The Mathematics and Physics of *Coupling-Gated Cortical Coherence*

A Comprehensive Plain-English Primer

Companion document to the manuscript and its Supplemental Material

Abstract

This primer explains, from the ground up, every piece of advanced mathematics and physics used in the manuscript and its Supplemental Material. It is written for readers with a general interest in the mind who want to follow the argument without prior knowledge of the specialist vocabulary or reading Lean 4 source or chasing down each cited result. Its parts explain how boundaries can arise, why finite memory loses information, what forgetting costs in heat, how an agent changes its future inputs, how rhythms can align, and how local descriptions might assemble into one state that can represent itself. They then explain the computer-checked proofs, simulations, brain-recording analysis, physical calibration and proposed interpretation in terms of experience. Explanations begin with the underlying idea and concrete examples before introducing notation. Most topics follow one rubric: what the object is, why the framework needs it, what is actually proved, and what is not. No proof is reproduced in full; two are sketched because their structure is unusual. The aim is literacy, not mastery.

This document is a companion, not part of the publication. Where it differs from the manuscript or the Supplemental Material, those are authoritative.

Contents

Part I. Orientation **2**

1 What the argument is, and where the mathematics sits **2**
1.1 Nine nodes, two thermodynamic branches 3
1.2 How to read this primer 4

2 Notation **5**
2.1 How to read the symbols 5
2.2 Symbols used in the framework 6

Part II. Topology, symmetry breaking, and boundaries **7**

3 Spontaneous symmetry breaking **7**

4 Topological defects: the Kibble classification **8**

Part III. Information, entropy, and the cost of forgetting **9**

5 Finite capacity **9**

6 Landauer's principle **10**
6.1 The Norton–Shenker distinction 11
6.2 One sum, two ways to compute it 11

Part IV. Memory, action and thermodynamics **13**

7 Mutual information and the data processing inequality **13**

8 The Still–Sivak–Bell–Crooks bound **13**
8.1 Deriving Still's bound from Landauer's . . . 14

9 Agents act on their future inputs **15**

10 Heat and entropy when an agent acts **16**
10.1 The two substeps share one intermediate law . . . 17
10.2 A reciprocal two-bit example . . . 18
10.3 Choosing a policy for a finite task . . . 19
10.4 A policy register that adapts . . . 19
10.5 Learning a task nobody told the register about . . . 21
10.6 A phase sensor can inform the learner . . . 22
10.7 Running another episode costs another episode . . . 22
10.8 A store you can carry, and one you can only report . . . 24
10.9 The source belongs in the budget . . . 24
10.10 What the measurement leaves behind . . . 25
10.11 Who pays for the clearing . . . 26
10.12 Who prepares the prior . . . 27
10.13 One register's ledger . . . 27
10.14 What a register's bath fixes, and what it does not . . . 28
10.15 How an action can set a scalar coupling amplitude . . . 29
10.16 How local actions install a coupling kernel . . . 30
10.17 How much installed energy the scalar coherent branch requires . . . 31
10.18 The learner pays to install its own coupling . . . 32

11 What does not follow: the free-energy reading **32**

Part V. Operators, gradients, and the squared-drift objective **33**

12 The squared-drift functional and its gradient **33**

Part VI. The Kuramoto model: dynamics **34**

13 The model **34**
13.1 Start with one clock hand . . . 34
13.2 Let the hands influence one another . . . 35
13.3 Add unpredictable timing jitter . . . 35
13.4 Measure how aligned the hands are: the order parameter . . . 36
13.5 Why an average arrow simplifies the interaction . . . 36

14 The Lyapunov potential and gradient descent **37**

15 The rotating frame **38**

16 Do the trajectories exist? Picard–Lindelöf, globally **39**

17 Where the trajectories go: Barbalat and Łojasiewicz **39**
17.1 The motion stops . . . 39
17.2 Where it stops . . . 40
17.3 The compactness objection, met without an invariant region . . . 40

18 The open gap: propagation of chaos **41**

Part VII. The synchronization transition **42**

19 The von Mises stationary density **42**
19.1 What becomes stationary? . . . 42
19.2 What is an ansatz? . . . 42
19.3 Why this particular density fits the model . . . 43
19.4 Order does not by itself cost dissipation . . . 43
19.5 Make the population reproduce its own average . . . 44

20 Why the threshold is $K_c = 2D$ **45**

21 The fold-and-cross proof of uniqueness **46**
21.1 First, what is a stationary branch? . . . 46
21.2 Why there cannot be two positive solutions at the same coupling . . . 47

Part VIII. The critical exponent and the square-root onset **48**

22 The exponent $\beta = 1/2$ **48**

23 From exponent to prediction **49**

Part IX. Finite populations, finite ramps, and what simulation adds **50**

24 Euler–Maruyama integration **50**

25 Ramp speed sets the apparent onset exponent **51**
25.1 What finite ramp speed means . . . 51
25.2 Crossing the threshold and observing escape are different events . . . 51

26 Spatial coupling and geometric frustration **53**

27 Squared-drift plasticity does not learn structure **54**

Part X. From discrete to continuum **56**

28 Building a coupling matrix from a field **56**

29 The negative result: coarse-graining produces no kernel **57**

Part XI. Sheaves, gluing, and the unity of a state **58**

30 What a sheaf is **58**
30.1 Assemble two overlapping maps . . . 58
30.2 The vocabulary for that assembly rule . . . 58

31 Compatibility is not phase agreement **59**
31.1 Agreement to within a tolerance . . . 60
31.2 Covers of content rather than of space . . . 61
31.3 Where the agreement comes from . . . 61
31.4 Three ways off the population count . . . 62
31.5 How soon can a patch respond? . . . 64

32 Where uniqueness comes from **65**

Part XII. Topological obstructions in the phase field **65**

33 Twisted gluing, and why it fails to add content **65**

34 Phase singularities and winding numbers **66**

Part XIII. Metric spaces, contractions, and the reflexive self **68**

35 Banach's fixed-point theorem **68**
35.1 An output's information ceiling 69
35.2 The error of a linear bottleneck 69
35.3 Approximate reconstruction 70

36 The metric on global sections, and why the obvious ones fail **70**
36.1 The germ–measure dictionary 71

37 The avatar: making the self-model run through a region **71**
37.1 Reconstructing more than one relevant state 72

Part XIV. Coupling support, atomless measures, and hardware **74**

38 Resource-matched coupling reallocation **74**

39 Point support and finite regions **75**

Part XV. Formal method: why the development declares no axioms **76**

40 What formalization is claimed to add **76**

41 Why a global axiom is the wrong shape for a physical postulate **76**

42 Witnesses, and what inhabitation does and does not prove **77**

43 The passive and active chain theorems **78**
43.1 A chosen physical model and its empirical interpretation 79

Part XVI. Circular statistics and the empirical test **79**

44 Doing statistics on angles **79**

45 Getting the phases: the analytic signal **80**

46 The parameter-free prediction, and the trap in testing it **81**

47 What the EEG analysis found **82**

48 The awakening experiment **83**

Part XVII. Dimensional analysis and the calibration problem **84**

49 Why the units do not match **84**

50 Why γ is not recoverable **84**

51 The empirical case for the field, and its critics **85**

Part XVIII. What the mathematics is asked to mean **86**

52 The commitment **87**

53 Motivations and the choices they require **87**

54 The costs **88**

55 What would count against it, and the observable that is missing **88**
55.1 Three components 89
55.2 Three ways it fails silently 89

56 Relation to other accounts **90**

Appendices **90**

A Master table: concept, role, and status **90**

B Glossary **94**

C Where to look next **98**

Part I. Orientation

1 What the argument is, and where the mathematics sits

The manuscript follows three requirements of its proposed account: coherence, compatibility and self-representation. Coupling can sustain collective timing; compatible local descriptions can form one global state; and an encoding within that state can reconstruct it. The second and third requirements need objects and assumptions beyond the phase model.

Two conditional mechanisms connect these requirements. Declared actuator modes have installation prices, so their stored energy bounds the mean coupling they can provide. For those modes, insufficient installed energy excludes a coherent stationary branch; the conversion depends on the hardware. Separately, patches that read phases through Lipschitz encoders shared on their overlaps have a content-disagreement bound determined by coherence. Their outputs can be vectors and vary across sites; overlapping patches must encode a shared site's quantity the same way. These results connect specified physical quantities; neither chooses a neural decoder or identifies a conscious state.

The cortex is the brain's outer neural sheet. Neurons communicate through electrical activity, including brief pulses called spikes; a population rhythm is repeated activity across a group of cells. EEG, or electroencephalography, records voltage differences at the scalp. The framework asks whether physical interactions can help these rhythms keep time together, and what more would be required for coordinated activity to carry one experienced situation.

The physical hypothesis concerns extracellular geometry and recovery after awakening. Extracellular means outside cells: the hypothesis concerns how the spaces between cells affect their electrical interaction. Coupling means how strongly one rhythm changes another's timing. The prediction first asks what happens when that strength is held fixed long enough for the statistics to settle; this is its *stationary* prediction. In recovery, coupling changes over time. Its rate of change is the *ramp speed*, and a system needs time to respond to that change.

There is also *nonidentifiability*: different causes can give the same measured pattern. Stronger interaction, less random timing jitter (*phase diffusion*), and a shared external influence (a *common drive*) can produce identical settled distributions of timing around a cycle. The test therefore needs phase dynamics and evidence distinguishing the interaction mechanism, as well as a concentration–coherence relation.

The representational argument asks what coherent activity carries. Here a phase is a position within a repeating cycle, like a clock hand's angle. Equal phases can coexist with incompatible spatial profiles: two regions can keep time together while assigning different values to a shared location. A plasticity control—a simulation that lets connection strengths change—preserves coherence while losing alignment with a specified template, or target pattern; that alignment statistic is distinct from overlap compatibility. The self proposal then asks whether the state assembled from local descriptions accurately represents itself. Identifying the relevant cover and encoding requires independent evidence. A *cover* is a collection of overlapping regions that together include the whole system; an encoding records its state, and a readout reconstructs a state from that record. Uniqueness within a specified cover and map does not choose a subject.

Thermodynamics studies energy, heat and the constraints on their conversion. Its results constrain a possible implementation of these requirements. Finite capacity and a symmetry-breaking boundary are independent starting objects, and their links to erasure, information and a continuum coupling require assumptions. Lean 4 checks two branches: one for passive predictive memory, and one for a specified process that acts on its environment. Each uses eight connecting hypotheses and supplied structures, and they share the passage from coarse-graining to the same fixed state. Their conclusion and its identification with experience have different statuses: the first is conditional mathematics, the second a philosophical commitment.

The main article develops coherence and the minimum cost of changing it, compatible con-

tent, and self-reconstruction together, then states what the composition connecting them establishes, listing the four edges E56–E89 it argues from. The full eight-edge chain, the installed-energy condition and the thermodynamic models are in the supplement. The article examines what their observables can establish before developing the cortical hypothesis and an awakening experiment. The supplement expands each of these in turn, with the detailed agency and resource models supporting that composition. This primer follows the mathematical prerequisites and explains those models individually. Their different stores, reservoirs and encoders should be identified only where a model explicitly connects them.

1.1 Nine nodes, two thermodynamic branches

The following table is a route map. A *node* is a claim; an *edge* is a proposed implication linking claims. A phase space is the collection of possible system states. An order parameter is a numerical summary of collective organization. A global section is one description assembled from compatible regional descriptions. A fixed point is a state a specified update leaves unchanged. These objects are developed with examples in the relevant parts. The branches share this route, with different thermodynamic content at n4:

n1	a finite phase space has bounded information capacity
n2	a broken symmetry forces a boundary, where the field leaves its vacuum manifold
n3	the boundary erases, and erasure costs heat
n4	passive: dissipation bounds nonpredictive memory; active: the named process's joint entropy and actual heat obey a budget
n5	discrete couplings coarse-grain to a continuum limit
n6	a specified physical field realizes that limiting coupling
n7	above threshold the field carries a unique coherent order parameter
n8	compatible local states glue to a unique global section
n9	that section is the unique fixed point of a self-prediction map

Each node is established or cited. Each *arrow* between consecutive nodes is an assumption. The passive branch labels them **E12** through **E89**; the active branch replaces E34 and E45 with **E34Active** and **E45Active**. Each branch therefore has eight connecting hypotheses:

Edge	Kind	What it asks for
E12	independent physical premise	a finite-capacity system realizes a field leaving its vacuum manifold
E23	bridge assumption	the boundary is a finite register whose update strands states
E34 / E34Active	bridge assumption	passive: predictive structure at the register's temperature and heat, with nonzero waste; active: physical transition assumptions and a heat allocation for the named process, the allocation itself derived from a ledger of one register's operations (§10.13)
E45 / E45Active	modelling assumption	passive: a convergent regime, supplied by named hardware holding a declared predictive system's own joint law; active: a scalar actuation law connects the process to regular spatial meshes (§10.15), or declared response modes carry a kernel on the product space (§10.16)
E56	physical commitment	the limiting coupling is realized by a specified physical field
E67	physical commitment	that field operates above the synchronization threshold
E78	bridge assumption	coherence drives a *given* cover to its relaxed equilibrium
E89	modelling assumption	a contracting map fixes the section glued on that cover

Two features of that table matter. First, the two *physical commitments* (E56, E67) are the empirically testable core, and they are bracketed by bridge assumptions on either side. These links connect distinct kinds of object; their ordering records the formal composition rather than a necessary physical sequence. Second, E12 is drawn as a premise rather than a gap because n1 and n2 are independent roots: finite capacity does not imply a symmetry-breaking boundary, and the composition merely accepts an implication between them.

The active branch matters because acting can create a correlation with the future: a controller helps determine the signal it will receive. Its bound concerns the joint controller–world entropy and the actual heat of that process. It does not reinterpret this newly created correlation as passive memory waste. Sections 9 and 10 explain the difference. Neither branch derives a learning rule or coupling convergence from a heat budget alone.

1.2 How to read this primer

The primer introduces vocabulary in the text, using ordinary examples before formal statements. The headings **What it is**, **Why the framework needs it**, **The statement** and **What it does not give** separate the idea, its purpose, the result and its limits. Worked walkthroughs sometimes need several steps between those headings. The notation table is a reference to return to, rather than a list of definitions to memorize first.

In Lean notes locate the computer-checked results. Lean is software that checks whether a proof follows from its stated assumptions; Mathlib is its library of mathematical definitions and theorems. Names in typewriter font are searchable source identifiers, not additional concepts to learn. A *hypothesis* is an assumption for a result; a *theorem* is a proved consequence; a *witness* is an example satisfying the assumptions; and a *counterexample* shows why a proposed stronger statement fails.

A structure or class is a record of named data and conditions. A field of that record is one entry, and an instance is an example meeting its conditions. To discharge a hypothesis means to supply its proof. The later formal-method part explains why packaging assumptions this way matters.

The part on formal methodology explains how physical postulates enter the conditional composition as assumptions on specified objects. This is the technical basis for keeping the manuscript's physical hypotheses distinct from its mathematical conclusions.

2 Notation

2.1 How to read the symbols

A function or map is a rule assigning an output to each input: $f(x) = x^2$ assigns 9 to 3. A scalar is a single number; a vector is an ordered list of numbers, pictured as an arrow when it has two or three components. A matrix is a rectangular table of numbers. A field assigns a value to each location, like a temperature map. A kernel $K(x, y)$ assigns an interaction strength to each *pair* of locations. The substrate is the set of locations being modelled. A functional assigns a number to a whole function or field, such as its total energy.

Sets and membership.
$x \in X$ means that x belongs to the set X; $|X|$ counts its elements when it is finite. $U \subseteq X$ means every member of U belongs to X. $U \cap V$ is their overlap; $U \cup V$ combines them. $\mathbb{R}$ denotes real numbers, $\mathbb{Z}$ integers, and S^1 a circle. An interval $[a, b]$ includes its endpoints; (a, b) excludes them.

Sums and averages.
$\sum_i x_i$ means add the listed values. $N^{-1} \sum_i x_i$ divides their sum by their count. An expectation $\mathbb{E}[X]$, also written $\langle X \rangle$, is an average weighted by the probability of each outcome. Absolute value $|x|$ ignores a number's sign; for a vector, length is written $\|x\|$ (or $|x|$ for a complex number).

Integrals and measures.
$\int f(x)\, d\mu(x)$ is a weighted total over locations. The measure μ specifies the weights: it can count sites or measure lengths, areas or probabilities. When μ has total weight one, the integral is an average. A density tells how that weight is spread out; its integral over a region gives the weight in that region.

Change and limits.
Δt is a time interval and Δx the change during it. dx/dt, or $\dot{x}$, is instantaneous change per unit time. $\partial f/\partial x$ varies x while holding the other inputs fixed. $x(t) \to L$ means the values approach L. $x \propto y$ means a constant multiple; $x \sim y$ means their ratio approaches one in the specified limit. $O(h^2)$ means an error bounded by a constant times h^2; $o(h^2)$ means an error whose ratio to h^2 tends to zero.

Exponentials and logarithms.
e^x is the exponential function, and $\log x$ or $\ln x$ is its inverse: $\log(e^x) = x$. A logarithm converts multiplication into addition. A square root $\sqrt{x}$ is the nonnegative number whose square is x.

Angles and complex numbers.
A full turn is 2π radians, a half turn π. The coordinates of a unit arrow at angle θ are $(\cos\theta, \sin\theta)$. The compact notation $e^{\mathrm{i}\theta} = \cos\theta + \mathrm{i}\sin\theta$ packages those two coordinates; $\mathrm{i}^2 = -1$. This imaginary unit is different from a subscript i labelling an oscillator. No physical quantity is imaginary.

