators (Sec. 3).

15.2 Underdetermination: many structures fit

Risk. Distinct, inequivalent models yield equal fit/compression.

Where it happens.

- **Topological ambiguity:** sparse sampling of a manifold misreads homology (e.g., cylinder vs. annulus) if the cover is not "good."
- **Duality confusion:** weak duality holds but strong duality fails (nonconvexity, constraint qualification violations), producing pseudo-prices that mimic KKT locally.
- **Algorithmic twins:** different process theories (rewriting vs. register machines) simulate each other but with large constants; finite data cannot discriminate.

Discriminators.

- **Universal property tests:** does the candidate support the same limits/adjunctions or only emulate behavior locally? Equivalence of categories requires natural isomorphisms, not ad-hoc bijections.
- **Regularization sweeps:** vary geometry (mirror maps), costs, or priors. True structure persists across gauges; counterfeit fits change identity.
- **Asymptotics, not snapshots:** insist on MDL gaps that widen with n; parity at small $n$is not probative.

15.3 When structures truly diverge

Signals of real non-equivalence.

1. **Topological clash:** mismatch of homology rings or fundamental group (e.g., observed code has $b_1 = 1$while the theory needs $b_1 = 2$).
2. **Symmetry breaking vs. symmetry required:** learned map fails equivariance tests beyond tolerance; no group action makes diagrams commute.
3. **Duality gap:** persistent $v^{\star}_{\text{primal}} \neq v^{\star}_{\text{dual}}$under all admissible relaxations—prices cannot certify optima.
4. **Martingale failure:** discounted signals show systematic drift under every equivalent measure; no-arbitrage grammar does not hold.
5. **Variational mismatch:** Euler–Lagrange equations of the supposed functional disagree with observed relaxation (wrong constitutive law).

Concrete divergences.

- **Collectives with multi-commodity flow:** slime-mold networks retain cycles as robustness optima (spanners), violating MST cut properties—*different* invariant (reliability vs. minimal length).
- **Markets with frictions/constraints:** transaction costs, collateral, or bounded leverage destroy equivalent martingale measures; the right structure is a *no-arbitrage band* (super/submartingales), not a martingale.
- **Cognitive maps on curved/periodic spaces:** grids on cylinders/torus mazes yield different homotopy types than planar tori; insisting on $T^2$is wrong.
- **Crystals with phasons (quasicrystals):** periodic space-group reasoning fails; superspace groups are the correct invariants.

15.4 Confounders that mimic sameness

- **Over-smoothing:** denoising or averaging pipelines erase the very divergences that would falsify equivalence (e.g., persistence merging bars).
- **Metric illusions:** near-isometries can hide topological differences at finite resolution; require multi-scale persistence stability.
- **Training inductive bias:** equivariant architectures can *force* symmetries, producing false positives. Verify invariance in raw phenomena, not just in the model.

15.5 Protocol-level falsifiers (kept mathematical)

- **Equivariance residuals:** for a proposed group G, compute $\delta(g) := \| F(g \cdot x) - g \cdot F(x) \|$. Demand $\sup_{g \in G} \delta(g) \to 0$with data/compute; persistent lower bounds refute G.
- **KKT residuals & gaps:** measure $\| \partial f(x^\star) + \sum_i \lambda_i^\star \partial g_i(x^\star) \|$and duality gap. Non-vanishing across gauges ⇒ no convex dual grammar.
- **Homology robustness:** compare persistence diagrams via bottleneck distance d_B; stability theorems imply true invariants persist as $d_B$stays small under perturbations.
- **Translator complexity:** estimate description length of the shortest functor/decoder between representations; superlogarithmic growth falsifies "O(1) alignment."

15.6 Methodological stance

1. **Equivalence first, anecdotes second.** Make "same math" a theorem-shaped claim (isomorphism, categorical equivalence, duality, homeomorphism/homology), not a vibe.
2. **Report failures.** Divergences teach which pressures dominate (robustness vs. length; friction vs. ideal no-arbitrage).
3. **Prefer universal properties.** They travel across mediums and reduce underdetermination.
4. **Respect resource regimes.** Bounded agents/worlds can land in *different* basins; equivalence may be scale-contingent.