2.2 Symbols used in the framework

Symbol	Meaning
θ_i	phase of oscillator i, an angle on the circle $[-\pi, \pi)$
ω_i	natural (intrinsic) frequency of oscillator i
K, K_{ij}	coupling strength; the coupling matrix. Inverse-time units
$K(x, y)$	continuum coupling kernel on a spatial substrate
D	phase diffusion coefficient (noise strength). Inverse-time units
K_c	critical coupling, $= 2D$ in this model
W_i	independent Wiener processes (Brownian motion)
r, r_N	order parameter: circular resultant length, in $[0, 1]$
ψ	collective phase (the direction of the mean resultant)
a	von Mises concentration parameter, $a = Kr/D$
I_n	modified Bessel function of the first kind, order n
$R(a)$	the Bessel ratio $I_1(a)/I_0(a)$
$\mathcal{S}(a)$	$\mathbb{E}_a[\sin^2\theta]$ under the von Mises law at concentration a
β	critical exponent of the order parameter; $= 1/2$ here
$\rho(\theta)$	stationary phase density
$V(\theta)$	Kuramoto Lyapunov potential
σ	squared-drift functional (*not*, in general, entropy production)
W_{diss}	dissipated work
$I(X;Y)$	mutual information
$I(X;Y \mid Z)$	conditional mutual information: dependence remaining when Z is known
X_t, S_t	internal and environmental states in the agency model
A_t, O_{t+1}	action and subsequent observation
Q, $\langle Q \rangle$	heat into the reservoir along one path and its mean
Σ	dimensionless entropy production of the finite path model; distinct from the squared-drift functional σ
$D_{KL}(P\|Q)$	Kullback–Leibler divergence
S, H	entropy (thermodynamic; Shannon)
k_B, T	Boltzmann constant; temperature
ν	fan-in of a communication graph: the most sites any one site hears from; distinct from the reset distribution ν of the Landauer ledger
k_BT	thermal scale of a register, in energy units
B	heat budget allocated to a named substep
$\mathcal{V}$	vacuum manifold (space of energy minima)
π_n	n-th homotopy group
$\mathcal{F}$	a sheaf; $\mathcal{F}(U)$ its sections over the open set U
μ	a measure on the substrate, normalized to total mass one; also a law on a finite register
Ξ	a configuration of the declared actuator hardware
$\mathcal{E}$	energy observable of the thermodynamic models
$\mathcal{K}_n$	mesh energy of a refining partition
$\mathcal{U}$	memory-update channel
$\mathcal{A}$	the avatar region an encoding restricts to
$\mathcal{C}$	content space: a real normed vector space
$\mathcal{T}$	timing sensitivity, the spike-time shift per unit field
E, R	encoder and readout; E is also the field amplitude in Part VI
γ	the rate calibration converting a field estimate to a coupling
Λ	contraction (Lipschitz) factor of a self-prediction map, < 1
τ	time interval over which the contraction rate is evaluated

Symbol	Meaning
λ	extracellular diffusion tortuosity (Part VI); also a kernel eigenvalue
$d_{\min}$	pairwise separation of a declared family of states
L	Lipschitz constant of an encoder or readout

Part II. Topology, symmetry breaking, and boundaries

3 Spontaneous symmetry breaking

Think of a landscape with two equally deep valleys. The landscape does not prefer left to right, but a ball resting in one valley has picked a side. Energy gives the landscape its height; a minimum is a lowest point. A local minimum is lowest nearby, while a global minimum is lowest among all allowed states. This example assumes the ball can lose energy so that it settles.

What it is. A physical system is described by an energy functional. A *symmetry* of that functional is a transformation—a rotation, a reflection, a sign flip—that leaves the energy unchanged. Symmetry is *spontaneously broken* when the energy functional has a symmetry that its lowest-energy states do not: the laws are symmetric, the outcomes are not. The standard toy model is the double well

$$V(v) = (v^2 - 1)^2,$$

which is invariant under $v \mapsto -v$ but has two minima, at $v = -1$ and $v = +1$. A ball in this potential must settle in one of them; whichever it picks, the reflection symmetry is gone from the outcome while remaining in the law.

The set of all lowest-energy states is the *vacuum manifold* $\mathcal{V}$. For the double well it is the two-point set $\{-1, +1\}$. In richer models it can be a circle, a sphere, or a more complicated space, and its shape is what governs everything in the next section. "Vacuum" here means lowest energy, not empty space. A manifold is a space that locally looks like an ordinary line, plane or higher-dimensional coordinate space; the traditional phrase also covers the simple set of minima in this example.

The spatial total depends on how locations are weighted. *Almost everywhere* means except on a set of zero weight under the chosen measure. For ordinary length on a line, an isolated point has zero weight. A *Dirac measure* instead puts all weight on one chosen point. In a two-point toy space, an energy average using that measure sees one point and ignores the other; minimizing the average need not minimize the ignored value.

Why the framework needs it. The framework's first move is that a conscious agent must be a *bounded* thing—something with an inside and an outside, separable from its environment. Rather than assuming a boundary, the framework wants one to be forced by physics. Symmetry breaking supplies the mechanism: if different regions of a substrate settle into different minima, something has to happen in between, and that something is a boundary.

The statement. Two results are proved, and the difference between them matters. The *pointwise vacuum* result says that a field globally minimizing the potential energy lies in the vacuum manifold *almost everywhere*—not everywhere. The almost-everywhere is not slack in the proof; it is necessary. On the double well over a two-point spacetime carrying a Dirac probability measure, the minimizing field sits at a vacuum value on the support of the measure and at 0 off it, so the pointwise version of the claim is false of the very field the theorem applies to.

What it does not give. No conserved quantity—one staying constant during motion—is constructed from this. Noether's theorem connects certain continuous symmetries to conserved quantities, but is not proved here. A symmetry group is a collection of transformations closed under combination and reversal; a metric measures distance. The symmetry group and the metric on the substrate appear in the definitions but are consumed by no theorem.

In Lean. `pointwise_vacuum_of_global_min` and `spontaneous_symmetry_breaking` in `Phase1_Primitives.lean`. The witness in `Examples` §11 instantiates the $\mathbb{Z}_2$ action $v \mapsto \pm v$ on the double well, proves the vacuum manifold degenerate, and proves the two minima exchanged by the symmetry. Unusually, its minimization hypothesis is *proved* rather than assumed—everywhere else in the development the existence of a global minimizer is a hypothesis, since no compactness or direct-method argument is formalized.

4 Topological defects: the Kibble classification

Topology studies what survives stretching and bending without tearing or joining. A connected space is in one piece; a connected component is one of its separate pieces. A continuous field has no abrupt jumps. If its value changes from -1 on one side of an interval to $+1$ on the other, it must pass through 0. If only -1 and $+1$ are energy minima, somewhere between the sides the field must leave the set of minima. That intervening region is the simplest example of a *defect*.

To describe more complicated defects, imagine drawing a loop on the space of allowed minimum values. Can it shrink to a point while staying in that space? A loop around a circle cannot; a loop drawn inside a filled disk can. *Homotopy* means continuous deformation. The symbols π_1 and π_2 classify deformations of loops and spheres respectively; π_0 records connected components. "Non-contractible" or "not null-homotopic" means that the relevant loop or sphere cannot shrink to a point there. The defect names below use the usual three-dimensional spatial setting.

What it is. Suppose a field takes values in the vacuum manifold $\mathcal{V}$ almost everywhere across some region of space. Different parts of the region may have chosen different vacuum values. Whether the field can interpolate smoothly between them depends on the *topology* of $\mathcal{V}$—specifically on its homotopy groups (Kibble, 1976):

- $\pi_0(\mathcal{V}) \neq 0$ ($\mathcal{V}$ is disconnected): **domain walls**. Two regions in different connected components cannot be joined without leaving $\mathcal{V}$ somewhere.
- $\pi_1(\mathcal{V}) \neq 0$ ($\mathcal{V}$ has non-contractible loops): **vortices** (line defects).
- $\pi_2(\mathcal{V}) \neq 0$: **monopoles** (point defects).

$\pi_0(\mathcal{V})$ is just the set of connected components—the crudest possible topological invariant, and the only one the framework uses.

Why the framework needs it. A domain wall separates regions choosing different minima. This supplies a candidate physical boundary for the framework. A forced wall need not enclose a bounded region: a plane dividing space also separates two domains. Further physical assumptions are needed to identify it with an agent's boundary.

The statement. The result taken is the π_0 case, and it is startlingly cheap:

> Let ϕ be a continuous field on a *connected* substrate, and let M be the candidate set of allowed minimum values within its value space. If $\phi(x_1)$ and $\phi(x_2)$ lie in different connected components of M—equivalently, if no preconnected subset of M contains both—then ϕ takes some value outside M.

The entire proof is: a continuous image of a connected set is connected. The image ϕ(substrate) is connected; if it stayed inside M it would be a connected subset of M containing points of two different components, which is impossible. So it leaves M. That exit point is the wall.

Note what is *absent* from the hypotheses: no energy potential, no Lagrangian, no symmetry group, no dynamics. Only continuity and connectedness.

An extension fills in values inside a region while retaining its prescribed boundary values. A contractible interior can be continuously shrunk to a point; a disk is an example. There is a second, differently directed result, the *extension obstruction*: boundary data that is not null-homotopic admits no continuous extension over a contractible interior. This says a defect *obstructs extension*; it does not say a defect must exist. The two directions are easy to conflate and the development keeps them apart deliberately.

What it does not give. Higher defects (π_1 vortices, π_2 monopoles) are recorded as available and not taken; the argument uses domain walls alone. And the theorem forces a *topological* separation. A register is a physical store of distinguishable states. E12 supplies a field realizing the boundary premise; E23 identifies that boundary with a register whose update erases information.

In Lean. `exists_notMem_of_no_common_preconnected` and `exists_notMem_of_connectedComponentIn` (`Phase1_Primitives.lean` §3), built on Mathlib's `IsPreconnected.subset_connectedComponentIn`. The witness (`Examples` §11.1) does not merely note that the double well has two minima—it *computes* the vacuum manifold, `wellVacuum_eq` proving DynamicalVacuum $V = \{-1, 1\}$ exactly. That matters: a larger vacuum set might have been connected, and the theorem would not apply. `wellVacuum_separated` then proves the disconnection in the form the theorem consumes, by observing that a preconnected subset of $\mathbb{R}$ spanning -1 and 1 must contain 0, which is not a minimum. Non-emptiness of the class of fields satisfying the hypotheses is checked rather than assumed: the clipped identity on $\mathbb{R}$ qualifies, and `kink_zero_notMem` locates its forced wall at the origin.

Part III. Information, entropy, and the cost of forgetting

5 Finite capacity

Suppose a device has four distinguishable states, labelled 00, 01, 10 and 11. It can record the answers to two yes/no questions, or two bits. Three such questions have eight possible answer strings, so four states cannot distinguish them all. This is a counting limit before it is a claim about heat, neurons or computation.

A probability distribution assigns each possible outcome a nonnegative weight, with all weights summing to one. A uniform distribution gives equal weight to every outcome. Entropy averages the surprise $-\log p$ of seeing an outcome of probability p: rare outcomes are more surprising. Using base-two logarithms measures entropy in *bits*; natural logarithms give *nats*. One unbiased yes/no answer carries one bit, or $\ln 2$ nats.

A map is *injective* if distinct inputs always give distinct outputs: no records merge. It is *surjective* if every possible output is reached: no state is missed. A *bijection* has both properties. The pigeonhole principle says more objects than boxes forces at least two into one box.

What it is. Shannon entropy $H(p)$ measures, in bits or nats, how much uncertainty a probability distribution carries—equivalently, how much information you gain on average by learning which outcome occurred. On a finite set X it is bounded:

$$H(p) \;\le\; \log |X|,$$

with equality exactly when p is uniform. A system with $|X|$ distinguishable states cannot hold more than $\log |X|$ nats, however cleverly it encodes.

A second, purely combinatorial fact does most of the work. If an environment delivers a stream of k perturbations drawn from an alphabet P, there are $|P|^k$ possible histories. If $|P|^k > |X|$, then by the pigeonhole principle the map from histories to internal states cannot be injective: two distinct histories leave the system in the same state. This holds *whatever* the update rule is, and even when every individual step of it is reversible.

Why the framework needs it. A bounded agent is bombarded by its environment. It cannot retain everything. Something is discarded, and the next section prices that.

What it does not give. Finite capacity is *assumed*, not derived—nothing here proves cortex has a finite effective register. And finite capacity does not imply that every update erases; it implies only that *some* history pair collides once the stream is long enough. The hypothesis is also fenced: on an *infinite* state space the corresponding step fails outright, as `Nat.succ` witnesses—it misses 0 yet is injective.

In Lean. `shannon_entropy_eq_log_card_iff` and `absorb_not_injective` in `Phase1_PhaseSpaceCapacity.lean`; the infinite-space refutation is recorded at `Examples` §19.

6 Landauer's principle

An eraser resets either 0 or 1 to 0. From the final register alone you cannot recover which value it held. A reversible operation, such as swapping 0 and 1, retains that distinction. *Logical irreversibility* is this loss of distinguishability in the update rule. Physical reversibility asks whether the complete device and its surroundings retain the information.

A heat bath is a large environment treated as staying at a fixed temperature. Heat is energy transferred through microscopic thermal motion; work is energy transferred through a controlled change, such as moving a piston. Dissipation is energy expenditure that is not recovered as useful work in the process. Thermodynamic entropy measures how many microscopic arrangements are compatible with a macroscopic description. The constant k_B converts information in nats to entropy units, and absolute temperature T is measured in kelvin.

What it is. Landauer (1961) observed that logically irreversible operations have a thermodynamic price. Erasing one bit of information—mapping two distinguishable states onto one—requires dissipating at least $k_B T \ln 2$ joules of heat into the environment. The reasoning: the microscopic laws are reversible, so phase-space volume is conserved (Liouville's theorem). If the system's own state space contracts, the volume has to go somewhere, and the only place available is the bath. Bath volume expanding means bath entropy rising, and $\Delta Q = T\Delta S_{\text{bath}}$ converts that to heat.

Why the framework needs it. This is what makes a bounded agent necessarily a *dissipative structure* rather than a passive one, and it supplies the thermal budget that the next part spends.

The statement. Three vocabularies for the same heat appear in the development, and they are deliberately connected rather than left side by side.

Combinatorial.

On a finite state space, a map that fails to reach every state is not injective (`is_erasure_of_not_surjective`). Under globally reversible system–bath dynamics, non-injectivity of the system update forces $\Delta Q > 0$ (`finite_phase_space_dissipates`, `landauer_from_reversibility`). A *second law* statement follows in the same file.

Measure-theoretic.

The same statement with continuous entropy and volume contraction in place of counting

arguments. The two are proved to coincide at counting measure on a finite register: `landauer_heat_bound_count_eq`.

Informational.
Wasted memory as erased entropy—see Section 8.1.

The bound is *not* a universal "$k_B T \ln 2$ per non-injective step". It is T times the entropy actually erased by the particular update; erasing one unbiased bit is the special case giving $k_B T \ln 2$.

What it does not give. The step from "the state space contracted" to "heat was dissipated" requires the thermodynamic postulate $\Delta Q = T\Delta S_{\text{bath}}$. That is carried as a class field, `StatisticalMechanics.heat_eq`, not proved—and `Examples` §1 exhibits a one-bit erasure system with an explicit bath discharging it, so the postulate is demonstrably satisfiable rather than merely asserted. Identifying such a register with a biological boundary is a further modelling step, and it is exactly what E23 carries.

6.1 The Norton–Shenker distinction

For example, a device can reset its visible bit while copying the old value into a hidden register. The visible operation merges two labels, but the complete outputs are still distinct. The heat accounting must include what happens to that hidden record. This is why the choice of physical state space matters to the principle.

What it is. The standard objection to naive readings of Landauer's principle (Norton, 2005; Shenker, 2000): *logical* irreversibility—a many-to-one map on abstract state labels—does not by itself entail heat. If the physical implementation is injective, nothing has been erased physically, and no heat need flow. A computation can look many-to-one at the level of the abstraction while the underlying device is perfectly reversible.

Why the framework needs it. Without this distinction the whole thermodynamic argument is a pun on the word "erasure".

The statement. The development defines erasure as non-injectivity on the *physical* state space, never on abstract labels. The distinction survives explicitly in the model: retaining an incoming perturbation is injective and free (`keepRecord_injective`), while refreshing the register that held it is not (`absorbStep_is_erasure`). In informational form the same distinction reappears as `nonpredictive_eq_zero_of_injective`: on a reversible update, the entropy erased is zero, so the bound forces wasted memory to vanish.

6.2 One sum, two ways to compute it

A *reduction* is a computation that reads many values and returns one summary of them: a sum, an average, a maximum. The mean-field coupling of this framework is a reduction—the coupling constant is the kernel averaged over both of its arguments, which on finitely many sites is a sum over pairs—so any implementation of the model performs one, whatever it is built from. The *register* is where the running total is kept. *Clearing* an array means overwriting it with a fixed value, which is what a working memory does when it moves on to the next job. Write $\mathcal{V}$ for the set of values one site can hold.

What it is. The previous subsection says a many-to-one *description* costs nothing and a many-to-one *device* costs heat. That leaves an obvious question about any particular device: did it have to be many-to-one? Take the sum over N sites and evaluate it twice. One evaluation accumulates the site values into the register and leaves the site array standing. The other computes exactly the same total and clears the array behind it. Both return the same number, and a reader looking only at the returned number cannot tell them apart.

Why the framework needs it. This is the analog energy-per-operation argument, which is usually made by gesture: a physical medium performs the sum by simply being in a state, where

a digital machine performs it by N accumulations and pays for each. The two evaluations are that argument with the gesture removed, and what they price turns out not to be what the gesture usually claims.

The statement. The charges are not close. Keeping the array is injective—read the array straight off the output, then subtract the total back out of the register—so it destroys exactly zero entropy and Landauer's bound charges it nothing at all. Clearing reaches only one array, the cleared one, so the reachable set has shrunk from the whole state space to a single register's worth of it, and the entropy destroyed is $N \log|\mathcal{V}|$: linear in the number of sites reduced, with a coefficient set by the alphabet each site is read to. The heat is at least temperature times that.

Stated that way it invites the Norton–Shenker reply one subsection earlier. The clearing evaluation erases because it was *defined* to clear, and the recording evaluation is the standing proof that nothing compelled it. The reply is closed by throwing both examples away and quantifying over implementations instead. Assume only that an evaluation leaves its results somewhere holding at most m states—a *memory budget*, which says where the answers may go and nothing whatever about how they are computed. The state space has not moved, so by pigeonhole everything the evaluation fails to reach is entropy it removed, and it destroys at least $\log(|X|/m)$. A map that reaches fewer states than the space holds cannot be surjective, hence on a finite space cannot be injective, hence is an erasure in the development's own sense; and the heat is at least temperature times $\log(|X|/m)$. The claim upgrades from *this implementation pays* to *every implementation inside a memory budget pays*, at a stated exchange rate.

Generality costs nothing here. Set the budget to one register's worth and the general bound returns exactly the $N \log|\mathcal{V}|$ the clearing evaluation was computed to owe, so it replaces the exhibited pair rather than standing weaker beside it.

What it does not give. This is a trade, not a barrier, and the trade is the whole of the content. Memory sufficient to retain every intermediate pays exactly zero—that is what the recording evaluation is—and letting the budget grow towards the whole state space sends the bound smoothly to zero. What is priced is the *exchange rate* between memory and dissipation, and the theorem closes neither end of it. Reversible computing is therefore a standing counterexample rather than an embarrassment: nothing here says a digital machine must take the clearing route, and a claim that it must would need an argument this development does not have.

Nor does any of it price a joule. The count is N times the logarithm of a site alphabet, so it says nothing until someone declares how a continuous field decomposes into site values and how many values a site carries. That is the same undetermined conversion factor as the κ of the installed-coupling argument, sitting in the same place, and a quantity linear in N with an undetermined coefficient is a scaling law and not an energy.

And the comparison is between two ways of evaluating one functional on one finite state space. It is not a comparison between two substrates, and it is certainly not a comparison between two ways of being conscious. Neither evaluation is claimed to be what any medium actually does.

In Lean. `Phase3_LandauerBridge.lean`, §4: `recordingSum` and `clearingSum` for the two evaluations; `recordingSum_injective` with `erasedEntropy_recordingSum` for the zero; `erasedEntropy_clearingSum` and `temperature_mul_le_heat_clearingSum` for the linear charge and its price, with `clearingSum_is_erasure` routing through `is_erasure_of_not_surjective`. The budget form is `erasedEntropy_ge_of_card_image_le`, with `is_erasure_of_card_image_lt` and `temperature_mul_log_le_heat_of_card_image_le`; `erasedEntropy_clearingSum_of_budget` checks it against the instance it generalizes.