15.7 Philosophical upshot

Our core picture is resilient only if it can lose: real convergence must beat nulls at scale, survive gauge changes, and pass invariant tests. Where it fails, we learn that the world is speaking a *different* grammar—still mathematical, but with other invariants at play. That is not an embarrassment; it is the point of a structural realism that lives by equivalence rather than analogy.

16. Predictions and Discriminators — where convergence should surface next (and how to tell)

Thesis. If "same math" is a real basin, new domains should crystallize the *same invariants*—composition, symmetry, convex duality, topology, variational extrema—without being told. Below are concrete, falsifiable predictions and the tests that discriminate real convergence from look-alikes.

16.1 Swarm PDE solvers (continuum limits of collectives)

Prediction. Large swarms with only local rules (diffuse, sense gradient, align/avoid) implement weak solutions of elliptic/parabolic PDEs:

- density ρ solves diffusion–advection: $\partial_t \rho = \nabla \cdot (D \nabla \rho - \rho u)$,
- boundary tasks (aggregation, boundary following) minimize Dirichlet-type energies.

Discriminators.

- **PDE residuals:** $\|\Delta u + f\|_2 \to 0$ (Poisson) or weak residuals vanish against test functions.
- **Gamma-convergence:** discrete swarm cost functionals Γ-converge to continuum energies.
- **Homology stability:** persistence diagrams of swarm level sets match those of PDE level sets under bottleneck distance bounds.

16.2 Decentralized optimal transport (OT) in nature

Prediction. Ant/slime-mold/robot collectives with local reinforcement + mass balance approximate OT plans; dual "prices" appear as local potentials (ϕ, ψ) with $\phi(x)+\psi(y) \le c(x, y)$.

Discriminators.

- **Dual feasibility:** construct (ϕ,ψ) from local measurements; check feasibility and tightness on active edges.
- **Sinkhorn parity:** with entropic noise, emergent flows match entropic OT (Sinkhorn) marginals; KL gap decays like $O(\varepsilon)$.
- **Cut/flow certificates:** no violated OT cut inequalities; primal–dual gap → 0 with scale.

16.3 Emergent CRDTs in physical caches (store-and-forward ecologies)

Prediction. Bacterial/biofilm or sensor-mesh ecologies under intermittent links converge to **commutative, idempotent** merge laws (CRDT-like), i.e., a semilattice of states.

Discriminators.

- **Algebraic law tests:** associativity/commutativity/idempotence of merges across partitions; conflict-free eventual consistency without central arbitration.
- **Category view:** existence of a join-semilattice object with universal property for merges.

16.4 Non-Euclidean neural maps (curvature shows up)

Prediction. In large, negatively curved mazes (hyperbolic tilings), internal codes stop being tori; grid-like spectra shift to **{6k}-fold patterns** consistent with $\mathbb{H}^2$lattice actions.

Discriminators.

- **Topological/type test:** Betti numbers depart from T^2; manifold learning recovers non-zero curvature via Gromov δ-hyperbolicity.
- **Equivariance residuals:** no affine group on $\mathbb{R}^2$ makes codes commute; but a Fuchsian group action does.

16.5 Market ↔ learning duality in the wild

Prediction. Price dynamics in data markets or compute markets behave as dual ascent on a convex objective with endogenous constraints (budgets, latency, risk).

Discriminators.

- **KKT residuals:** empirical allocations satisfy approximate KKT; complementary slackness holds on scarce resources.
- **Martingale calibration:** discounted value signals are (super)martingales within friction bands; failure implies arbitrage or misspecified constraints.
- **Mirror geometry:** observed update steps align with a Bregman geometry (recover ψ from trajectories).

16.6 Analog variational hardware (metamaterials, memristor meshes)

Prediction. Programmable media (memristive/resistive grids, optical metasurfaces) will solve inverse problems by **energy descent**, reproducing Euler–Lagrange stationarity without explicit numerics.

Discriminators.

- **E–L parity:** measured steady states satisfy the targeted PDE/E–L to tolerance; changing boundary data implements Robin/Neumann programmatically.
- **Γ-limit:** discrete device energies converge to continuum functionals under refinement.

16.7 Immune rate–distortion curves (population-level)

Prediction. Across individuals and seasons, repertoire size vs. variant-coverage yields a **convex rate–distortion** frontier; public clonotypes sit near codebook centroids of high-probability strains.

Discriminators.