Part IV. Memory, action and thermodynamics

7 Mutual information and the data processing inequality

Imagine X is a stored weather reading and Y is whether it is raining. If the reading always tells you the answer, it removes your uncertainty about Y. If the reading is unrelated to rain, it removes none. Mutual information measures this reduction on average; it concerns statistical dependence, not whether a message has meaning for a person.

The *joint law* P_{XY} lists probabilities for pairs of outcomes. The *marginals* P_X and P_Y describe each variable separately, obtained by adding over the other. Independence means the joint probabilities equal the products of those separate probabilities, written $P_X \otimes P_Y$. The Kullback–Leibler (KL) divergence $D_{KL}(P\|Q)$ averages $\log(P/Q)$ using probabilities from P. It is nonnegative, zero when the distributions agree, and can be infinite if Q rules out an outcome that P permits. It measures distributional discrepancy, but is not an ordinary distance: reversing P and Q can change its value.

What it is. Mutual information $I(X;Y)$ measures how much knowing X tells you about Y. Formally it is the KL divergence of the joint law from the product of its marginals:

$$I(X;Y) \;=\; D_{KL}(P_{XY} \,\|\, P_X \otimes P_Y)\,,$$

which is zero exactly when X and Y are independent.

The *data processing inequality* says information cannot be created by processing. If $X \to Y \to Z$ forms a Markov chain—Z depends on X only through Y—then $I(X;Z) \le I(X;Y)$.

Why the framework needs it. The passive branch splits what a system remembers into a part that predicts the future and a part that does not. The chain here is $X_t \to S_t \to S_{t+1}$: the system state X_t at time t was built from the environmental signal S_t, and the environment evolves by its own dynamics, unaffected by the system. Then $I(X_t;S_{t+1}) \le I(X_t;S_t)$: what the memory says about the future is at most what it says about the present. The difference,

$$\underbrace{I(X_t;S_t)}_{\text{total memory}} \;-\; \underbrace{I(X_t;S_{t+1})}_{\text{predictive memory}} \;=\; \textit{nonpredictive}\ \text{memory} \;\ge\; 0,$$

is well-defined and non-negative.

What it does not give. Here the arrows describe conditional dependence, not a causal claim that the memory generates the signal. A Markov kernel is a rule assigning probabilities to the next state for each current state. The Markov structure encodes a *no-back-action* assumption: the system does not act on its environment. This is not decoration. The development fences it with a counterexample—a Markov kernel on the pair that writes the system's state into the signal and thereby *raises* mutual information, breaking the inequality. Acting systems are treated by the separate feedback identity in Section 9; the passive assumption applies to the chosen variables and interval, rather than to every information-processing system.

In Lean. Mathlib carries no mutual information, so it is defined from `klDiv` in `Phase3_PredictiveThermodynamics.lean`; the inequality is `predictiveInfo_le_mutualInfo`.

8 The Still–Sivak–Bell–Crooks bound

Suppose a memory records two independent unbiased bits of the present, but only one helps predict the next signal. It carries two bits of memory, one bit of prediction, and one nonpredictive bit. Under the physical assumptions of this bound, the last bit contributes a lower bound

$k_B T \ln 2$ on dissipated work. This is a constraint on an implementation's cost; it does not tell the system which bit to keep or how to learn.

What it is. Still et al. (2012) proved that nonpredictive memory is thermodynamically expensive:

$$\boxed{W_{\text{diss}} \;\geq\; k_B T\,[\,I(X_t;S_t) - I(X_t;S_{t+1})\,]}$$

The work a system dissipates is at least the thermal cost of the information it holds about the present that is useless for the future. Memory that fails to predict is paid for in heat.

Why the framework needs it. Under a specified selection or optimization process, a cost for wasted memory could favor its removal. The framework's proposed connection to predictive structure is called *structural resonance*. The bound supplies a constraint that such a process would face; it does not supply the process. Retaining no information also meets a bound on memory waste, so useful prediction needs additional requirements.

The statement. The bound itself is a physical postulate—a class field, `PredictiveDissipation.still_bound`, not a theorem. Everything downstream *is* derived: the data processing inequality above; non-negativity $W_{\text{diss}} \geq 0$ (which follows rather than being assumed); the statement that a given dissipation budget bounds wasted memory; and the statement that a reversible quasistatic drive, dissipating nothing, leaves every retained bit predictive. Quasistatic means changing the controls slowly enough to pass through equilibrated states; slow operation alone does not guarantee zero dissipation in an arbitrary driven system.

A related exact result is also carried: the Kawai–Parrondo–Van den Broeck equality (Kawai et al., 2007),

$$\langle W_{\text{diss}} \rangle = k_B T\, D_{KL}(\text{forward} \,\|\, \text{reverse}),$$

from which `nonpredictive_le_arrow` follows: the irreversibility of the dynamics is at least the memory that fails to earn its keep.

The angle brackets in that equality mean an average over repeated realizations. "Forward" and "reverse" denote the distributions compared under a process and its appropriately reversed protocol. Their KL divergence quantifies how distinguishable the two directions are under the equality's assumptions.

What it does not give. This is the single most over-read result in the framework, and the manuscript is emphatic about it. The bound does *not*:

- specify a learning rule;
- maximize prediction;
- require the internal representation to reproduce environmental geometry;
- entail Bayesian inference or semantic content.

Bayesian inference updates probabilities in light of observations; semantic content concerns what a representation is about. Neither is supplied just by a numerical information-cost bound. Section 27 reports a simulation making this concrete: a coupling update that removes most of a squared-drift objective at the instant it acts maintains coherence while moving *away* from the structure it is expected to learn—further from it than from almost every relabelling of the nodes.

In Lean. `Phase3_PredictiveThermodynamics.lean`. The class is inhabited twice over one correlated two-bit law (`Examples` §18)—a frozen signal permitted to dissipate nothing, and a scrambled one forced to dissipate its entire memory—so the theorems about it are not vacuous.

8.1 Deriving Still's bound from Landauer's

A *bridge assumption* identifies quantities from two otherwise separate models. If wasted memory is no larger than erased entropy, a heat budget that pays for erasure also pays for that waste.

Without this comparison, two inequalities mentioning information and heat need not concern the same system.

What it is. Two thermodynamic charges have been introduced: Landauer charges a register for *destroying* information; Still charges a structure for *keeping* information it cannot use. They are the same heat viewed twice, and the development connects them.

The statement. `PredictiveDissipation.ofLandauer` builds a predictive structure out of a given register: its thermal scale is that register's temperature, its dissipated work is that register's Landauer heat, and Still's bound is *derived* from `landauer_bound` rather than postulated a second time. The single hypothesis is the identification

$$\underbrace{I_{\mathrm{mem}} - I_{\mathrm{pred}}}_{\text{memory wasted}} \;\le\; \underbrace{S(\mathrm{id}) - S(t)}_{\text{entropy erased}}.$$

What it does not give. That identification is a restriction, not bookkeeping. On an injective update the right-hand side is zero, so the inequality forces the waste to vanish—the Norton–Shenker distinction in informational form. The witness is the one-bit eraser, where a register destroying $\log 2$ pays for a memory wasting $\log 2$, with equality. Connecting the budget to Landauer heat therefore requires nonpredictive information to be bounded by erased entropy; the formal bridge derives the bound *under* that identification and does not derive the identification. That is E34.

9 Agents act on their future inputs

A weather recorder can observe rain without changing it. A controller connected to a lamp can change what it subsequently observes: if its internal bit says "on" and its action turns the lamp on, the future lamp state carries information about the controller. This correlation can arise even when the controller and lamp were initially independent. It reflects the effect of action on the future, which the passive Markov assumption excludes.

What it is. The agency model specifies four probability rules, called *channels*. A *policy* chooses an action from the current internal state; world dynamics responds to that action; an observation channel supplies a reading; and a memory update uses the reading and the sampled action. A policy here is a specified rule, not necessarily a learned rule or an optimal one. Writing $\sim$ for "is sampled according to", the channels are

$$\begin{aligned} A_t &\sim \pi(\cdot \mid X_t), & S_{t+1} &\sim K(\cdot \mid S_t, A_t), \\ O_{t+1} &\sim G(\cdot \mid S_{t+1}), & X_{t+1} &\sim \mathcal{U}(\cdot \mid X_t, A_t, O_{t+1}). \end{aligned}$$

Here K is a transition-probability rule, rather than the coupling strength of the oscillator model. The same sampled action is supplied to the world and the memory update. These equations describe one perception–action update; they do not assert that the system returns to its starting state.

Why the framework needs it. When actions influence the signal, the difference $I(X_t; S_t) - I(X_t; S_{t+1})$ can be negative. Calling it a nonnegative memory waste would impose the wrong constraint. The information accounting must allow the controller to create correlations with the future.

The statement. *Conditional mutual information* $I(X; Y \mid Z)$ measures the dependence between X and Y remaining when Z is known, averaged over the possible values of Z. For the

finite states considered here,

$$I(X_t; S_t) - I(X_t; S_{t+1}) = I(X_t; S_t \mid S_{t+1}) \\ - I(X_t; S_{t+1} \mid S_t).$$

The first conditional term asks what memory says about the present signal beyond what the future signal tells us. The second asks what memory says about the future beyond what the present signal tells us. Each term is nonnegative; their difference can have either sign. The passive model makes the second term zero and recovers its memory-waste formula exactly.

The comparison holds memory at its pre-action value X_t. The later internal state X_{t+1} belongs to the memory-update channel, and is not substituted for X_t halfway through the identity. Lean uses real subtraction for the signed difference, retaining negative values.

What it does not give. Conditional dependence alone does not identify a causal effect. That reading requires a causal model in which S_t includes the environmental state needed by the declared dynamics; a partial observation need not suffice. The four channels specify how a candidate agent operates. They do not determine its goal, explain how its policy was learned, or identify its internal variables with neural content. An information identity also supplies no heat cost until the physical transitions and reservoir are specified.

In Lean. `Phase3_Agency.lean` defines `Agency` and composes its channels. `Feedback.information_balance` proves the signed identity, and `passive_feedback_zero` and `passive_signedWaste` recover the passive case. The deterministic `writingAgent` writes its bit into an independent world bit, senses the result and updates memory; its signed difference is strictly negative.

10 Heat and entropy when an agent acts

To determine the cost of changing a lamp, a model must describe physical transitions and their relation to a reservoir. The finite thermodynamic model takes one elementary part of the update: the internal state x stays fixed while the world changes from s to s'. The transition probabilities may depend on x, so back action is permitted throughout.

What it is. Let $p(x, s)$ be the initial joint probability and $K_x(s' \mid s)$ the probability of the world transition. Adding over all possible starting world states gives the final law:

$$p'(x, s') = \sum_s p(x, s) K_x(s' \mid s).$$

A *path law* assigns probabilities to whole transition records, here the triples (x, s, s'). The model compares

$$P_F(x, s, s') = p(x, s) K_x(s' \mid s),$$
$$P_R(x, s, s') = p'(x, s') K_x(s \mid s').$$

The forward law starts at the initial distribution. The reverse benchmark starts at the final distribution and uses the same autonomous channel to make the return transition. Both laws have total probability one. Autonomous means that the transition rule is held fixed during this elementary update.

The statement. The KL divergence of these path laws is nonnegative. Expanding the logarithm of their ratio separates the joint entropy change from the transition ratios:

$$\Sigma = D_{KL}(P_F \| P_R) = H(p') - H(p) + \sum_{x,s,s'} P_F(x, s, s') \log \frac{K_x(s' \mid s)}{K_x(s \mid s')} \geq 0.$$

This is an identity derived from the specified probabilities. To interpret the last sum as reservoir entropy, a further physical assumption is needed: *local detailed balance*,

$$q(x, s, s') = k_B T \log \frac{K_x(s' \mid s)}{K_x(s \mid s')}.$$

It relates the relative likelihood of a transition and its reverse to the heat q entering a reservoir at temperature T. The mean heat $\langle Q \rangle$ is q averaged over the forward paths. Under this assumption the balance becomes

$$\Sigma = H(p') - H(p) + \frac{\langle Q \rangle}{k_B T} \geq 0.$$

Here H uses natural logarithms and Σ is dimensionless. This capital Σ belongs to the physical path model; it is distinct from the lowercase squared-drift functional σ used in the plasticity study.

Why the framework needs it. A positive thermal scale and an upper heat budget B yield the active node:

$$k_B T > 0, \qquad k_B T[H(p) - H(p')] \leq \langle Q \rangle \leq B.$$

Thus the joint entropy reduction is at most $B/(k_B T)$. The reduction itself need not be positive: the system's joint entropy may increase. This constraint uses the same process's distributions and actual heat, allowing action without substituting a signed information difference into a passive work bound.

What it does not give. The theorem assumes finite state spaces, strictly positive initial masses and transition probabilities, and this autonomous reversal rule. Zero-probability transitions or time-dependent driving need corresponding support and reversal conditions. This elementary theorem covers one substep with the controller fixed. The composition below includes the memory-changing substep, with the world fixed in turn. Additional storage or control operations require their own dynamics and accounting.

Work needs its own first-law relation. With q positive into the reservoir and w the work supplied to the joint system, that relation is $w = \mathcal{E}(x, s') - \mathcal{E}(x, s) + q$. Lean derives the corresponding average relation from this pathwise assumption. A mean heat bound alone does not identify work.

In Lean. `Phase3_AgencyThermodynamics.lean` builds the normalized laws and proves `FiniteFeedbackStep.entropy_balance`, `heat_bound` and `mean_first_law`. `Phase3_FiniteInformation.lean` relates finite probability sums to the same measure-theoretic KL divergence used for mutual information. `history_toAgency` identifies the thermodynamic paths with the history generated by the explicit action channels. `Chain.ActiveBound` retains the process, heat observable, thermal scale and budget.

10.1 The two substeps share one intermediate law

Let $p_0(x, s)$ be the initial joint law. Actuation gives $p_1(x, t) = \sum_s p_0(x, s) K_x(t \mid s)$. During sensing, the world t is fixed and memory changes from x to y through a channel $\mathcal{U}_t(y \mid x)$. This substep must start at $p_1(x, t)$, written in world–memory order as $p_1^{\mathrm{swap}}(t, x)$. Its output, returned to memory–world order, is

$$p_2(y, t) = \sum_x p_1(x, t) \mathcal{U}_t(y \mid x).$$

Lean proves that this is the law produced by the composed perception–action channels. The second initial law is constructed from the first final law; an independently chosen distribution cannot replace it.

At one common thermal scale $k_BT > 0$, add the two substep balances:

$$\Sigma_A + \Sigma_M = H(p_2) - H(p_0) + \frac{\langle Q_A \rangle + \langle Q_M \rangle}{k_BT} \geq 0.$$

The intermediate entropy cancels because a coordinate exchange preserves joint entropy. Neither p_1 nor p_2 has to be an equilibrium distribution. If the total mean heat is at most B, the complete update therefore satisfies $H(p_0) - H(p_2) \leq B/(k_BT)$.

Energy accounting uses one observable $\mathcal{E}(x, s)$ for both substeps, evaluated in the appropriate coordinate order. Under each substep's first law, with heat positive into the bath and work positive into the system, the same intermediate expected energy cancels:

$$\langle W_A \rangle + \langle W_M \rangle = \mathbb{E}_{p_2}\mathcal{E} - \mathbb{E}_{p_0}\mathcal{E} + \langle Q_A \rangle + \langle Q_M \rangle.$$

Strict positivity and local detailed balance carry the same roles as in the elementary theorem. Action and observation alias existing states; extra physical registers and time-dependent control need further accounting. The supplied upper heat budget does not identify its source in a register or establish the coupling-convergence assumption of the active chain.

In Lean. `FiniteFeedbackCycle` lives in `Phase3_AgencyThermodynamics.lean`. `cycle_law` identifies the output distribution; `entropy_balance`, `heat_balance`, `mean_first_law` and `entropy_budget` prove the complete update's balances and budget consequence.

10.2 A reciprocal two-bit example

Consider an internal bit x and a world bit s. In units $k_BT = 1$, agreeing bits have interaction energy zero and disagreeing bits have energy $\log 3$. Each elementary update redraws one bit, agreeing with the fixed other bit with probability $3/4$. The transition rule and heat are specified so that local detailed balance holds. This is a chosen theoretical model; no recording is needed to verify its conditional mathematical consequences.

Start with two independent unbiased bits, giving probability $1/4$ to each pair. Actuation uses action $a = x$ to update s. Afterwards each agreeing pair has probability $3/8$ and each disagreeing pair $1/8$. Each bit remains unbiased individually, while the pair has become correlated. The signed passive information difference is negative because the future world now carries information about the initially independent controller.

The actuation releases mean heat $(\log 3)/4$, with entropy production $\log 2 - (\log 3)/2 > 0$. Sensing then uses observation $o = s$ to redraw x by the same thermal rule. Individual internal states can change while this joint distribution is preserved; sensing has zero mean heat and entropy production. The composed channels give the same final law as this two-substep calculation.

The general composition also applies when memory starts biased. Under the same thermal channels, the intermediate law is then asymmetric and sensing changes the joint law, while actuation still releases positive heat. This example makes the coordinate swap observable: a second step initialized without it is incompatible even though all its probabilities are positive. The symmetric equilibrium law alone cannot detect that error.

Both substeps obey the first law with zero external work: their total mean heat comes from the decrease in mean interaction energy. The action and observation are aliases of the existing bits, with no extra physical storage register in this model. Separate devices carrying those signals would require their own accounting. Relaxing an initially nonequilibrium distribution is not a continuing energy supply. The preference for agreement is specified by the interaction; the model does not learn which relation to prefer.

In Lean. `Examples/AgencyCycle.lean` specializes the general balances and budget to the reciprocal model and rejects a mismatched intermediate law and a zero total heat budget. `Exam`

ples/AgencyThermodynamics.lean checks the actual transition law, nontrivial action and sensing, signed information, heat, entropy production, energy and work. ThermalAgency.agent_cycle_law checks the composed update and cycle_energy_budget checks its energy source.

10.3 Choosing a policy for a finite task

A goal can be supplied explicitly and policies compared against it. Consider the task of making a lamp true after one complete update. The internal bit is initially unbiased and independent of the lamp, which starts true with probability 1/4. A fixed world channel makes the lamp agree with the chosen action with probability 3/4. The memory channel then makes the internal bit agree with the new lamp with the same probability. Every policy faces these same channels and initial conditions.

There are four deterministic rules from a bit to a bit: always choose false, copy the internal bit, invert it, or always choose true. The baseline is always false. Performance is the final probability that the lamp is true, with a task target of at least 1/2. This objective is an input to the model.

The thermodynamic comparison uses the complete update. In units $k_B T = 1$, each transition's outward heat is its forward/backward log-ratio. Its work is the change in the common interaction energy plus this heat. Agreement has energy zero and disagreement has energy $\log 3$. This is the declared reservoir and work model for the compared policies.

Policy	Lamp success	Total mean heat	Total mean work
Always false	1/4	$(\log 3)/4$	0
Copy	1/2	$(\log 3)/4$	0
Invert	1/2	$3(\log 3)/4$	$(\log 3)/2$
Always true	3/4	$3(\log 3)/4$	$(\log 3)/2$

These are exact expectations proved from the finite transition laws. The initial mean interaction energy is $(\log 3)/2$ and its final value is $(\log 3)/4$ under every policy. Their difference accounts for the heat released at zero mean work; individual paths can still exchange work.

At a heat budget of $(\log 3)/4$, copying is the unique feasible optimum. It doubles success over the baseline and meets the task target. The baseline has the same mean heat and work yet misses the target: feasibility alone does not select successful behaviour. Always choosing true performs better but exceeds this budget. Copying and inversion even end with the same joint distribution, while inversion dissipates more. Computing costs from the endpoint alone would miss that difference.

More generally, a nonempty finite set of feasible deterministic policies has a maximizer for any supplied terminal reward. Strict improvement needs a better feasible policy, as this example provides. The theorem compares fixed policies for one update. The register model below supplies adaptation dynamics and their heat accounting; the comparison supplies no continuing power source or identification with cortex. The active chain's resource allocation is derived separately in §10.13. The scalar actuation model in §10.15 supplies a spatial connection through an additional constitutive law, and §10.16 supplies one through declared hardware instead.

In Lean. FiniteControlProblem in Phase3_AgencyThermodynamics.lean defines reward and heat on the actual FiniteFeedbackCycle and identifies its law with the explicit policy, world, observation and memory channels. exists_optimal proves finite feasible maximization. Examples/AgencyControl.lean proves the exact policy table, budgeted_improvement, unique_optimum and the counterexamples involving feasibility and equal outputs.

10.4 A policy register that adapts

Comparing policies does not say how a physical controller changes its policy. A small learning model makes that next step explicit. Add a separate register bit: false stands for the baseline lamp policy, true for copying memory. Both policies fit the task's heat budget. The question is how the probability of using copy changes, and what changing the register costs.

The model starts with the policies' expected rewards already known. At thermal scale one, assign the register energy $\log 3$ when false and zero when true. This energy is exactly $4 \log 3$ times the gap between copy's expected reward and the selected policy's reward. The objective is therefore built into the energy landscape. The model concerns adaptation toward that encoded objective; it does not estimate an unknown reward from experience.

At each learning update the register follows this fixed channel:

Current register	Next false	Next true
False	5/8	3/8
True	1/8	7/8

The ratio of the forward and reverse switching probabilities is three. Consequently, each transition's forward/backward log-ratio equals the energy lost by the register. That loss is the outward heat in this reservoir model, and the pathwise first law gives zero external work during relaxation.

Let p_n be the probability that the register holds copy after n updates. Each update starts from the preceding update's actual output. With $p_0 = 1/4$,

$$p_{n+1} = \frac{3}{8} + \frac{1}{2} p_n, \qquad p_n = \frac{3}{4} - \frac{1}{2}\left(\frac{1}{2}\right)^n .$$

The expected reward of the policy it holds is then

$$J_n = (1 - p_n)\frac{1}{4} + p_n \frac{1}{2} = \frac{7}{16} - \frac{1}{8}\left(\frac{1}{2}\right)^n .$$

This expectation increases at every finite update. The distribution converges to probability 3/4 of copy, and that distribution is stationary under the same channel. Its expected reward 7/16 remains below copy's 1/2 and below the task target. Convergence need not mean optimal learning.

The resource accounting uses those same changing probabilities. Update n releases mean heat $(\log 3)(1/2)^n/4$, and the total through N updates is

$$Q^L_{<N} = \frac{\log 3}{2}\left[1 - \left(\frac{1}{2}\right)^N\right].$$

This equals the decrease in the register's mean energy and stays below $(\log 3)/2$. The finite entropy balances telescope as well. The initial preparation contains the energy that is released; the model supplies no continuing power source.

The task has a different clock. J_n evaluates how well the held policy would perform on the lamp task's prepared initial law, independently of the register. Both policies have task heat $(\log 3)/4$ and zero mean task work, so every mixture retains those expectations. The learning updates themselves execute no lamp trials. Preparing either initial law and implementing the register's readout in an actuator need their own physical accounting. A finite total for register relaxation cannot pay for arbitrarily many task episodes. Section 10.5 takes up the separate question of a register whose values are not supplied in advance.

Two checks delimit the improvement claim. A register holding copy can switch to the baseline with probability 1/8, lowering that trajectory's reward. And the same channel, started

with copy probability 7/8, lowers expected reward because it relaxes toward 3/4. The proved improvement depends on the chosen initial preparation as well as the objective and channel.

In Lean. `FiniteFeedbackStep.iterate` and its summed entropy and first-law theorems live in `Phase3_AgencyThermodynamics.lean`. `Examples/PolicyLearning.lean` identifies energy with the supplied reward, proves the recurrence from the actual channel, and checks reward improvement, the stationary limit, transition heat and cumulative energy loss. Its regressions retain reward-decreasing paths, reject a reset initial law and show expected reward decrease for the alternative preparation.

10.5 Learning a task nobody told the register about

The model above starts from the answer: the energy landscape is built from rewards that are already known. The obvious objection is that a learner is precisely a thing that does not know them. So remove the supply and see what survives.

Which of two actions pays is an environmental bit. The learner's register bit is read out directly as the action it executes. The world answers each action with a noisy outcome: success with probability 3/4 if the action taken was the rewarding one, 1/4 otherwise. The update rule is win-stay, lose-resample — keep the register on success, redraw it uniformly on failure.

What makes this a different model, rather than the previous one relabelled, is where the fences sit. They are in the types. The update has the shape observation → register → register law, and the readout the shape register → action. Neither has a slot for the environmental bit, so neither can consult it, and no expected reward can be decoded into the action. The reward function is used to score the process and appears in none of its channels.

Acting, observing and updating compose into a single channel on the joint law of parameter and register, which is the same elementary bipartite update the rest of this section uses. At $w = 1$ it is

Current register	Next 0	Next 1
0	5/8	3/8
1	1/8	7/8

and its mirror image at $w = 0$. Every mass of the actual joint law is available in closed form from the uniform prior: each pair with $r = w$ carries $3/8 - \frac{1}{8}(1/2)^n$ and each pair with $r \neq w$ carries $\frac{1}{8} + \frac{1}{8}(1/2)^n$. The expected reward of the action *actually executed* is therefore

$$J_n = \frac{3}{4} - \frac{1}{4}\left(\frac{1}{2}\right)^n,$$

rising strictly from 1/2 and never reaching certainty.

Two baselines say what did the work. Freeze the register and it stays at 1/2 forever. Keep the update rule but make the world uninformative — success with probability 1/2 whatever is done — and it also stays at 1/2, and spends nothing doing so. The improvement is bought by the observations, not by the update rule. The information those observations carry is conditional on the action: at a fixed register value the outcome distribution differs between the two environments, while at the prior the outcome's marginal says nothing at all.

The accounting has a feature the previous model does not. Cumulative heat is

$$\sum_{n<N} \langle Q_n \rangle = \frac{\log 3}{4}\left[1 - \left(\frac{1}{2}\right)^N\right] \leq \frac{\log 3}{4},$$

and because the register's two states are energetically degenerate, all of it is work supplied from outside. That is not an artefact of choosing a flat energy. A register energy that made these updates free would force the drive's heat to be the same at both values of the environmental

bit — and a drive that behaves identically in the two environments cannot be what tells the register which one it is in. Learning from experience is paid for in work; relaxation toward an encoded objective is not. The entropy the learner removes fits inside the same $(\log 3)/4$.

Two more checks delimit the claim. The register leaves the rewarding action with probability $1/8$ at every update, so expected improvement is not improvement along every path. And each action is executed exactly half the time at every horizon: the learner acquires no preference between the actions, only a correlation with the parameter. What is still outside the model is preparing the prior, supplying the work, running task episodes beyond the declared updates, building the readout, and any claim that these variables are cortical ones.

In Lean. `FiniteObservationalLearner` and its generic results live in `Phase3_ObservationalLearning.lean`: `zero_work_needs_parameter_independent_heat` is the obstruction above, and `cumulative_entropy_budget` telescopes the entropy balance. `Examples/ObservationalLearning.lean` proves every mass of the law, the reward formula and its limit, the heat and work sums, and the improvement over both baselines. Its regressions retain a reward-decreasing transition, an unchanged action marginal, a conditional-only informativeness of the outcome and the rejection of a reset law.

10.6 A phase sensor can inform the learner

What it is. A sensor needs a rule converting activity into an observation. Here it samples one oscillator uniformly and converts its phase into a symbol. If three of four oscillators read as success, the probability of observing success is $3/4$. The observation law is computed from the phase configuration and readout.

The statement. In the positive example, a matching action produces three phases at zero and one at π; a mismatching action reverses the counts. Reading zero as success produces exactly the observational learner of the preceding subsection. A control always has three zeros and one π, regardless of the task's unknown parameter or executed action. Its observations carry no task information and its reward stays at $1/2$.

Both populations have the same coherence, $r^2 = 1/4$, in every case. The informative one learns because its phase configuration depends on the task in a way its readout can detect. A constant readout destroys that advantage even with the informative phase configuration unchanged.

What it does not give. The phase law and sampling hardware are supplied. No phase dynamics produces the configurations, and greater coherence is not shown to improve learning. In general the channel distinguishes two parameters exactly when some readout symbol occurs a different number of times at the executed action.

In Lean. `Phase3_PhaseSensor.lean`: `PhaseSensor.channel`, `channel_ne_iff_fiberCount_ne`, `FiniteObservationalLearner.ofPhaseSensor`. `Examples/PhaseSensor.lean`, §27, proves equality with the existing learner and the matched-coherence control.

10.7 Running another episode costs another episode

A learner's register can release only a finite amount of energy while its decoded policy improves. Running a task again also involves acting on the world and preparing it for reuse. A continuing finite-agent model makes those operations explicit on one evolving law.

The state has three bits: an unknown rewarding action, a register that selects the action, and a task flag. The cycle has three stages. Acting moves the flag according to the register's action. Learning reads that flag through a noisy observation and writes the register. Reset returns the flag to a standard probability distribution for the next episode. The coordinate not being driven undergoes noisy drift. Each output is the next input; no stage receives a freshly prepared copy of the task.

The register, observation and update cannot directly consult the unknown parameter. The actuator's response depends on whether the chosen action is the rewarding one, so the observation can convey useful information. Reward scores what happens and enters none of the channels. Starting with three uniform bits, the register agrees with the unknown rewarding action with probability 65/128 after the first complete cycle. Later cycles obey

$$a_{m+1} = \frac{65}{128} + \frac{a_m}{128}.$$

Thus the acquired advantage remains above chance at every completed cycle of the mathematical protocol. The flag returns to probability 1/4 after each reset; the register's changing correlation with the task carries over.

That recursion has a closed solution. Its fixed point is 65/127, and the distance to it shrinks by a factor of 128 per cycle:

$$\frac{65}{127} - a_m = \left(\frac{65}{127} - \frac{65}{128}\right)\left(\frac{1}{128}\right)^m.$$

So this learner's expected performance climbs to 65/127 and stops there, short of the certainty an optimal player would have. Two cautions come with that number. It is an expectation over completed cycles, not a statement about any single run; and it belongs to these particular channels, which nothing here compares with the best available policy. The cost bound below adds a third: reaching the limit takes unboundedly many cycles, and each one costs a fixed positive amount, so no finite allowance pays for the convergence.

Now give the process an initial work allowance B. The remaining allowance after N stages is $B - W_{<N}$, where $W_{<N}$ is expected cumulative work. A run fits that allowance only if it stays nonnegative at every earlier stage too. For one common energy and temperature, the first law and entropy balance add across the actual sequence:

$$k_B T[H(p_0) - H(p_N)] \le Q_{<N} = W_{<N} + \langle \mathcal{E} \rangle_0 - \langle \mathcal{E} \rangle_N.$$

The store and the system's energy loss therefore bound the joint entropy the run can remove. If each stage costs at least a positive amount, a finite store funds only finitely many stages.

The bit witness assigns zero energy to every joint state, so heat is paid entirely by work. Heat is the transition log-ratio at thermal scale one for each composite stage. The first cycle costs $(15/32)\log 3 + (1/16)\log(5/3)$, including reset, and every cycle costs at least $(3/16)\log 3 + (1/16)\log(5/3)$. An allowance $4\log 3$ covers every stage of the first cycle and cannot cover 22 complete cycles. This bound does not locate the exact stage at which the allowance runs out.

The controls explain the roles of observation and preparation. Make the observation independent of the flag and performance remains at chance, with no learning-stage heat. Replace reset by idle drift and some learning survives, but the flag fails to return to its standard law and performance in the second cycle is lower. The reset has a cost and an effect on the next action.

What it does not give. The allowance bounds expected work at each prefix, not a battery on every sample path; §10.8 carries the store on the trajectory instead. The initial law and store are supplied resources. Reset prepares later episodes; preparing the original prior is outside this accounting. The observation is marginalized inside the learning channel, so its local detailed balance does not derive a separate physical sensor memory or its erasure; §10.10 supplies that register. Battery dynamics, fabrication and cortical identification require further models.

In Lean. `FiniteProtocol`, `ContinuingProcess` and `ContinuingAgent` live in `Phase3_ContinuingAgent.lean`. The summed first law and entropy balance lead to `entropy_reduction_le_stored` and `horizon_le_of_cost`. `Examples/ContinuingAgent.lean` proves the cycle law, persistent reward, stage costs, a funded first cycle, finite-store exhaustion and the observation/reset controls.

10.8 A store you can carry, and one you can only report

The allowance above is a number outside the process: it is compared against an average of the work, so what it bounds is an average. That is weaker than it sounds. A store can cover its expected draw at every stage and still be overdrawn on a run that actually happens.

Make the store part of the state instead. The state now carries a reading β, and each transition the process can execute draws w^- from the store and receives a declared supply w^+, with

$$\beta(\text{after}) = \beta(\text{before}) - w^- + w^+.$$

The condition to check is that the draw is covered where it is made: $w^- \le \beta(\text{before}) + w^+$ on every transition of positive probability out of a state the process reaches. If it holds at every stage, and the reading starts nonnegative, then *no state the process ever reaches* carries a negative reading. Nothing here needs every transition to have positive probability, so a deterministic gate — the kind §10.14 uses — is allowed, and the draw itself may be random.

Averaging the same identity over the executed paths returns the earlier bookkeeping: mean reading in, less work drawn, plus work supplied. So the pathwise statement implies the expected one. The converse fails, and one witness shows how. A store reading 1 either idles or spends a unit, with equal probability; the same from a reading of 0. Over two stages it draws 1 in expectation against an allowance of 1, so the expected allowance is never exceeded — and the path that spends twice, which has probability 1/4, ends at -1. The allowance was never a battery.

That leaves the question of what keeps a process running. Two conditions bracket it. If the supply covers each draw as it is made, every horizon is solvent and cumulative work is unbounded without difficulty. If instead every executed transition draws at least $c > 0$ more than it supplies, then a run solvent at N stages satisfies $Nc \le \beta_0$: the process stops. Between them sits the interesting case. A cycle that spends a unit and is then given it back satisfies neither condition, is solvent at every horizon, and has drawn more work by some horizon than any allowance you care to declare. Its supply is doing the work: remove it and the reading no longer records what happens to the store at all.

What it does not give. The generic supply is an input. The next subsection accounts for a finite source behind it, and §10.10 for the measurement's own register. Preparing the initial resources and every cortical identification remain distinct tasks.

In Lean. `PathwiseStore` and its predicates `Ledgered`, `Funded` and `Solvent` live in `Phase3_ContinuingAgent.lean`, with `solvent_of_funded`, `mean_balance_eq`, `totalDraw_le_of_solvent`, `solvent_forall_of_replenished` and `horizon_le_of_net_cost`. `Examples/PathwiseStore.lean` carries the gambler and the charger, including `sustained_not_solvent`, `charger_draw_unbounded` and `unsupplied_not_ledgered`.

10.9 The source belongs in the budget

Recharging a buffer moves the resource question to whatever charges it. Put that source in the state too, with usable energy R, and check that each transfer into the buffer is energy lost by the source:

$$R(\text{after}) = R(\text{before}) - w^+.$$

Add this to the buffer's ledger. The transfer w^+ cancels, leaving

$$(\beta + R)(\text{after}) = (\beta + R)(\text{before}) - w^-.$$

If both resources stay nonnegative, the total work drawn cannot exceed their initial combined energy in expectation. If every executed stage before N draws at least $c > 0$ and the initial combined reading is at most b, then $Nc \le b$. A direct transfer of one unit followed by idling

saturates the bound. Only the stages of that finite run need have positive cost. Moving the source inside the account does not create more energy.

Energy being available is different from its arriving when requested. The finite witness has a source, a buffer and an output load, each storing zero or one unit, and a selector bit that stores no energy. Initially the source is full, the buffer and load empty and the selector uniform. If the selector is on, the first gate swaps source and buffer; if it is off, nothing moves. The second gate swaps buffer and load. Both gates are reversible and conserve the total energy. Half the runs deliver one unit to the load, and half leave it in the source. The mean work actually delivered is $1/2$.

Three checks make the resource claim precise. Reporting a unit delivered on every run violates the buffer's ledger. Preparing the source empty gives zero delivery with the same gates. And the preceding subsection's charger, which keeps drawing without bound, cannot have any nonnegative finite source coordinate whose loss pays for its supply. Its continuing input has to cross that finite resource boundary.

What it does not give. This is one uncertain transfer; the selector is prepared once and is carried through both gates. Subsequent stages idle. The work is energy deposited in an ideal output load, and no heat bath is introduced. Preparing the source and selector, controlling the gates, externally refuelling the source and identifying this hardware with the learning agent's actuator remain separate modelling tasks.

In Lean. `SourceLedgered`, `withSource_ledgered`, `totalSupply_eq_source_loss`, `totalDraw_le_initial_resources` and `horizon_le_of_finite_source` belong to `PathwiseStore` in `Phase3_ContinuingAgent.lean`. `Examples/FiniteSupply.lean` proves the gates' reversibility and energy conservation, `law_eq_coin`, `draw_two`, `undelivered_mass`, `forced_not_ledgered`, `empty_source_draw` and `charger_no_finite_source`.

10.10 What the measurement leaves behind

In the agent of §10.7 the observation is used and thrown away inside a single channel. A real measurement leaves a mark: a register that now holds something about the world and will hold it until something clears it. Clearing it is an operation like any other, and it costs.

Give the state a memory coordinate and run four operations in order. The readout drives the actuator. A measurement writes the memory from the world the action moved. The register is updated from the memory — only from the memory, which is a fact about the update's type, not a promise about its behaviour: two worlds leaving the same memory drive the register identically. Then a clearing operation pushes the memory toward a declared standard law ν while the world is prepared for the next episode.

The clearing operation moves three coordinates at once, and its transition log-ratio splits exactly into their three own. So there is a well-defined answer to what clearing the memory cost, rather than a leftover. When the erasure lands on ν whatever it is given, that answer is

$$\langle Q_M \rangle = k_B T \, [H(\mu) - H(\nu)] + k_B T \, D_{KL}(\mu \| \nu),$$

where μ is what the memory held. The first term is Landauer's: the entropy you removed. The second is the price of using one fixed channel on a law it was not built for. Gibbs' inequality says the second term is nonnegative, which recovers the usual bound — but the identity says more than the bound does. Set $\mu = \nu$ and the cost is zero: a memory already standard is free to clear. Make μ *more* ordered than ν and the cost is negative: the operation takes heat out of the reservoir. Clearing dissipates because of what the memory holds, not because clearing is clearing.