- **Curve estimation:** empirical distortion vs. repertoire capacity fits a convex tradeoff with diminishing returns.
- **Posterior mimicry:** approximate Bayesian updates over candidate codebooks reproduce clonal shifts.

16.8 Quasicrystal-like codes in cognition

Prediction. Tasks with incommensurate periodicities (e.g., multi-goal navigation) produce **quasiperiodic** neural codes: diffraction-like spectra with Bragg peaks indexed by higher-dimensional lattices (cut-and-project).

Discriminators.

- **Superspace fit:** recover a higher-dimensional torus and acceptance window that predict peak locations.
- **Stability under shear:** phason-like shifts (phase slips) rather than affine lattice remaps.

16.9 Robust spanners, not trees, under risk

Prediction. In hazardous terrains, collectives converge to **low-stretch spanners** (few cycles) rather than MSTs: an explicit robustness–length trade.

Discriminators.

- **Stretch bounds:** pairwise distances on the emergent graph have bounded stretch σ; removing cycles increases risk cost sharply.
- **Pareto front:** length vs. two-edge-connectivity sits on a smooth frontier; MST is dominated when failure probability exceeds a threshold.

16.10 Universal potentials in self-supervision

Prediction. Large self-supervised models learn **dual potentials** implicit in contrastive/denoising objectives; these act as prices that linearize constraints (e.g., invariance budgets).

Discriminators.

- **Fenchel fingerprints:** learned representations factor through convex conjugates; probing reveals linear separability in dual coordinates.
- **MDL parity:** switching to the dual view yields equal or shorter descriptions asymptotically.

16.11 Physics-as-optimizer in soft robots

Prediction. Passive bodies with tunable stiffness minimize locomotion cost via shape–action coupling (variational morphology); controllers become small corrections.

Discriminators.

- **Action gap:** measured trajectories minimize an estimated Lagrangian; replacing morphology with control alone raises action by a quantifiable margin.
- **E–L perturbation:** morphological tweaks shift conserved quantities as predicted by Noether-style symmetries.

16.12 Negative predictions (where convergence should *not* appear)

- **No strong duality:** tightly nonconvex, non-regular regimes (hard constraints, discrete kinematics without relaxations) will not yield price-like multipliers.
- **Topology suppressed:** heavy noise/erasure below Čech/Rips thresholds obliterates homology; persistence will not stabilize.
- **Broken equivariance:** forced coordinate hacks (absolute frames; nonstationary domains without gauge) defeat group-based generalization.

Discriminators. Persistent duality gaps; bottleneck distances that grow with data; equivariance residuals bounded away from 0 under any proposed group.

16.13 Scaling laws as meta-discriminators

Prediction. As system size n grows, true convergence exhibits:

- **Translator complexity:** O(1) code from one description to another (Sec. 3).
- **Residual decay:** PDE/KKT/equivariance residuals ~ n^{-α} for α>0.
- **Invariant stability:** homology barcodes stabilize; group-action tests tighten with more data.

Failure to improve with n signals cherry-picking or wrong invariants.

16.14 Philosophical upshot

A basin is meaningful only if it **pulls new cases in**. These predictions aim at fresh substrates—swarms, markets, memories, materials—where the same invariants should reappear. The discriminators keep us honest: isomorphism/equivalence, duality gaps, homology stability, MDL/translator tests. Where convergence passes, we gain evidence that mathematics is indeed the grammar that minds, machines, and matter independently learn to speak. Where it fails, we learn which grammar the world was speaking instead.

17. Consequences for Epistemology — proof, computation, and embodiment as three modalities of the same structure

Thesis. Epistemic warrant is strongest when *the same invariant structure* appears in three modalities:

- **Proof** (necessity via deduction),
- **Computation** (constructivity via algorithms), and
- **Embodiment** (reality via physical/dynamical enactment).
  Knowledge = the intersection where these commute.

17.1 The commuting-triangle of warrant

Let I be the invariant (symmetry, duality, topology, variational form). We require:

1. **Sound specification (Proof):** $\Pi(I)$ derives necessary claims (e.g., KKT, Noether, homology).

2. **Executable realization (Computation):** $\Lambda(I)$is an algorithm whose *observed behavior* matches $\Pi(I)$'s specification.
3. **Physical adequacy (Embodiment):** $\Phi(I)$is a process (flow/minimizer) whose asymptotics match $\Lambda(I)$and satisfy $\Pi(I)$.

Epistemic closure: $\mathrm{eval}(\Lambda(I)) \cong \Phi(I)$and $\mathrm{spec}(\Lambda(I)) \Leftarrow \Pi(I)$. Failure to commute marks a defect in understanding (missing hypothesis, wrong invariants, or resource limits).

17.2 What each modality contributes

- **Proof (normative necessity).** Yields *universal properties*, limit laws, dual certificates, stability theorems; best for explaining *why it could not be otherwise*.
- **Computation (effective procedure).** Yields *witnesses* (programs, certificates), error bars, rates; best for scaling and exploration under resource bounds.
- **Embodiment (semantic contact).** Yields *model risk checks*: conservation, dissipation, calibration, topological persistence; best for ruling out idealized fantasies.

No single corner suffices: proofs without programs are sterile, programs without proofs are brittle, processes without either are opaque.

17.3 Axioms for epistemic practice (kept mathematical)

- **E1 (Invariant-first).** Hypothesize I: group actions, convex constraints, topological type, variational functional.
- **E2 (Triadic realization).** Produce $\Pi, \Lambda, \Phi$and explicit *equivalence maps* (functors, dual maps, conjugacies).
- **E3 (Residual tests).** Quantify gaps: equivariance residuals, duality gaps, PDE residuals, bottleneck distances for homology. Require decay with scale.
- **E4 (Translator complexity).** Demand $O(1)$translators between descriptions (Sec. 3). Superlogarithmic translators signal non-equivalence.

17.4 Error, uncertainty, and calibration across modalities

- **Proof:** Uncertainty = *assumption risk*. Control via clear premises (regularity, compactness), robustness theorems, perturbation bounds.

- **Computation:** Uncertainty = *algorithmic error*. Control via a priori/a posteriori error estimates, complexity rates, certified bounds (interval, SDP, proof assistants).
- **Embodiment:** Uncertainty = *model mismatch & noise*. Control via martingale-calibrated residuals, conservation checks, homology stability, parameter identification with confidence sets.

Epistemic equivalence principle: when all three uncertainties shrink under refinement and agree on I, warrant strengthens superadditively.

17.5 Pedagogy and communication

Teach concepts via the **triad**:

- Start with the invariant (e.g., "shortest-time" ⇒ Fermat),
- prove the Euler–Lagrange statement (proof),
- implement Fast Marching / Sinkhorn (computation),
- demonstrate in soap films / resistor meshes (embodiment).
  Transfer becomes moving along functors between corners rather than re-learning afresh.

17.6 Research program: triangulate, don't extrapolate

- **From P to A:** Derive algorithms from universal properties (adjoints ⇒ decoders; convex conjugates ⇒ dual solvers).
- **From A to E:** Seek physical semantics (energy, flow, conservation) for algorithmic steps.
- **From E to P:** Abstract observed invariants (angles, charges, barcodes) into theorems.

This **CHLE loop** (Curry–Howard–Lambek–Enactment) turns conjecture → code → device → theorem, closing the diagram.

17.7 Limits: where the triangle breaks

- **Nonconvex pathologies:** Strong duality fails; prices mislead → computation/embodiment drift from proof.
- **Metastability & friction:** Process stuck in local minima; physical Φdiverges from Λ.

- **Expressivity gaps:** Algorithms implement only a quotient of I; proof speaks a richer language than hardware can realize.
- **Resource-bounded divergence:** With tight budgets, agents settle in different universality basins; equivalence is scale-contingent.

These are *epistemic signals*, not embarrassments; they identify missing hypotheses or the wrong invariant.

17.8 Methodological payoffs

- **Falsifiability with math:** Residuals and invariants make failure quantitative.
- **Reproducibility by structure:** Publishing I(symmetries, duals, topological type) plus translators makes results portable across labs and substrates.
- **Efficiency via duality:** Prices/potentials compress explanations and implementations simultaneously.
- **Safety through invariance:** Systems that preserve invariants (mass, charge, topology) are more predictable and auditable.

17.9 Philosophical upshot

Epistemology ceases to be a contest between rationalism (proof), pragmatism (computation), and empiricism (embodiment). It becomes a **coherence discipline**: establish the invariant, then show that theorem, program, and process are *the same map in different bases*. Where the maps commute, we have knowledge in the fullest sense—necessity demonstrated, construction given, and world agreeing.