The bit witness pins this down without ever computing the joint law. Every channel mass is $1/4$ or $3/4$. Whatever the agent has learned, the flag it leaves is a mixture of the actuator's two masses, so it lies in $[1/4, 3/4]$; the measurement mixes that into the memory and the drift

moves it halfway to uniform, so the memory's mass at each clearing lies in $[7/16, 9/16]$. With $\nu = (3/4, 1/4)$ the cost is $[3/4 - \mu(0)] \log 3$, hence at least $(3/16) \log 3 > 0$ at every cycle. Learning does not make the measurement free. In the first cycle the memory is exactly uniform, and the two terms are $(3/4) \log 3 - \log 2$ and $(1/2) \log(4/3)$, adding to $(1/4) \log 3$. The first is positive for the reason that $16 < 27$.

Two controls say what the cost is and is not about. Blind the measurement — let it record a coin instead of the world — and the memory's marginal is unchanged, so the heat is identical. You pay to clear a register that learned nothing, exactly as much as one that learned something. Replace the clearing by the symmetric drift and you pay nothing, and the measurement survives into the next episode; the erasure returns the standard state whatever was written.

What it does not give. The four operations are stages of one protocol, so the cumulative entropy balance of §10.7 charges the whole run including the clearing. Local detailed balance at the memory's reservoir is an input, as everywhere here. Gate fabrication and cortical identification remain separate. Section 10.11 supplies the source these operations draw on; §10.12 accounts for reaching the declared prior.

In Lean. `memoryHeat`, `memoryHeat_const`, `memoryHeat_ge_entropy_drop` and `MemoryAgent` live in `Phase3_SensorMemory.lean`, with `learn_register_independent`, `memoryLaw_clear`, `memoryLaw_record`, `clear_stageHeat_eq` and `clear_memoryHeat_const`. `Examples/SensorMemory.lean` proves `erase_cost`, `memMass_clear_bounds`, `clear_cost_floor`, `first_cost_decomposes`, `blind_same_cost`, `idle_clear_free` and `over_ordered_absorbs`.

10.11 Who pays for the clearing

The clearing costs. Nothing so far says where the energy comes from. The store of §10.8 is carried on the trajectory and the source of §10.9 is a separate three-bit witness; neither runs on the agent that does the clearing. Put them on one protocol: let the world the agent acts on carry a charge as well as a task flag, and let the operation that prepares the next episode spend a unit of it.

The horizon bound of §10.8 does not reach this agent as it stands, because it asks for a positive net draw at *every* executed transition and this agent draws at one operation in four. The fix is to let the cost depend on the stage. A charge paid once in each period of p operations sums to mc over m periods, however it sits inside the period, so a run that is solvent through m cycles and starts no higher than b satisfies

$$mc \leq b.$$

A finite source funds finitely many clearings. In the witness the charge is three units and each clearing spends one, so three cycles are funded, the bound is attained, and the reading is below empty at the sixteenth operation. The memory channels are the ones of §10.10, so every clearing still costs between $(3/16) \log 3$ and $(5/16) \log 3$: what the source funds is what that section prices.

Two things the model does not do are worth stating, because both are properties of the model rather than of this witness. First, the erasure is a channel of the memory alone — that was the fence that made its cost well defined — so it cannot be conditioned on the charge. Past the third cycle the agent keeps clearing, keeps costing at least $(3/16) \log 3$, and draws nothing. The model says the run is unfunded, not that it stops. Second, an agent whose channels all have full support cannot be funded at all: if every transition is possible then a reading that never rises is squeezed between every pair of states in both directions, so it is constant and every net draw is zero. The witness's charge moves deterministically for that reason, not for convenience.

What it does not give. The unit the clearing draws is a declared reading of the agent's state. It is not identified with that stage's log-ratio heat; doing so needs the reservoir's own energy

as a coordinate, as in §10.14. Preparing the charge, refuelling it from outside, fabricating the gates and any cortical identification are separate. The preparation model in §10.12 requires a target with full support.

In Lean. `balance_le_of_stage_cost`, `sum_period_indicator`, `horizon_le_of_periodic_cost`, `periodic_horizon_le_of_finite_source` and `net_draw_eq_zero_of_positive` live in `Phase3_ContinuingAgent.lean`; `MemoryAgent.drawnStore`, `clearings_le_of_source` and `positive_source_never_falls` in `Phase3_SensorMemory.lean`. `Examples/FundedMemory.lean` proves `charge_invariant`, `clear_spends_one`, `cycles_le_three`, `not_solvent_four`, `clear_cost_floor`, `clear_unfunded` and `not_positive`.

10.12 Who prepares the prior

An agent's prior specifies the law from which its operations start. To account for reaching it, prepend a preparation channel to the same protocol. Here that channel lands on one hardware law *whatever it is given*, so the clearing identity of §10.10 also gives its mean heat. Its target must give every hardware state positive probability.

For a uniform target the mean log-ratio heat is zero. The relative entropy of any μ from the uniform law is $\log|S| - H(\mu)$; this cancels the entropy change in the clearing identity. The result holds for any starting hardware law and any thermal scale. Interpreting this expression as reservoir heat uses local detailed balance. It does not price the construction or control of the channel.

The sensor-memory witness starts with a skewed hardware law with full support and randomizes it to the agent's uniform prior. The starting and target laws are proved different. Preparation is stage zero, every later law is the agent's own, and the existing cumulative entropy bound covers the whole run. The choice of the starting hardware law remains an input.

Two limits follow from what the channel can access. It writes the hardware and leaves the unknown parameter's marginal unchanged, so it cannot turn a certain parameter into the uncertainty this agent needs. A channel that reads neither coordinate produces a product law: it cannot create a correlation between the parameter and register. Both limits have concrete counterexamples in Lean.

For contrast, preparing the memory's standard state $(3/4, 1/4)$ from uniform has mean heat $(1/4)\log 3$ at thermal scale one. For a target biased to ε, the corresponding expression is $(1/2 - \varepsilon)\log[(1-\varepsilon)/\varepsilon]$, with $0 < \varepsilon < 1$. Lean proves this family has no finite upper bound. Its exact endpoint has zero reverse probabilities and fails the positive-target premise. This result concerns the specified channels; it does not establish the cost of every implementation of exact preparation. In particular, preparing the sharp charge of §10.11 remains a separate problem.

What it does not give. Local detailed balance at the preparing reservoir is an input, as everywhere here. Funding the preparation from a source, fabricating the gates, preparing a sharp law and any cortical identification are separate.

In Lean. `ProbDist.uniform`, `KL_uniform`, `memoryHeat_uniform`, `FiniteProtocol.withPreparation`, `withPreparation_law_succ`, `preparation_meanHeat`, `preparation_parameter_marginal` and `preparation_product` live in `Phase3_Preparation.lean`. `Examples/Preparation.lean` proves `prior_eq_uniform`, `prepared_law`, `preparation_free`, `prepared_entropy_budget`, `standard_prep_cost`, `biased_prep_cost`, `sharp_prep_unbounded`, `blank_ne_prior`, `known_parameter_not_preparable`, `sharp_target_not_positive` and `correlated_not_preparable`.

10.13 One register's ledger

The active chain's fourth node needs an *upper* bound on a controlled substep's heat. Landauer's principle gives a *lower* bound on an erasure's heat, which is the opposite direction, and two numbers computed in unrelated models do not bound one another at all. What closes the gap is a model in which one register owns both quantities.

A ledger for a register records its own update, the operations it performs on its environment, and one accounting identity: the heat the register dissipates is the total those operations deliver to its reservoir,

$$Q_{\mathrm{reg}} = \sum_i \langle Q \rangle_i .$$

Every operation in the ledger has that register as its controller and obeys local detailed balance at that register's temperature, so the shares and the total are quantities of one physical system rather than of several.

Call an operation *compressive* when it does not increase the joint register–environment entropy. The heat bound of §10, applied to such an operation, gives $\langle Q \rangle_i \geq k_B T[H(p) - H(p')] \geq 0$: a compressive operation cannot draw heat out of the reservoir. If every operation but one compresses, the remaining share is therefore at most Q_{reg}. That is the allocation the edge asks for, derived from the identity and the second law instead of assumed.

The witness is the chain's own one-bit register, at temperature one and with Landauer heat $\log 2$. Its operations drive a lamp bit toward the register's bit: the lamp agrees with the held register bit with probability a, starting from a law of agreement probability w, with interaction energy $\log[a/(1-a)]$ for disagreement. Such an operation delivers mean heat $\log[a/(1-a)](a-w)$. At $a = 2/3$, $w = 4/15$ that is $(2/5)\log 2$; at $a = 4/5$, $w = 1/2$ it is $(3/5)\log 2$. The two make exactly the register's $\log 2$, and the controlled one is the entropy-*increasing* of the pair: its own second law gives it no bound in either direction.

Neither input can be dropped. The same family at $a = 8/9$, $w = 1/9$ preserves the joint entropy exactly, so it is compressive, and still delivers $(7/3)\log 2$ — more than the whole dissipation. And a second ledger for the same register, whose other operation starts at $w = 5/6$ and therefore draws $(1/6)\log 2$ out of the reservoir, satisfies every field of the structure while allocating $(7/6)\log 2$ to its controlled operation.

What it does not give. The index labels the operations; it does not sequence them. Composing them into one evolving joint law, with its preparation, resets and power source, is a separate model. A microscopic bipartite dynamics does not supply the accounting identity either: §10.14 builds one and shows what it leaves open. The ledger says nothing about the coupling-convergence assumption of the next edge, and nothing about cortex.

In Lean. `RegisterLedger` in `Phase3_AgencyThermodynamics.lean` carries the update, operations, local detailed balance and the identity. `opHeat_nonneg_of_compressive` and `opHeat_le_dissipation` derive the allocation, `ledger_bathEntropy` exhibits the total as the register's own bath entropy change, and `Chain.e34Active_of_ledger` discharges the edge. `Examples/RegisterBudget.lean` proves the family's heat, final law and entropy, the witness ledger, and the two regressions.

10.14 What a register's bath fixes, and what it does not

What it is. The ledger charges a register's operations against the heat that register dissipates. Write down the bath that receives the heat, and ask which of the ledger's premises the microscopic dynamics gives you.

Take one register bit and one bath bit. Each of the four maps of the register — identity, negation, and the two erasures — is carried by a permutation of the *pair* that performs it when the bath is prepared in its ground state: an injective map moves the register and leaves the bath alone, and an erasure swaps the register's bit into the bath. Nothing is lost microscopically; the register's information goes somewhere. Let the bath's levels be 0 and $2\log 2$, let the register's be degenerate, take the register uniform and the thermal scale one, and read heat off the paths the gate actually runs as the bath's mean energy gain. It is $\log 2$ for either erasure and 0 for either injective map, and those are the values Landauer's heat equation demands of this bath, whose reachable states the model records.

For any sequence of operations on one law, the bath's own ledger is a first law and nothing more:

$$\sum_{k<N} \langle \Delta E_B \rangle_k = \langle E_B \rangle_N - \langle E_B \rangle_0,$$

with no positivity anywhere and no entropy inequality out of it. A reported heat ledger differs from that gain by the transfer it fails to count, and the difference is available as an identity rather than an estimate.

Three things then go wrong for a budget indexed by the register's map. First, run the erasing gate a second time: the bath is not prepared any more, so the gate hands the bit back, returns the $\log 2$ and restores the law it started from. The two shares are $\log 2$ and $-\log 2$ about a total of zero, so a share is neither nonnegative nor bounded by the total, and an eraser is reusable only if something prepares its bath again. Second, a different reversible gate leaves the register exactly where it is and flips the bath: the same register map as idling, at $2\log 2$ instead of 0. The work accounts for the difference — $\log 2$ for one erasure, $2\log 2$ for that drive, zero for two erasures — and the register's map does not determine any of it. Third, a swap is its own inverse, so forward and reverse transition probabilities agree on every path the gate runs, and the log-ratio heat that local detailed balance would assign is zero at every thermal scale, while the bath gains $\log 2$. And a deterministic gate sends one state to one state, so it is not a strictly positive channel either.

What it does not give. So positivity, local detailed balance and the accounting identity are physical inputs of the resource model, not consequences of microscopic reversibility. This model constructs no ledger and weakens no result that uses one; it says what such a model has to be given. The numbers are those of the declared energy gap and preparation. Preparing the bath, supplying the work, fabricating the gate and identifying this bath with the reservoir the ledger's operations exchange heat with are all still open, and nothing here concerns cortex.

In Lean. `FiniteProtocol.energyTransfer`, `sum_energyTransfer` and `reported_heat_eq` in `Phase 3_ContinuingAgent.lean` are the generic ledger and its correction. `Examples/RegisterBath.lean` carries `lift` with `lift_injective` and `lift_realizes`, `operationHeat_formula`, a file-local `Stati sticalMechanicsBool` instance, `erase_law_two` and `two_stage_heat` for the reused eraser, `same_up date_different_heat` and `driven_work` for the drive, and `gate_channel_zero`, `gate_not_positive` and `erase_stageHeat_not_bath_heat` for the two premises.

10.15 How an action can set a scalar coupling amplitude

What it is. A heat allowance can limit how much joint entropy a process removes. To turn that limit into one on coupling, the model must say how the process changes coupling. Consider a specified spatial profile $b(x)$ and a conversion gain $g \geq 0$. The constitutive law sets the density to

$$f_A(x) = g\, d\, b(x), \qquad d = H(p) - H(p').$$

The process fixes d through its initial and final joint distributions; the model supplies the gain and profile. The drive is signed, so a process that increases joint entropy can give a negative density under this law.

Why the framework needs it. Two deductions then use different premises. Regular spatial meshes with cells shrinking to zero, together with continuity and integrability, give

$$\mathcal{K}_n(f_A) \longrightarrow L_A = g\, d \int_S b\, d\mu.$$

The process's budget gives $d \leq B/(k_B T)$. If the total profile is nonnegative, this bounds the limit by $g(B/(k_B T)) \int_S b\, d\mu$. The budget constrains the amplitude; the mesh assumptions establish

convergence. Refinement does not execute the channel again, so n is a spatial index and not the learning clock.

The noisy two-bit actuator has positive drive $d = (3/4)\log 3 - \log 2$. With unit gain and $b(x) = 1 + |x - 1/2|$ on $[0, 1)$, its first two grid energies are $3d/2$ and $5d/4$. They differ, so this is a changing spatial approximation. The stationary version of the same channel has $d = 0$ and gives zero density. Doubling the gain raises the limit without changing the process or heat. Replacing the profile by one changes the first energy to d. These are exact identities proved in Lean.

What it does not give. This is a supplied law for an ensemble entropy difference. It does not say how a local actuator measures that difference or pays for installing the profile. The gain and profile are additional physical choices, and a scalar density on space is not a pairwise coupling kernel. The model discharges E45Active for its own energy sequence; the microscopic mechanism and the kernel are built separately in §10.16, and cortical identification remains a separate question. Choosing the gain to fit the existing field witness's strength permits composition of the chain but does not predict that strength from heat.

In Lean. `Phase3_ActuatedCoupling.lean` defines the arrangement and proves `actuated_coarse Grains`. In `Chain.lean`, `e45Active_of_actuatedCoupling` supplies the spatial edge and `actuated_limit_le_budget` supplies the limit bound for the same process, temperature, heat and budget. `Examples/ActuatedCoupling.lean` fences process, gain and profile dependence.

10.16 How local actions install a coupling kernel

What it is. A scalar density is not a pairwise coupling. To build one, the model has to say what the hardware is. Suppose it carries finitely many spatial response modes. Each mode i has a fixed profile $\phi_i(x)$ saying where on the substrate it responds, and a price u_i saying what it costs to hold open. A configuration z of the controller and its switches sets the occupancies $c_i(\Xi)$, and those same numbers fix two things at once:

$$K_\Xi(x, y) = \sum_i c_i(\Xi)\, \phi_i(x)\, \phi_i(y), \qquad U(\Xi) = \sum_i u_i\, c_i(\Xi).$$

The first is a kernel on pairs of points; the second is the energy stored in the hardware as configured. A transition $z \to z'$ changes the kernel only through the modes whose occupancy it changes, and its external work is $U(\Xi') - U(\Xi) + Q$.

Why the framework needs it. This is what the scalar model leaves out. The kernel the executed step installs is the average of K_Ξ over the law that step actually leaves, so its change is the mean of the pathwise changes rather than a function of an entropy difference. The mean of the transition works is the installation cost, and the first law splits it into the change in expected stored energy and the channel's heat: closing a mode is paid for, not free.

The spatial limit is then a separate, spatial argument. Divide the substrate into finitely many measurable cells and pick one sample point in each. The reconstruction that uses the sampled weight throughout each *cell pair* has a cell-mass weighted energy exactly equal to its integral against the product measure. As the sample error shrinks, a continuous kernel on a compact substrate gives

$$\mathcal{K}_n \longrightarrow \iint K\, d\mu\, d\mu.$$

The index n counts refinements of space, not executed updates.

One response mode on $[0, 1]$ with profile $(4/3)(1 + x)$ and unit price, switched by the same noisy two-bit channel, installs $(8/9)(1+x)(1+y)$ before and $(4/3)(1+x)(1+y)$ after; the switch closing takes the kernel at the origin from 0 to $16/9$. The mean work is $(1 + \log 3)/4$: expected stored energy rises from $1/2$ to $3/4$, and $(\log 3)/4$ goes to heat. Uniform cells give first energies

4/3 and 25/12 with limit 3. Reading a declared predictive system's own joint law on the same hardware installs $(16/9)(1+x)(1+y)$ with limit 4, so which law the hardware holds changes the answer.

What it does not give. Profiles, prices, occupancy readout, the reservoir convention and the substrate measure are declared hardware. They are what a heat bound does not supply, and the passive edge makes the same point from the other side: two declared predictive systems differing by their whole dissipated work install the same kernel, so efficiency picks out neither the hardware nor the partition. One prepared update is accounted for. Fabricating the hardware, preparing its initial law, powering it indefinitely, and identifying the substrate or the modes with a physical field are separate questions.

In Lean. `Phase3_LocalActuator.lean` defines `LocalActuator` and `KernelArrangement` and proves `kernel_update`, `kernel_update_local`, `expected_kernel_update`, `installation_first_law` and `co arseGrains`. `Phase2_KernelMesh.lean` proves `energy_eq_integral` and `energy_tendsto`. In `Chai n.lean`, `e45Active_of_localActuator` and `e45_of_localActuator` supply the active and passive spatial edges. `Examples/MicroscopicCoupling.lean` holds the thermal-switch witness and its regressions.

10.17 How much installed energy the scalar coherent branch requires

What it is. The integral of the installed kernel is also its mean-field coupling. Fubini's theorem says that averaging over pairs at once and averaging over each point in turn give the same number. With a substrate measure of total mass one, that number is the coupling strength used by the scalar threshold $K_c = 2D$.