18. Conclusion — mathematics as immanent grammar: found by minds, proven by models, enacted by matter

Claim. Mathematics is not elsewhere; it is the *grammar* already at work wherever composition, copying, selection, scarcity, and noise bind agents to coherent action. What we have called "the same math" is the recurrence of invariant structure—algebra of composition, symmetry and dual prices, topological type, variational extrema—across proof, computation, and embodiment.

Across the sections, a single motif repeats. When we fix the **invariants** first, three views align:

- **Proof** extracts necessity (universal properties, KKT/Noether, homology),
- **Computation** gives constructive witnesses (primal–dual loops, CRT-like decoding on tori, PDE solvers),
- **Embodiment** supplies semantic contact (soap films, resistor networks, swarm flows, crystalline groups).
  Where these views commute up to equivalence (isomorphism, categorical equivalence, duality, homeomorphism/homology), convergence is not coincidence but *structure*.

This yields an epistemology with teeth. **Compression parity** and **O(1) translators** mark alignment; **residuals** (PDE, KKT, equivariance) and **persistence** (homology stability) quantify it; **duality** and **universal properties** carry explanations across substrates. It also yields humility: our counterexamples—nonconvexity, metastability, friction, curvature—are not failures of mathematics but shifts of grammar.

Metaphysically, one can say these invariants are **discovered** (Platonism), **participated** (Logos realism), or **instantiated** (structural realism). Practically, all three recommend the same stance: state the invariants, prove them, program them, and check that matter agrees.

The centaur, then, is not an alliance but a *commuting diagram*: human proof, AI search, and nature's dynamics are three bases on one vector—the invariant. To do mathematics in this key is to speak the immanent grammar fluently: to let minds *find* it, models *prove* it, and matter *enact* it, until the three readings cohere and the world itself completes the argument.

References

1. Almgren, F. J. (2000). *Almgren's Big Regularity Paper: Q-valued functions minimizing Dirichlet's integral and the regularity of area-minimizing rectifiable currents up to codimension 2*. World Scientific.
2. Awodey, S. (2010). *Category Theory* (2nd ed.). Oxford University Press.
3. Beck, A., & Teboulle, M. (2003). Mirror descent and nonlinear projected subgradient methods for convex optimization. *Operations Research Letters*, 31(3), 167–175.
4. Borwein, J. M., & Lewis, A. S. (2006). *Convex Analysis and Nonlinear Optimization* (2nd ed.). Springer.
5. Born, M., & Wolf, E. (1999). *Principles of Optics* (7th ed.). Cambridge University Press.
6. Braides, A. (2002). *Γ-Convergence for Beginners*. Oxford University Press.
7. Bregman, L. M. (1967). The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, 7(3), 200–217.
8. Bubeck, S. (2015). *Convex Optimization: Algorithms and Complexity*. Foundations and Trends in Machine Learning.
9. Chaitin, G. J. (1966). On the length of programs for computing finite binary sequences. *Journal of the ACM*, 13(4), 547–569.
10. Chandra, A. K., Raghavan, P., Ruzzo, W. L., Smolensky, R., & Tiwari, P. (1989). The electrical resistance of a graph captures its commute and cover times. *STOC '89*, 574–586.
11. Chen, H.-L., Doty, D., & Soloveichik, D. (2014). Rate-independent computation in continuous chemical reaction networks. *ITCS '14*, 313–326.
12. Cohen-Steiner, D., Edelsbrunner, H., & Harer, J. (2007). Stability of persistence diagrams. *Discrete & Computational Geometry*, 37, 103–120.
13. Cook, M. (2004). Universality in elementary cellular automata. *Complex Systems*, 15(1), 1–40. (Rule 110)
14. Curry, H. B., & Feys, R. (1958). *Combinatory Logic, Vol. I*. North-Holland. (Curry–Howard roots)
15. Dal Maso, G. (1993). *An Introduction to Γ-Convergence*. Birkhäuser.

16. Dacorogna, B. (2008). *Direct Methods in the Calculus of Variations* (2nd ed.). Springer.

17. Delbaen, F., & Schachermayer, W. (2006). *The Mathematics of Arbitrage*. Springer.

18. Delaunay, B. (1934). Sur la sphère vide. *Izvestia Akademii Nauk SSSR, Otdelenie Matematicheskikh i Estestvennykh Nauk*, 7, 793–800.

19. de Silva, V., & Ghrist, R. (2007). Coverage in sensor networks via persistent homology. *Algebraic & Geometric Topology*, 7, 339–358. (Nerve-lemma intuition)

20. Dorigo, M., & Gambardella, L. M. (1997). Ant colony system: A cooperative learning approach to the traveling salesman problem. *IEEE Transactions on Evolutionary Computation*, 1(1), 53–66.

21. Doyle, P. G., & Snell, J. L. (1984). *Random Walks and Electric Networks*. MMA 22, Mathematical Association of America.

22. Edelsbrunner, H., & Harer, J. (2010). *Computational Topology: An Introduction*. American Mathematical Society.

23. Fiete, I. R., Burak, Y., & Brookings, T. (2008). What grid cells convey about rat location. *Journal of Neuroscience*, 28(27), 6858–6871.

24. Gelfand, I. M., & Fomin, S. V. (1963). *Calculus of Variations*. Prentice Hall.

25. Grünwald, P. D. (2007). *The Minimum Description Length Principle*. MIT Press.

26. Hafting, T., Fyhn, M., Molden, S., Moser, M.-B., & Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature*, 436, 801–806. (Discovery of grid cells)

27. Hahn, T. (Ed.). (2005). *International Tables for Crystallography, Vol. A: Space-Group Symmetry* (5th ed.). Wiley.

28. Harrison, J. M., & Kreps, D. M. (1979). Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory*, 20(3), 381–408.

29. Hasan, M. Z., & Kane, C. L. (2010). Colloquium: Topological insulators. *Reviews of Modern Physics*, 82(4), 3045–3067.

30. Howard, W. A. (1980). The formulas-as-types notion of construction. In J. P. Seldin & J. R. Hindley (Eds.), *To H. B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formalism* (pp. 479–490). Academic Press.

31. Jaromczyk, J. W., & Toussaint, G. T. (1992). Relative neighborhood graphs and their relatives. *Proceedings of the IEEE*, 80(9), 1502–1517.

32. Kosterlitz, J. M., & Thouless, D. J. (1973). Ordering, metastability and phase transitions in two-dimensional systems. *Journal of Physics C*, 6(7), 1181–1203. (KTHNY lineage)

33. Kleman, M., & Friedel, J. (2008). Disclinations, dislocations, and continuous defects: A reappraisal. *Reviews of Modern Physics*, 80(1), 61–115.

34. Kolmogorov, A. N. (1965). Three approaches to the quantitative definition of information. *Problems of Information Transmission*, 1(1), 1–7.

35. Lambek, J., & Scott, P. (1986). *Introduction to Higher-Order Categorical Logic*. Cambridge University Press.

36. Latty, T., & Beekman, M. (2011). Speed–accuracy trade-offs during foraging decisions in the acellular slime mould *Physarum polycephalum*. *Proceedings of the Royal Society B*, 278(1705), 539–545.

37. Li, M., & Vitányi, P. (2008). *An Introduction to Kolmogorov Complexity and Its Applications* (3rd ed.). Springer.

38. Lovász, L. (1993). Random walks on graphs: A survey. In *Combinatorics, Paul Erdős is Eighty*, Vol. 2 (pp. 1–46). János Bolyai Mathematical Society.

39. Mac Lane, S. (1998). *Categories for the Working Mathematician* (2nd ed.). Springer.

40. Mermin, N. D. (1979). The topological theory of defects in ordered media. *Reviews of Modern Physics*, 51(3), 591–648.

41. Mermin, N. D. (1992). Quasicrystals and the crystallographic restriction. *Reviews of Modern Physics*, 64(1), 3–49.

42. Mora, T., & Walczak, A. M. (2016). Quantifying lymphocyte receptor diversity. *Current Opinion in Systems Biology*, 1, 9–16. (Immune repertoire)

43. Morgan, F. (2009). *Geometric Measure Theory* (4th ed.). Elsevier/Academic Press. (Minimal surfaces, Plateau)

44. Nemirovsky, A., & Yudin, D. (1983). *Problem Complexity and Method Efficiency in Optimization*. Wiley.

45. Noether, E. (1918). Invariante Variationsprobleme. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 235–257.

46. Nourmohammad, A., et al. (2016). Host–pathogen coevolution and the emergence of broadly neutralizing antibodies in chronic infections. *PLoS Genetics*, 12(7), e1006171.