Write $m_i = \int \phi_i \, d\mu$ for the spatial mass of mode i and $p(z)$ for the law of the hardware's configuration. The product of its two profiles integrates to m_i^2. The mean coupling and the energy held in the modes are therefore

$$K = \sum_{\Xi} p(\Xi) \sum_i c_i(\Xi) m_i^2, \qquad U = \sum_{\Xi} p(\Xi) \sum_i u_i c_i(\Xi).$$

Declare $\kappa > 0$ with $m_i^2 \le \kappa u_i$ for every mode. If occupancies are nonnegative, multiply that inequality by each occupancy and average:

$$K \le \kappa U, \qquad \kappa U \le 2D \implies K \le 2D.$$

The scalar phase model then has only the uniform self-consistent stationary density and no growing Fourier mode about it. A coherent stationary branch requires $U > 2D/\kappa$.

Why the framework needs it. This connects the hardware's energy account to the scalar coherence criterion. For the existing profile $(4/3)(1+x)$, the spatial mass is 2 and the price is 1, so the sharp conversion factor is 4. At noise $D = 1$, agreement probabilities 1/4, 1/2 and 3/4 hold energies 1/4, 1/2 and 3/4 and give couplings 1, 2 and 3. The theorem excludes the first two, including the threshold itself. It does not decide the third merely from the energy bound; its coupling can be computed separately.

What it does not give. Energy sitting in the modes is not heat dissipated in putting it there. Reversible transfer can move energy with zero entropy production, as the register–bath witness illustrates. The profiles, prices and substrate fix the allowed conversion factors; a heat budget supplies none of them. Exceeding the energy line is necessary, not sufficient: a dearer half-filled mode can sit above it while its coupling stays at the threshold. Negative occupancy would invalidate the bound by allowing one mode to cancel another's price without cancelling its coupling. Finally, the scalar model requires a justified reduction from a heterogeneous kernel. E56 still supplies the physical field identification, and these witnesses identify no cortical substrate.

In Lean. `Phase9_InstalledCoupling.lean` proves `continuumEnergy_eq_mean_field_coupling`, `continuumEnergy_eq_modes`, `continuumEnergy_le_installedEnergy`, `no_coherence_of_installedEnergy` and `stationary_eq_incoherent_of_installedEnergy`. `Chain.installedEnergy_required_by_chain` derives the necessary energy from E56 and E67 for this arrangement. `Examples/InstalledCoupling.lean` contains the numerical witnesses and the rejections.

10.18 The learner pays to install its own coupling

What it is. The learning register can be identified with the actuator's switch. Its configuration is the same pair of bits: the unknown parameter and register. The mode is occupied when those bits agree. With the mode price and profile already specified, the law used to compute reward now also determines stored energy and installed coupling.

The statement. After n updates, this exact example has

$$U_n = \frac{3}{4} - \frac{1}{4}\,2^{-n}, \qquad K_n = 4U_n = 3 - 2^{-n}.$$

The work of the next update includes the increase in storage and its heat:

$$\langle W_n \rangle = \frac{1+\log 3}{8}\,2^{-n}, \qquad W_{<N} = \frac{1+\log 3}{4}(1-2^{-N}).$$

The kernel update is likewise averaged over the learner's own executed paths. Replacing its final law with an unrelated one gives a different kernel.

The uninformative phase sensor uses the same hardware, prior, update rule and work allowance. It stays at $U_n = 1/2$, $K_n = 2$, with zero mean work. At $D = 1$ and $\kappa = 4$ this is exactly the scalar exclusion line. The learner starts there too, and one informative update raises its energy above the line.

What it does not give. The energy bound alone gives a necessary condition. The exact coupling formula can be compared with the stationary scalar threshold separately; neither calculation describes a coherent trajectory. No evolving phase dynamics feeds the installed kernel back into the sensor. Work here is an expectation under the declared reservoir model, not a demonstration of a physical power source.

A separate two-state process closes the phase–register loop. The register installs a signed coupling to a binary phase; a bath-supported phase step changes the phase, a sensor reads it, and that reading updates the register for the next step. The joint law keeps phase–register correlations. With a fixed readout trained on one family, an informative sensor predicts the next phase better on held-out families than blind and replayed controls with the same marginal phase order. A finite work store pays declared preparation, sensing, reset and installation prices; heat to the named bath is recorded for phase flips alone. The example shows a closed loop under supplied toy prices, not a cortical coupling or a heat theorem for the whole protocol.

In Lean. `Examples/LearnedCoupling.lean`, §28: `installedEnergy_learned`, `continuumEnergy_learned`, `kernel_update_on_executed_step`, `work_learned`, `cumulative_work_learned`, `uninformed_no_coherence`, `learned_above_line`.

11 What does not follow: the free-energy reading

An upper bound is a ceiling, not a target. If an error is at most 10 and the ceiling falls to 3, the error may still be 2. Only a ceiling tending to zero forces a nonnegative error to zero. This elementary distinction is the issue here. In predictive modelling, *variational free energy* is a quantity used to assess an approximate probabilistic explanation of observations; it is not

automatically the same quantity as physical free energy, the thermodynamic work budget under specified conditions.

What it is. A tempting inference, and one associated with free-energy formulations (Friston, 2010): if a system minimizes its dissipation, and dissipation bounds the divergence between its internal model and the world, then the divergence must go to zero—the system converges on a true model of its environment.

The statement. The inference is invalid, and the development says exactly where. The assumption available is a bound of the form

$$D_{KL}(P\|Q) \;\le\; \Delta t\,\sigma,$$

carried as the class field `StructuralResonance.kl_bound`. Descending σ does not force $D_{KL} \to 0$ unless the right-hand side itself tends to zero. A driven system settles into a non-equilibrium steady state with $\sigma > 0$, and the bound then permits *any* divergence up to $\Delta t\,\sigma$. The witness `boolDetuned` attains the assumed bound with $D_{KL} = \sigma = \log 2$: a positive limiting divergence is permitted, so the limit cannot be read off the bound.

Separately, Gibbs' inequality proves $D_{KL} \ge 0$. It does not identify relative entropy with physical entropy production. Conflating the two is the other half of the same error.

What it does not give. The framework replaces the invalid inference with the correct, weaker bound and declines the conclusion. It does not claim the free-energy conclusion is false—only that it does not follow from these premises.

Part V. Operators, gradients, and the squared-drift objective

12 The squared-drift functional and its gradient

Consider two sites with systematic phase velocities 1 and 3 in chosen units. Their squared velocities sum to 10. If an interaction redistributes the velocities to 2 and 2, the sum falls to 8, even though both sites still move. Squaring penalizes uneven motion. *Drift* here means the systematic part of phase change; random jitter is treated separately. An *objective* is the quantity an update rule tries to lower. Plasticity means changing the interaction strengths themselves.

The continuum notation below replaces a list of sites by locations x and y. The inner integral adds the influences on one site; the outer integral totals their squared effect across sites. $d\mu$ specifies the spatial weighting, so it plays the role of counting or averaging in a finite model.

What it is. Fix the phases and ask how fast the system is being pushed around. Each site x has a drift—the instantaneous velocity its phase is being driven at:

$$v(x) \;=\; \omega(x) + \int K(x,y)\,\sin(\theta_y - \theta_x)\,d\mu(y).$$

The *squared-drift functional* totals the squares of these:

$$\sigma(K) \;=\; \frac{1}{D}\int \Big[\omega(x) + \int K(x,y)\sin(\theta_y - \theta_x)\,d\mu(y)\Big]^2 d\mu(x).$$

Read as a function of the kernel K with phases held fixed, this is a *quadratic functional of an affine image of* K: the inner integral is a linear operator applied to K, then ω is added (affine), then the result is squared and integrated (quadratic).

Why the framework needs it. The framework wants a quantity that a plasticity rule could descend, and wants to know whether descending it organizes the coupling usefully.

The statement. Functions of this shape have gradients computable by Hilbert-space methods. Let us unpack those words. A gradient points in the direction of steepest increase; moving in its negative direction is gradient descent. An inner product $\langle u, w\rangle$ generalizes the dot product of vectors and measures their alignment. A Hilbert space is a space of vectors or functions with such an inner product and no missing limits. An operator is a rule acting on these objects; linear means it respects addition and scaling, and bounded means it cannot amplify size without a fixed bound. The adjoint moves an operator from one side of an inner product to the other, as transposing a matrix does. Let A be the bounded linear operator sending a kernel to its drift field, and A^* its adjoint—the operator satisfying $\langle Au, w\rangle = \langle u, A^*w\rangle$ for all u, w. Then

$$\nabla\sigma \;=\; \frac{2}{D}\,A^*(AK+\omega)\,,$$

exactly as for a finite-dimensional least-squares problem, and the gradient flow $\dot{K} = -\nabla\sigma$ decreases σ at fixed phases.

A second, more surprising result concerns *symmetric* (reciprocal) kernels. For symmetric K, the integrand $K(x,y)\sin(\theta_y - \theta_x)$ is antisymmetric under swapping x and y, so its double integral vanishes: the *mean* drift is the same for every phase field. Since the mean is pinned, minimizing the sum of squares amounts to minimizing the variance, and a phase-locked field achieves the minimum. Reciprocity is necessary—a one-way two-site coupling strictly beats its own phase-locked state.

What it does not give. Three limitations, all stated in the source. First, σ is *not* in general a thermodynamic entropy-production rate. The Lean identifier is named `entropy_production_rate`, and the supplement notes explicitly that the name does not establish the identification: an actual entropy-production rate requires probability-current information absent from this calculation. Second, descent is proved at *fixed phases*; joint noisy phase-and-coupling dynamics is a different problem, and Section 27 shows the guarantee does not survive it. Third, whether biological couplings follow this flow at all is an additional modelling question.

In Lean. `Phase8_ContinuousField.lean`: `kernelCLM`, `continuumDriftCLM`, the Hilbert-space adjoint, and `structural_resonance_decreases_sigmaContinuum`, with an atomless witness on the unit interval (`Examples` §20). The symmetric-kernel result is `phase_locked_minimizes_entropy_of_symm`, its mean-matching step `total_drift_eq_of_symm`, and its necessity counterexample `Examples` §16.

Part VI. The Kuramoto model: dynamics

13 The model

13.1 Start with one clock hand

An *oscillator* is something that repeats a cycle. Its phase says where it is within that cycle: at the peak, halfway to the next peak, and so on. Picture a clock hand going round a dial. The angle $\theta_i(t)$ labels the position of oscillator i at time t; adding 2π describes the same position after one full turn. The phase model keeps this timing coordinate while leaving out the changing height, or *amplitude*, of an oscillation. It therefore assumes that a meaningful repeating rhythm has already been identified. A neural population or recording channel is not automatically a single oscillator; that identification is part of modelling the data.

An isolated, noiseless hand follows

$$\dot{\theta}_i = \omega_i, \qquad \theta_i(t) = \theta_i(0) + \omega_i t.$$

The first equation says its angular speed is constant; the second adds the angle travelled to its starting angle. The *natural frequency* ω_i is the angular speed it would have without the other oscillators. An ordinary frequency f in cycles per second corresponds to angular frequency $\omega = 2\pi f$ in radians per second. Equal natural frequencies mean equal free-running speeds, not necessarily equal starting angles.

13.2 Let the hands influence one another

For uniform attractive coupling, the deterministic model is

$$\dot{\theta}_i = \omega_i + \underbrace{\frac{K}{N}\sum_{j=1}^{N}\sin(\theta_j - \theta_i)}_{\text{change in speed caused by the population}}.$$

Deterministic means an initial state and the rule determine the whole motion. The phase difference $\theta_j - \theta_i$ is the relative angle of hand j as seen from hand i. For a small positive difference, j is ahead, the sine is positive, and the interaction speeds i up. If j is slightly behind, the interaction slows i down. Equal phases give zero correction. The sine makes this rule repeat every full turn, as a rule for angles must. Exactly opposite hands also give zero sine; the equation permits special balanced configurations as well as alignment.

$K \geq 0$ controls the strength of this attraction. Dividing by N makes the sum an average: increasing the population does not automatically multiply the total coupling experienced by each oscillator. In a network, K_{ij} replaces K/N and specifies the strength from site j to site i. Positive entries favor alignment; negative entries favor separation locally. The threshold derived later concerns uniform attractive coupling.

A two-hand example makes the feedback explicit. Take identical frequencies, $N = 2$, and relative phase $\delta = \theta_2 - \theta_1$. Subtracting their equations gives $\dot{\delta} = -K\sin\delta$. A small positive separation therefore decreases, and a small negative separation increases toward zero. Their *relative* timing converges even while both hands keep turning.

13.3 Add unpredictable timing jitter

What it is. The framework uses the noisy, or *stochastic*, version of this rule:

$$d\theta_i = \Big[\omega_i + \sum_j K_{ij}(t)\sin(\theta_j - \theta_i)\Big]dt + \sqrt{2D}\,dW_i, \qquad r_N = \Big|\frac{1}{N}\sum_i e^{i\theta_i}\Big|. \tag{1}$$

Here W_i are independent Wiener processes (Brownian motions), so $\sqrt{2D}\,dW_i$ is white noise of strength set by the *phase diffusion* $D > 0$. The uniform mean-field convention is $K_{ij} = K/N$.

A Wiener process models accumulated random increments. Over a time step Δt, its increment has mean zero and variance Δt; increments on disjoint time intervals are independent. Multiplication by $\sqrt{2D}$ makes the random phase increment have variance $2D\Delta t$, or standard deviation $\sqrt{2D\Delta t}$. Variance is mean squared departure from the mean; standard deviation is its square root, in the original units. "White noise" describes the idealization of uncorrelated driving increments, not a literal differentiable random velocity.

Read the stochastic differential equation (SDE) as: phase change equals systematic speed times elapsed time, plus a random increment. Larger D means more phase spreading. Independent W_i means each oscillator gets its own jitter; a shared random drive would be a different

model. With angles treated as dimensionless, both K and D have inverse-time units, so K/D is a dimensionless comparison of attraction with spreading.

13.4 Measure how aligned the hands are: the order parameter

A numerical summary of collective organization is called an *order parameter*. "Order" here refers to alignment, not to a command or to the order in which events happen. Draw a unit arrow for each phase, add their horizontal and vertical components, and divide by the number of arrows:

$$C_N = \frac{1}{N}\sum_i \cos\theta_i, \qquad S_N = \frac{1}{N}\sum_i \sin\theta_i, \qquad r_N = \sqrt{C_N^2 + S_N^2}.$$

The complex notation $z_N = C_N + \mathrm{i}S_N = r_N e^{\mathrm{i}\psi_N}$ packages this same average arrow. Its length r_N measures alignment and its direction ψ_N is the collective phase. When the average arrow is zero, its direction is undefined. The subscript N reminds us this is a finite sample.

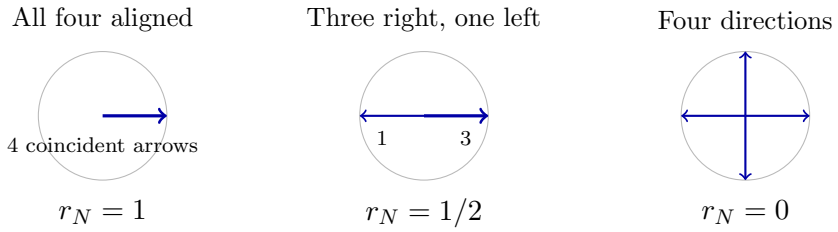


Figure 1: Three exact four-oscillator examples. Average the arrows, then take the length. In the middle, $(3-1)/4 = 1/2$. In the last, horizontal and vertical components each cancel. These are geometric examples, not simulation results.

The *order parameter* r_N is the length of the average of the unit vectors $e^{i\theta_i}$. If all phases coincide the vectors add up and $r_N = 1$; if they are spread uniformly they cancel and $r_N \approx 0$. It is the measure of phase coherence used in the framework's physical test. It contains no information about what the activity represents. Nor does $r_N = 0$ prove an absence of all organization: two equal groups pointing in opposite directions also cancel. Random finite samples usually cancel only approximately.

Phase locking in general means phase differences stay constant; the hands maintain fixed offsets. Perfect *in-phase* synchronization is the special case where all offsets vanish and $r_N = 1$. Several formal results here use "phase-locked" for this stronger aligned condition. With independent noise, sustained statistical coherence need not mean exact locking of each individual pair.

13.5 Why an average arrow simplifies the interaction

For uniform coupling there is an exact finite-population identity:

$$\frac{1}{N}\sum_j \sin(\theta_j - \theta_i) = S_N\cos\theta_i - C_N\sin\theta_i = r_N\sin(\psi_N - \theta_i).$$

It follows by expanding the sine of a difference. Thus each oscillator feels a pull toward the average direction, with strength Kr_N. When the arrows almost cancel, this pull is small even if individual interactions exist. A more aligned population produces a stronger collective pull, which can reinforce alignment. Diffusion opposes this concentration.

This is the *mean field*: the population's aggregate influence. Writing the finite sum this way is exact. Replacing its fluctuating average with a deterministic density for an infinite population is an additional limiting step, explained in Section 18.

Why the framework needs it. Because it is the simplest dynamics in which a genuine phase transition between incoherence and coherence occurs as a coupling parameter is raised— which is precisely the structure the coupling-gated recovery hypothesis needs.

A phase transition is a qualitative change in the possible collective states as a control parameter crosses a threshold. In this phrase "phase" means a population regime; θ_i instead means position within an individual cycle. The sharp stationary transition belongs to the infinite-population description. Finite populations have fluctuating order on either side.

What it does not give. One modelling point is made repeatedly and is easy to miss. *Frequency heterogeneity and phase diffusion are separate parameters.* Classical Kuramoto analyses often use a spread of natural frequencies ω_i as the disordering influence; the framework's stationary theory assumes *identical* frequencies and gets its disorder from D. A frequency-spread width cannot be substituted for D when applying that theory, and the simulations are run with diffusion rather than a fixed-frequency-spread surrogate for exactly this reason.

14 The Lyapunov potential and gradient descent

The next several sections study deterministic motion: set $D = 0$, hold the couplings fixed, and use symmetric interactions $A_{ij} = A_{ji}$. Symmetry here means the influence is reciprocal. It allows these interactions to come from one potential landscape. The noisy stationary theory resumes in Part VII. Random kicks need not move downhill at every instant.

What it is. A *Lyapunov function* is a quantity that never increases along a trajectory; finding one is the standard way to show a dynamical system settles rather than wanders. For the Kuramoto system at zero natural frequency the function is

$$V(\theta) \;=\; -\tfrac{1}{2}\sum_{i,j} A_{ij}\cos(\theta_j - \theta_i),$$

and the dynamics is exactly gradient descent on it: the velocity $\dot\theta_i$ equals $-\partial V/\partial\theta_i$. Differentiating along a trajectory gives the identity

$$\dot V \;=\; -\sum_i \dot\theta_i^2 \;\le\; 0,$$

so V decreases at a rate equal to the total squared velocity, and stops decreasing only when the motion stops.

Why the framework needs it. This converts a question about trajectories into a question about a landscape, and the landscape can be characterized.

The statement. Two directions are proved, for *positive* couplings: `phase_locked_minimizes_potential` (a phase-locked configuration attains the global minimum) and `potential_min_implies_phase_locked` (any minimizer is phase-locked). Together: minimizers of V are exactly the phase-locked states.

What it does not give. Positivity is not a technical convenience. A network of all-positive couplings favors all phase arrows aligning, like magnetic moments aligning in a ferromagnet. Its minimum is unique apart from rotating all phases together. It does not supply many competing states of almost equal energy (a *near-degenerate landscape*) of the kind that a memory argument from geometric frustration would require. The development is careful about where positivity enters:

- Results that assume it: `phase_locked_minimizes_potential`, `potential_min_iff_phase_locked`, the class field `ThermodynamicCover.A_pos` (hence every conclusion at the gluing node), and `kuramoto_tendsto_global_minimum`, which needs the strictly stronger uniform bound $0 < a \le A_{ij}$.

- Results that do *not*: existence and uniqueness of trajectories, the Lyapunov identity itself, monotone decrease of the potential, its convergence, and $\sum_i \dot{\theta}_i^2 \to 0$.

So *that* the motion stops holds for frustrated systems too; *where* it stops does not. Because positivity is a field of one structure rather than a hypothesis scattered through the development, it enters the chain at one node and not before it.

15 The rotating frame

Imagine filming runners moving around a circular track. A camera turning at their shared speed makes their positions appear still if they maintain fixed offsets. Their actual running has not stopped; the observer's coordinate system has changed. A *rotating frame* is that moving coordinate system.

If every oscillator has natural frequency Ω, define its angle relative to the turning camera by

$$\phi_i(t) = \theta_i(t) - \Omega t.$$

Then $d\phi_i = d\theta_i - \Omega\, dt$. The common natural-frequency term cancels, while $\phi_j - \phi_i = \theta_j - \theta_i$, so the coupling term is unchanged. The additive noise is unchanged too. The resulting noisy equation is

$$d\phi_i = \sum_j K_{ij} \sin(\phi_j - \phi_i)\, dt + \sqrt{2D}\, dW_i.$$

For the average arrow, z_N is multiplied by $e^{-\mathrm{i}\Omega t}$. Its direction turns, but its length r_N stays the same. Thus phase differences and coherence can be studied without carrying the common rotation through every equation.

Subtracting a common speed is distinct from *centering* a snapshot on its measured mean phase. The first removes Ωt; the second subtracts $\psi_N(t)$. A noisy finite population's mean direction can wander even in the frequency-rotating frame. Density comparisons may also center each snapshot to compare distribution shapes.

What it is. For deterministic equations on real, unwrapped phases, nonzero natural frequencies make the potential acquire a term $-\sum_i \omega_i \theta_i$, which is linear in the phases and hence *unbounded below*: the potential has no minimum at all, and a minimum-attaining argument is not available. The standard repair is to change to a co-rotating coordinate frame, $\theta_i \mapsto \theta_i - \Omega t$, which subtracts the common rotation. Unwrapped angles count complete turns rather than resetting after each lap. A nonzero linear term can keep falling as that count increases. "Unbounded below" means no finite number bounds its values from below.

The statement. For *identical* natural frequencies $\omega \equiv \Omega$, the reduction is exact: a trajectory (the path of states through time) maps to one of the zero-frequency system. The interaction potential depends only on differences, so its value is unchanged by rotation, and the deterministic descent identity applies in the reduced frame. Under the convergence hypotheses of Section 17, the limit is aligned with $r^2 = 1$. Phase differences, phase-locking and the order parameter magnitude are all proved frame-invariant: changing coordinates preserves these relative observables.

What it does not give. Eliminating every natural frequency this way requires identical frequencies. A spread of frequencies leaves residual detunings $\omega_i - \Omega$ in the reduced system, to which the unboundedness result still applies, and phase-locking then depends on the threshold—which this file does not treat.

In Lean. `Phase4_RotatingFrame.lean`: `kuramoto_potential_unbounded_below`, `is_kuramoto_trajectory_rotate`, `dynamic_potential_descent`, `rotating_frame_chain`. Run end to end on a two-oscillator system in `Examples` §7.

16 Do the trajectories exist? Picard–Lindelöf, globally

An equation is a proposed rule, not yet a guarantee of a well-defined motion. *Existence* asks whether any path follows it from the prescribed start. *Uniqueness* asks whether two different paths could share that start. "Local" means for a short time interval; "global" means for every finite time. A vector field is a rule assigning a velocity vector to each state; it need not be a physical electromagnetic field.

What it is. Before asking where a differential equation goes, one must know it has a solution and only one. The Picard–Lindelöf theorem gives both, locally, when the vector field is Lipschitz—meaning it does not change faster than linearly in the state.

The statement. The Kuramoto vector field is *globally* Lipschitz, with constant $2\sum_{ij}|A_{ij}|$: sin is 1-Lipschitz and the index set is finite, so there is no smallness caveat, and the natural frequencies drop out of the constant entirely. Hence uniqueness on all of $\mathbb{R}$: two solutions agreeing at one instant agree forever, in both time directions.

Existence globally in time takes one extra step, because Mathlib's `IsPicardLindelof` caps a local solution's lifetime by the radius of the ball on which the field's bounds hold divided by its largest velocity bound there, and carries no global-in-time theorem. The field here is bounded by $\sum_i|\omega_i|+\sum_{ij}|A_{ij}|$ on the *whole* state space, so the cap can be defeated by enlarging the ball *after* choosing the window. Solutions on every window $(t_0-T,\, t_0+T)$ are produced, uniqueness makes the family coherent on overlaps, and they glue into a solution on all of $\mathbb{R}$.

In Lean. `kuramotoField_lipschitz`, `kuramotoField_norm_le`, `kuramoto_exists_on_window`, `is_kuramoto_trajectory_exists`, and the combined statement `kuramoto_cauchy_problem`: the trajectory predicate is inhabited exactly once for each initial state.

17 Where the trajectories go: Barbalat and Łojasiewicz

This section describes the longest single chain of analysis in the development, and it is worth following because two of its steps replace textbook tools that Mathlib does not carry.

17.1 The motion stops

If a height only decreases and cannot fall below the valley floor, its value approaches a limit. This is *monotone convergence*. An integral over all future time is an *improper integral*, defined by taking the limit of integrals over longer finite intervals. A finite total squared velocity is a budget on motion, but by itself still allows isolated bursts.

What it is. The potential decreases and is bounded below (at zero frequency), so by monotone convergence it converges: $V(\theta(t)) \to V_\infty$ along *every* trajectory. Integrating the Lyapunov identity gives the quantitative form—the energy dissipated over $[0,T]$ is exactly the potential dropped—so the improper integral

$$\int_0^\infty \sum_i \dot\theta_i^2 \, dt$$

converges, to the total drop.

The statement. Finiteness of that integral is *not* the same as $\sum_i \dot\theta_i^2 \to 0$: an integrable function can spike forever on ever-narrower intervals. The classical bridge is **Barbalat's lemma** (roughly: a uniformly continuous function with a convergent integral tends to zero). Mathlib does not have it, so it is proved here, in the form: *a non-negative Lipschitz function whose integrals $\int_0^T g$ are bounded above tends to zero.* The proof idea is geometric—a spike of height ε drags a whole window of width $\varepsilon/2K$ up with it, and only finitely many such windows fit under a finite integral.

Applying it needs the dissipation rate to be Lipschitz in time, which is checked without differentiating it: velocities are bounded, the field is Lipschitz in the configuration, and the configuration is Lipschitz in time because its derivative *is* the bounded field. Conclusion (`velocity_sq_tendsto_zero`): along every trajectory of a zero-frequency system, $\sum_i \dot\theta_i^2 \to 0$.

What it does not give. This locates no limit. It says the motion stops, not where.

17.2 Where it stops

An *equilibrium* of a deterministic equation is a state with zero velocity. An *invariant region* is one a trajectory cannot leave after entering. Compactness is a mathematical form of having no escape to infinity or missing boundary points; in ordinary finite-dimensional coordinates, closed bounded regions are compact. These concepts help locate eventual motion, but they do not mean every equilibrium is a global minimum.

What it is. Locating the limit looks like a job for the *LaSalle invariance principle*, which Mathlib does not carry, and which would want a compact invariant set that the state space $V \to \mathbb{R}$ does not supply. Neither turns out to be needed. LaSalle is what one needs when a decreasing functional is *all* one has; here the flow is a gradient flow whose minimizers are already characterized (Section 14), and the right tool is a *Łojasiewicz* (or Polyak–Łojasiewicz) estimate: an inequality bounding how far the functional is above its minimum by how fast it is currently decreasing.

The statement. The estimate proved is: on the region where every pairwise phase difference lies within a quarter turn,

$$2a^2 W \;\le\; \Big(\sum_{ij} A_{ij}\Big) \sum_i \dot\theta_i^2,$$

where W is the excess of the potential over its minimum and $a > 0$ is a lower bound on the couplings. The proof uses the two sites carrying the largest and smallest phase, without assumptions about the curvature of the whole landscape.

Combined with $\dot W = -\sum_i \dot\theta_i^2$ this is the differential inequality $\dot W \le -cW$ with $c = 2a^2/\sum_{ij} A_{ij}$, which integrates elementarily ("$W(t)e^{ct}$ has non-positive derivative") to exponential decay $W(t) \le W(0)e^{-ct}$. The rate transfers to the velocity via $|\sin d| \le \sqrt{2(1-\cos d)}$, giving $|\dot\theta_i(t)| \le Be^{-ct/2}$, which is integrable on $(0,\infty)$; since $\theta_i(t) = \theta_i(0) + \int_0^t \dot\theta_i$, the trajectory itself converges, and $W \to 0$ forces the limit to be phase-locked—a *global* minimizer with $r^2 = 1$.

17.3 The compactness objection, met without an invariant region

Here the direct route is to control the initial energy. If it is too low for any pair to reach a quarter-turn separation, energy descent keeps the pair within that range. Continuity prevents it from jumping across the forbidden angle. This is what the argument calls *branch-pinning*: keeping a phase difference in one chosen interval of unwrapped angles.

The quarter-turn region has to be shown forward-invariant, or the estimate applies only at the initial instant. It is, and the argument is neat: because all terms of the excess are non-negative, a single pair satisfies $\frac{1}{2}a(1-\cos(\theta_j-\theta_i)) \le W$, so an initial excess below $a/2$ —which monotone decrease preserves—keeps every pairwise cosine strictly positive for all time. A phase difference escaping $[-\pi/2, \pi/2]$ would have to cross $\pi/2$ in absolute value by the intermediate value theorem, and there the cosine vanishes. This is branch-pinning by continuity on a connected half-line, and it replaces a quotient by the phase-shift symmetry.

What it does not give. The two hypotheses are on the *initial data* and neither can be removed. A splay state spreads phases evenly around the circle; a twisted state makes phase advance systematically around a spatial loop. Such balanced configurations can be equilibria too—on a ring with a translation-invariant symmetric kernel a twisted state is stationary exactly (`windi`

ng_is_kuramoto_trajectory), its drift cancelling by pairing each separation with its negative—so "every trajectory reaches the phase-locked state" is simply false. What is proved is the arc-condition version.

The hypothesis is not merely convenient, then; it is forced. A twisted state has global order zero and a global minimiser of the potential has order one, so a twisted state is not a minimiser at all (char_not_potential_min): the conclusion of the convergence theorem is false of it, and no version of the theorem can reach it. What the twisted state does instead is answered by linearising there. The derivative of the field at a twist is the same translation-invariant coupling with each separation reweighted by the cosine of the phase the twist advances across it (hasDerivAt_charJacobian); the reweighted kernel is still even, so the same modes diagonalise it and their rates are its eigenvalues. Where the twist advances by no more than a quarter turn across every separation the coupling reaches, all rates are non-positive; where one rate is positive, that mode grows without bound and the potential falls along it. This is a statement about rates at the linearisation and not about trajectories: it says which twists cannot persist, not which one a trajectory reaches. It is witnessed on three sites (Examples §17), where no closed-form solution exists—the trajectory is produced by the existence theorem and its limit by the convergence theorem, the whole numerical content of checking the hypotheses being $\cos\frac{1}{2} > \frac{3}{4}$.

18 The open gap: propagation of chaos

A finite simulation keeps a list of N angles. A continuum description keeps instead a density $\rho(\theta, t)$: integrating it over an arc gives the fraction or probability of phases in that arc. A histogram with ever more samples suggests such a density, but interactions can correlate the samples. To justify a closed equation for one representative oscillator, one must control those correlations as the population grows.

"Chaos" here means statistical independence in the limiting construction, not unpredictability from sensitive deterministic dynamics. Factorization means the joint probability becomes a product of individual probabilities. The one-particle marginal is the distribution of one oscillator after the others have been averaged out. A closed equation determines its evolution without needing a separate equation for every pair, triple and larger group.

What it is. Everything in Part VII is about a *mean-field* description—a single density over phases, as though each oscillator felt an averaged influence from an independent population. Justifying that from the finite-N system (1) is the problem of *propagation of chaos* (Sznitman, 1991): showing that if oscillators start independent they stay asymptotically independent as $N \to \infty$, so the joint law factorizes and the one-particle marginal obeys a closed equation.

Why the framework needs it. Without it, the stationary theory of Part VII and the finite system of Part VI are two different objects with a suggestive resemblance.

The statement. No theorem in the development closes this gap, and the manuscript says so plainly. Exchange symmetry alone does not justify factorization, particularly for an ensemble with an unpinned collective phase.

What it does not give. Two finite-N controls probe the gap without closing it, and both are reported as negative or partial:

- **Subcritical.** As N grows, mean order falls toward the finite-size floor and absolute circular pair correlation decreases—consistent with independence. But the matched-sample sliced-Wasserstein distance between the joint law and the product of its marginals barely moves, exposing that statistic's finite-sample floor rather than measuring convergence.

- **Supercritical.** Mean order stabilizes, but lab-frame pair correlation stays substantial rather than following the $1/N$ reference. The reason is structural, not a failure of the simulation: above threshold the ensemble has an *unpinned collective orientation*, a shared latent variable

that every realization draws once. Unconditional independence and independence *conditional on that orientation* are different claims, and only the first is refuted. After centering on each realization's collective orientation, the $N = 1000$ density matches the von Mises target closely.

This is a negative result for the *unconditional* protocol and settles nothing about the conditional mean-field limit.

For intuition, imagine photographs of the same cluster of arrows, each photo rotated through a randomly chosen angle. Across photos, knowing one arrow's direction tells you about the others, although the cluster shape is unchanged. Conditioning means specifying that shared orientation before comparing the arrows. The sliced-Wasserstein diagnostic above measures differences between distributions by projecting them onto lines and comparing the projected samples; its sampling floor is the discrepancy finite samples produce even when the underlying distributions agree.

Part VII. The synchronization transition

19 The von Mises stationary density

19.1 What becomes stationary?

Imagine watching many clock hands whose speeds are continually perturbed. Individual hands keep moving, but the fraction in each part of the dial can remain unchanged. A *stationary density* describes that unchanging probability pattern. Stationary therefore does not mean every noisy hand stops. The statement is made in the rotating frame, at fixed K and D, and for a chosen collective orientation ψ.

The density integrates to one around the circle. Its collective arrow is

$$z = \int_{-\pi}^{\pi} e^{\mathrm{i}\theta} \rho(\theta)\, d\theta = r e^{\mathrm{i}\psi}.$$

This is the continuous counterpart of averaging N arrows: r has the same meaning as r_N, but refers to a distribution rather than one finite list.

19.2 What is an ansatz?

An *ansatz* is a proposed form for a solution, with quantities still to determine. For example, proposing a straight line $y = mx + b$ leaves its slope m and height b to be found. One then checks whether that form satisfies the problem. Calling a form an ansatz does not prove it, and does not mean it is an arbitrary guess: an equation or symmetry can motivate the choice.

The proposed circular density is proportional to $e^{a\cos(\theta-\psi)}$. In the stated stationary model this form can be derived, as the next subsection explains. The cosine is largest at ψ, so positive a weights nearby angles more heavily. At $a = 0$ every angle gets equal weight; increasing a concentrates the distribution around ψ. The quantity a is called its *concentration*. It is different from r: a specifies the density's shape and is unbounded above, while r is the length of its average arrow and lies between zero and one.

What it is. The *von Mises* distribution is the circular analogue of the Gaussian:

$$\rho(\theta) = \frac{e^{a\cos(\theta-\psi)}}{2\pi I_0(a)}.$$

It is peaked at ψ with *concentration* $a \geq 0$; at $a = 0$ it is uniform, and as a grows it tightens. The normalizing constant involves I_0, a *modified Bessel function of the first kind*—the functions

that arise in circular integrals. No prior knowledge of Bessel functions is needed here: the two used in self-consistency are defined by

$$I_0(a) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{a\cos u}\,du, \qquad I_1(a) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \cos u\, e^{a\cos u}\,du.$$

$2\pi I_0(a)$ is the total unnormalized weight, so dividing by it makes ρ integrate to one. The ratio I_1/I_0 is the average horizontal component of a phase arrow after centering on ψ. The vertical components cancel by symmetry. A Gaussian is the familiar bell-shaped distribution on a line; von Mises plays a similar role on a circle.

19.3 Why this particular density fits the model

The mean-field equation for a density is the *Fokker–Planck equation*. It tracks probability moving around the dial, rather than each oscillator's random path. Writing $b(\theta) = Kr\sin(\psi-\theta)$ for systematic drift,

$$\frac{\partial\rho}{\partial t} = -\frac{\partial J}{\partial\theta}, \qquad J = b\rho - D\frac{\partial\rho}{\partial\theta}.$$

J is the *probability current*: net probability crossing an angle per unit time. Its first term transports probability with the drift; the second spreads it from denser to less dense regions. If more current enters a small arc than leaves, its probability increases. This is the content of the first equation, a conservation law.

For the zero-current stationary solution, choose the mean direction as zero and set $J = 0$. With $b = -Kr\sin\theta$, this gives

$$D\rho' = -Kr\sin\theta\,\rho, \qquad \frac{\rho'}{\rho} = -\frac{Kr}{D}\sin\theta.$$

The left side is the derivative of $\log\rho$. Since the derivative of $\cos\theta$ is $-\sin\theta$, integration yields $\log\rho = (Kr/D)\cos\theta + \text{constant}$. Exponentiating and normalizing gives the von Mises form with $a = Kr/D$. This elementary calculation explains the ansatz under the mean-field and zero-current conditions.

The second condition is a consequence here rather than a further assumption. Stationarity makes J constant, since $\partial J/\partial\theta = 0$. On a circle, with a drift that is the slope of a periodic potential, that constant has to be zero: multiplying ρ by $e^{V/D}$ turns the current into the derivative of a single periodic function, and a periodic function's derivative integrates to zero over a full turn, while the factor it is multiplied by is strictly positive. The formalization proves this (`stationary_current_zero`) and then reads off the density (`stationary_iff_vonMises`). Periodicity is what does the work: a stationary density on a line, or under a drift that is not the slope of a potential, can carry nonzero current, so the word "stationary" alone does not justify setting $J = 0$.

19.4 Order does not by itself cost dissipation

The ordered stationary state has a nonuniform density and zero probability current. Its current-based entropy production, $\sigma_J = \int J^2/(D\rho)$, is therefore zero even when its order is positive. Keeping that density fixed in time also keeps its time-averaged cost zero. Positive order alone cannot force continuous heat production in this model.