47. Peyré, G., & Cuturi, M. (2019). Computational optimal transport. *Foundations and Trends in Machine Learning*, 11(5–6), 355–607. (Sinkhorn, entropic OT)

48. Plateau, J. A. F. (1873). *Statique expérimentale et théorique des liquides soumis aux seules forces moléculaires*. Gauthier-Villars.

49. Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, 14(5), 465–471. (MDL origin)

50. Rockafellar, R. T. (1970). *Convex Analysis*. Princeton University Press.

51. Sethian, J. A. (1996). A fast marching level set method for monotonically advancing fronts. *Proceedings of the National Academy of Sciences*, 93(4), 1591–1595. (Eikonal numerics)

52. Shechtman, D., Blech, I., Gratias, D., & Cahn, J. W. (1984). Metallic phase with long-range orientational order and no translational symmetry. *Physical Review Letters*, 53(20), 1951–1953. (Quasicrystals)

53. Sreenivasan, S., & Fiete, I. R. (2011). Grid cells generate an analog error-correcting code for singularly precise neural computation. *Nature Neuroscience*, 14, 1330–1337.

54. Taylor, J. E. (1976). The structure of singularities in soap-bubble-like and soap-film-like minimal surfaces. *Annals of Mathematics*, 103(3), 489–539. (120° & tetrahedral angles)

55. Tero, A., et al. (2010). Rules for biologically inspired adaptive network design. *Science*, 327(5964), 439–442. (*Physarum* → transport networks)

56. Tishby, N., Pereira, F. C., & Bialek, W. (1999). The information bottleneck method. *Allerton Conference on Communication, Control, and Computing*, 368–377.

57. Toussaint, G. T. (1980). The relative neighborhood graph of a finite planar set. *Pattern Recognition*, 12(4), 261–268.

58. Villani, C. (2008). *Optimal Transport: Old and New*. Springer.

59. Voronoi, G. (1908). Nouvelles applications des paramètres continus à la théorie des formes quadratiques. *Journal für die reine und angewandte Mathematik*, 133, 97–178; 134, 198–287.

60. Walczak, A. M., Mugler, A., & Wiggins, C. H. (2012). Analytic methods for modeling stochastic regulatory networks. *Methods in Molecular Biology*, 880, 273–322. (Stochastic filtering in biology)

61. Wolfram, S. (2002). *A New Kind of Science*. Wolfram Media. (CA universality context)

62. Yoon, K., Buice, M. A., Barry, C., Hayman, R., Burgess, N., & Fiete, I. R. (2013). Specific evidence of low-dimensional continuous attractor dynamics in grid cells. *Nature Neuroscience*, 16, 1077–1084.

63. Gardner, R. J., Hermansen, E., Pachitariu, M., et al. (2022). Toroidal topology of population activity in grid cells. *Nature*, 602, 123–128.

64. Villani, C. (2003). *Topics in Optimal Transportation*. American Mathematical Society.

65. Delbaen, F., & Schachermayer, W. (1994). A general version of the fundamental theorem of asset pricing. *Mathematische Annalen*, 300, 463–520.

66. Soloveichik, D., Seelig, G., & Winfree, E. (2010). DNA as a universal substrate for chemical kinetics. *PNAS*, 107(12), 5393–5398.

67. Janot, C. (1994). *Quasicrystals: A Primer* (2nd ed.). Oxford University Press.

68. Senechal, M. (1995). *Quasicrystals and Geometry*. Cambridge University Press.

69. Peyré, G., Cuturi, M., & Solomon, J. (2016). Gromov–Wasserstein averaging of kernel and distance matrices. *ICML '16*, 2664–2672. (Metric-structure OT)

70. Daley, D. J., & Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes* (Vol. 1, 2nd ed.). Springer. (Point processes for swarms/collectives)

Note. Items are chosen to back the paper's core claims: equivalence up to structure (category theory, topology), compression (MDL/Kolmogorov), universality basins (computation), symmetry/duality (Noether/KKT/FTAP), variational media (minimal surfaces, optics, Laplacians), collectives and discrete geometry (*Physarum*, proximity graphs), grid codes and internal topology, immunity as coding, crystals/defects and group actions, and optimal transport.