There is a quantitative statement when current is specified. For a continuous positive normalized density and continuous current,

$$\left(\int_{-\pi}^{\pi} J\right)^2 \le D\sigma_J.$$

To see it, put $m = \int J$ and expand the nonnegative integral $\int (J - m\rho)^2/\rho = D\sigma_J - m^2$. A uniform density driven at constant angular velocity ω attains equality: $J = \omega/(2\pi)$ and $\sigma_J = \omega^2/D$. The formalization checks both this stationary witness and an ordered zero-cost witness. Applying the bound to cortical power requires the physical drive, its current and a thermal scale; the squared-drift objective used for coupling plasticity is a separate quantity. For the cosine moment of a sufficiently regular evolving density, the continuity equation also gives $\dot{m}^2 \le D\sigma_J$. This limits density change. For the rotation-invariant magnitude r of the density's average arrow, a stronger statement keeps track of how much room the density has to change:

$$\dot{r}^2 \le D\sigma_J(1 - r^2).$$

Over a finite interval on which $0 < r < 1$, it gives the arcsine endpoint floor $\int_0^\tau \sigma_J\, dt \ge [\arcsin r(\tau) - \arcsin r(0)]^2/(D\tau)$. The proof needs a positive normalized periodic density, a classical continuity equation and valid differentiation of its moments. The bound prices a change in order, not its maintenance: the ordered stationary counterexample still has zero current cost. Saved finite-volume trajectories check the floor at an exploratory order threshold, with no conversion of current cost to cortical heat. It places no general limit on an external coupling ramp: the uniform density solves the time-dependent equation with zero current cost even for $K(t) = e^{at}$ and arbitrary slope $K'(0) = a$.

19.5 Make the population reproduce its own average

Why the framework needs it. It is the stationary density of the noisy mean-field Kuramoto dynamics: the shape the phase distribution settles into when a coherent mean field competes with diffusion. Every distributional prediction the framework makes flows from it.

The statement. For identical frequencies in the rotating frame, the positive normalized stationary density has the form above with

$$a = \frac{Kr}{D}, \qquad r = R(a) := \frac{I_1(a)}{I_0(a)}. \tag{2}$$

The second equation is a *self-consistency* condition and deserves unpacking. The concentration a is proportional to the coherence r, because a more coherent population exerts a stronger mean field. But the coherence is itself determined by the density: the average of $e^{i\theta}$ under a von Mises law at concentration a is exactly $R(a)$. So r must reproduce itself, and the fixed points of $r \mapsto R(Kr/D)$ are the possible stationary states. Sakaguchi (1988) is the source for the threshold this yields.

The calculation can be read as a loop: propose r, use it to obtain the pull Kr, use attraction versus diffusion to obtain a density, then average that density's arrows. A *self-consistent* proposal returns the same r you started with. A fixed point of this calculation is a solution of an algebraic equation. Repeating this calculation numerically is not the same as evolving the physical oscillators in time.

Following these stationary solutions across choices of K traces a *stationary branch*: a family of possible settled densities, each at its own fixed coupling. Section 21 explains how to read that graph.

What it does not give. The formal stationary theorems take this density as their starting point. The calculation above is an explanatory derivation, not a Lean proof connecting the SDE to its long-time distribution. The repository script `simulations/fokker_planck_von_mises.py` checks the stationary algebra symbolically and numerically. It proves neither the passage from the finite system to a density nor the density's dynamical selection. Stability analysis asks how small perturbations evolve; spectral methods study the growth or decay of their different modes. These are two distinct missing links: Section 18 asks whether the finite system

approaches a mean-field description; the long-time question asks which density that description selects.

In Lean. `Phase8_SelfConsistency.lean`. Because Mathlib has no modified Bessel functions, `besselRatio` is written as a ratio of interval integrals. `vonMisesDensity` is proved a probability density on $[-\pi, \pi]$; `circularOrderParameter_vonMises` proves its order parameter equals $R(a)$ (real, because $\mathbb{E}_a[\sin\theta] = 0$ by symmetry); and `fixedPoint_iff_selfReproducing` states the equivalence—r solves $r = R(Kr/D)$ exactly when the density it induces has order parameter r.

20 Why the threshold is $K_c = 2D$

Near a uniform density, a small concentration gives approximately $R(a) = a/2$. In the self-consistency loop, a small proposed r therefore returns roughly $(K/2D)r$. If $K < 2D$, that returned value is smaller; if $K > 2D$, it is larger. The critical value $K_c = 2D$ is where the initial slope of this response equals one. The following inequalities turn that local intuition into a statement about all the stationary solutions. This slope argument alone is not a proof of how finite noisy trajectories evolve.

What it is. The claim is that a positive self-consistent stationary r is impossible at or below $K = 2D$ and, above it, takes exactly one value. Zero also remains a solution above threshold. The number 2 is not fitted; it falls out of a single inequality about the von Mises family.

The statement. Integration by parts is the product rule for derivatives used in reverse inside an integral. It exchanges a derivative on one factor for a derivative on the other, plus endpoint terms. Here the key identity is

$$\int_{-\pi}^{\pi} \cos\theta \, e^{a\cos\theta} \, d\theta \;=\; a \int_{-\pi}^{\pi} \sin^2\theta \, e^{a\cos\theta} \, d\theta,$$

exact because $\sin(\pm\pi) = 0$ kills the boundary term. Dividing by the normalizer, this factorizes the Bessel ratio:

$$R(a) \;=\; a\,\mathcal{S}(a), \qquad \mathcal{S}(a) := \mathbb{E}_a[\sin^2\theta].$$

The subscript a means average with the von Mises probabilities at that concentration. Thus $\mathcal{S}(a)$ is the average squared vertical component of a centered phase arrow. A moment is an average of a power of a variable; the cosine moments below are averages or unnormalized integrals of powers of cosine. Now everything turns on $\mathcal{S}$. Since $\sin^2 \le 1$ trivially, $R(a) \le a$, which would put a threshold at $K = D$. The sharper fact is

$$\mathcal{S}(a) \;\le\; \tfrac{1}{2}, \qquad \text{strictly, for } a > 0,$$

which gives $R(a) < a/2$ and hence a threshold at $K = 2D$. **That factor of** $1/2$ **is exactly what separates a threshold at** D **from one at** $2D$, and the development keeps the crude version around to make the point visible.

Proving $\mathcal{S}(a) \le 1/2$ amounts to showing the second cosine moment $I_2(a) \ge 0$, done by *folding* the interval $[-\pi, \pi]$ onto $[0, \pi/4]$ through the two reflections $\theta \mapsto \pi - \theta$ and $\theta \mapsto \pi/2 - \theta$, after which the integrand becomes a product of two manifestly non-negative factors. Strictness for $a > 0$ comes from positivity of the folded integrand on the interior of that interval.

With $R(a) = aE(a)$ in hand, the two halves of the threshold follow:

Subcritical ($K \le 2D$).

The only $r \ge 0$ solving $r = R(Kr/D)$ is $r = 0$. Because $R(a) < a/2$, the map $r \mapsto R(Kr/D)$ lies strictly below $r \mapsto (K/2D)r \le r$ for $r > 0$, so it cannot cross. The threshold case $K = 2D$ falls on the *incoherent* side, so the two regimes exhaust $K \ge 0$ with no gap.

Supercritical ($K > 2D$).

There is an r with $0 < r \leq 1$ solving it, and no series expansion is needed. Writing the condition through $\mathcal{S}$: since $D/K < 1/2 = \mathcal{S}(0)$ and $\mathcal{S}$ is continuous, the map overshoots for small $r > 0$; $R(a) \leq 1$ (being a mean of $\cos\theta$) makes it undershoot at $r = 1$; and the intermediate value theorem supplies the crossing.

Three further results complete the picture:

- **Uniqueness.** At most one positive solution exists—see Section 21.
- **Continuity at threshold.** For every $\varepsilon > 0$ there is $\delta > 0$ such that on $(2D, 2D + \delta)$ every coherent solution has $r < \varepsilon$. The branch emerges *continuously* from zero: no discontinuous jump. (The proof needs no monotonicity of $\mathcal{S}$—on a compact interval bounded away from the origin the extreme value theorem gives a maximum strictly below $1/2$, and beyond that the crude bound $\mathcal{S}(a) \leq 1/a$ suffices.)
- **Strict growth.** r increases strictly with K above threshold.

Together: the non-negative solution set is exactly $\{0\}$ at or below threshold and $\{0, r\}$ above it. A *bifurcation* is a qualitative change in the solution structure as a parameter varies. The onset is often drawn as a pitchfork using a signed amplitude. Here r is a nonnegative length, and different mean directions ψ give a whole family of rotated coherent densities with that same length. Uniqueness concerns the positive magnitude.

What it does not give. Everything here is a statement about a stationary *density*. Of the coherent branch the development proves existence, uniqueness, continuous emergence and strict growth—but **not that it is dynamically selected**. No theorem in these files mentions a trajectory, so none could. The order parameter reached is that of a density, not of a trajectory, and no theorem relates the two—that relation is the mean-field limit of Section 18.

There is also a modelling caution about the spatial version. Replacing the finite sum by an integral against a kernel $K(x, y)$ requires the substrate measure to be *normalized to total mass one*, and this is load-bearing rather than cosmetic. The predicate compares the kernel's double average against $2D$; on a substrate of total mass m that average carries a factor m^2, so without normalization the comparison tracks the substrate's *size* rather than its coupling and can be satisfied by any kernel at all. `Examples` §9 exhibits a system a factor of two below threshold whose unnormalized double integral nonetheless clears $2D$. Reduction of a heterogeneous spatial kernel to a scalar K is a further modelling step, and the threshold does not automatically transfer to every spatial or signed network.

21 The fold-and-cross proof of uniqueness

21.1 First, what is a stationary branch?

Fix D. For each chosen K, solve the stationary equation and mark its solutions on a graph with K horizontally and r vertically. Following one continuous family of solutions as K changes traces a *branch*. The horizontal line $r = 0$ is the incoherent branch. Above K_c, a positive coherent branch emerges and rises. "Stationary" describes each solution at its own fixed coupling; "branch" collects these separate solutions. The horizontal axis is coupling, not elapsed time.

Object	Question it answers
Stationary solution	What density can remain unchanged at this fixed K?
Stationary branch	How do those solutions vary across choices of K?
Stability	Does a small disturbance shrink or grow under the dynamics?
Trajectory	What does this initial population actually do through time?

Finding a branch does not prove stability or say which trajectory approaches it. A ball exactly balanced on a hilltop illustrates the distinction: the position is an equilibrium, but a small displacement takes it away.

The first of those two questions is answered for this equation, one linear order deep. Disturb the uniform density a little: the disturbance splits into harmonics, the coupling touches only the first one, and that harmonic changes at rate $K/2 - D$ while the n-th shrinks at rate $-Dn^2$. So a growing disturbance exists exactly above $K = 2D$ — the hilltop, in this picture, appears at the same coupling at which the coherent branch appears. Disturb the coherent density instead, and the corresponding quantity is strictly negative in every direction except one. The exception is rotation: sliding the whole density around the dial gives another stationary density, so nothing decays along that direction and nothing can. Stability here therefore means stability of the shape, not of its position. None of this says which trajectory a population follows, which is the second question and remains open.

21.2 Why there cannot be two positive solutions at the same coupling

The positive-branch equation can be written $\mathcal{S}(a) = D/K$. If $\mathcal{S}$ strictly decreases as a rises, a horizontal line of height D/K can meet it at most once. That is the geometric meaning of the uniqueness proof.

This is the one proof worth describing in outline, because its structure is unusual and the supplement devotes a section to it.

What it is. Uniqueness of the coherent branch reduces to the *injectivity* of $\mathcal{S}$ on $(0, \infty)$, and what is proved is the stronger statement that $\mathcal{S}$ is *strictly decreasing* on $[0, \infty)$.

The statement. One route is to differentiate. An exponential family weights outcomes by an exponential of a parameter times a statistic; changing the parameter is called *tilting*. Covariance measures whether two quantities tend to be above their means together: $\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$. For this family, differentiation gives

$$\mathcal{S}'(a) \;=\; -\,\mathrm{Cov}_a\big(\cos^2\theta,\ \cos\theta\big)\,,$$

and one would show that covariance positive. That route was not taken, and it was the harder one: it costs differentiation under the integral sign, and the covariance is not sign-definite pointwise—symmetrizing with an independent copy gives $\frac{1}{2}\mathbb{E}[(X - X')^2(X + X')]$ with $X = \cos\theta$, an integrand changing sign with $X + X'$.

In the proof below the integrals, both over $[-\pi, \pi]$, are

$$Z(a) = \int_{-\pi}^{\pi} e^{a\cos\theta}\, d\theta, \qquad C_2(a) = \int_{-\pi}^{\pi} \cos(2\theta) e^{a\cos\theta}\, d\theta.$$

$\cosh x = (e^x + e^{-x})/2$ is the hyperbolic cosine. Folding means pairing reflected angles so their contributions can be added on a smaller interval. The target is the two-point inequality $C_2(a)Z(b) < C_2(b)Z(a)$ for $a < b$, and two moves settle it.

Fold first. The same two reflections used for $I_2 \geq 0$ carry both integrals onto $[0, \pi/2]$, replacing $e^{a\cos\theta}$ by $W(a, \theta) = 2\cosh(a\cos\theta)$ and confining $\cos\theta$ to $[0, 1]$. There, $\cos 2x - \cos 2y$ and $W(b, x)W(a, y) - W(a, x)W(b, y)$ have the *same sign*, both being governed by $\cos x$ against $\cos y$. This is the same $X \mapsto -X$ symmetrization the covariance form needed, performed once at the start rather than inside a double integral. Its entire analytic content is one identity: $\cosh(ap)\cosh(bq) \leq \cosh(bp)\cosh(aq)$ for $0 \leq a \leq b$, $0 \leq q \leq p$, which follows from $2\cosh X \cosh Y = \cosh(X+Y) + \cosh(X-Y)$ and cosh increasing in $|\cdot|$.

Then cross. Rather than integrating that inequality over a square—which would need a product measure and Fubini's theorem, which permits exchanging the order of two integrals under suitable conditions—let α be the mean of $\cos 2x$ under $W(a, \cdot)$, and let $x_0 \in [0, \pi/2]$ be

a point where $\cos 2x_0 = \alpha$, supplied by the intermediate value theorem. Evaluating at $y = x_0$ and integrating in x alone gives the conclusion, because the term carrying $W(b, x_0)$ vanishes by the choice of α. Strictness comes from positivity of the integrand on a subinterval missing x_0.

Nothing in the proof differentiates anything. No product measure and no Fubini appear.

The same crossing argument run with $\cos\theta$ in place of $\sin^2\theta$ needs no fold at all—$R(a) = \mathbb{E}_a[\cos\theta]$ is the mean of the very variable the family is tilted by—and yields `besselRatio_strictMono`: the Bessel ratio R is strictly increasing on all of $\mathbb{R}$. Mathlib carries this in no form.

The statement. The consequences are the uniqueness results (`coherent_fixed_point_unique`, `supercritical_fixed_point_existsUnique`, `supercritical_solution_set`) and strict growth of the branch (`coherent_branch_strictMono`), which uses both halves: $\mathcal{S}$ decreasing puts the concentrations of two couplings in the order $a_1 < a_2$ through $\mathcal{S}(a_i) = D/K_i$, and R increasing carries that order to $r_1 < r_2$.

What it does not give. Because nothing is differentiated, $\mathcal{S}$ and R are obtained as strictly monotone rather than as smooth. The rate at which the branch leaves zero therefore needs a different file—the next section.

Part VIII. The critical exponent and the square-root onset

22 The exponent $\beta = 1/2$

An exponent describes how a quantity changes when its input is rescaled. For $r \propto \sqrt{K - K_c}$, making a small coupling excess four times larger makes r about twice as large. This is the meaning of exponent 1/2. It describes the branch close to threshold, not the entire curve up to full alignment. A critical exponent is an exponent at such a transition.

What it is. Near a continuous phase transition, the order parameter typically behaves like a power of the distance to threshold: $r \sim (K - K_c)^\beta$. The exponent β is a *universal* feature of the transition—it does not depend on microscopic details, only on the class the transition belongs to. For mean-field theories $\beta = 1/2$, giving a square-root shape.

Why the framework needs it. This supplies the framework's proposed near-onset shape. A square-root graph has *infinite initial slope*: its slope becomes arbitrarily large as the positive coupling excess approaches zero. Its height still starts continuously at zero. This is a statement about a stationary graph, not an instantaneous physical jump. Converting it to a recovery-time prediction requires the additional tracking assumptions in the next section.

The statement. The calculation uses a Taylor expansion: near a point, approximate a smooth function by a polynomial built from its derivatives there, keeping track of the remaining error. Differentiating under an integral means exchanging the order of differentiation and averaging. Dominated convergence justifies such limit exchanges when one common integrable bound controls the quantities. These are the tools behind the following steps.

1. **Differentiate under the integral.** Define the cosine moments $\mathrm{vmMoment}\, n\, a = \int_{-\pi}^{\pi} \cos^n\theta\, e^{a\cos\theta}$. Differentiating in a raises the cosine power by one. The interchange of derivative and integral is justified by dominated convergence, dominated by the constant $e^{|a|+1}$ on a unit ball around the parameter—this is where compactness of $[-\pi, \pi]$ and boundedness of cos do the work. Iterating gives two continuous derivatives.

2. **Rewrite $\mathcal{S}$ through the moments.** $\mathcal{S}(a) = 1 - \mathrm{vmMoment}\, 2\, a / \mathrm{vmMoment}\, 0\, a$, a quotient with positive denominator, so $\mathcal{S}$ inherits two derivatives.

3. **Evaluate at zero.** The moments at $a = 0$ are elementary integrals of powers of cos: $2\pi,\ 0,\ \pi,\ 0,\ 3\pi/4$. They give

$$\mathcal{S}(0) = \tfrac{1}{2}, \qquad \mathcal{S}'(0) = 0, \qquad \mathcal{S}''(0) = -\tfrac{1}{8},$$

the first derivative vanishing by the symmetry that kills the odd moments.

4. **Taylor.** Hence

$$\mathcal{S}(a) \;=\; \frac{1}{2} - \frac{a^2}{16} + o(a^2).$$

The remainder is $o(a^2)$, not $O(a^4)$; the stronger form would need a fourth derivative and nothing requires it.

5. **Substitute into the branch equation.** On the coherent branch $\mathcal{S}(Kr/D) = D/K$. Writing $a = Kr/D$ and rearranging gives

$$K - 2D \;=\; Ka^2\big(\tfrac{1}{8} - 2q(a)\big)$$

with q the normalized remainder. Letting $K \to 2D$ along *any* family of positive coherent solutions with $r \to 0$ sends $q \to 0$ and yields

$$\frac{r^2}{K - K_c} \;\longrightarrow\; \frac{1}{D}, \qquad \text{i.e.} \qquad r \sim \sqrt{\frac{K - K_c}{D}}. \tag{3}$$

The conclusion applies to any family of positive solutions approaching the threshold in the stated way. Lean expresses that generality using a *filter*, a device specifying what counts as "sufficiently close" in a limit. This mathematical filter is unrelated to filtering a recording. "Eventually" means after entering that sufficiently close regime. The file also constructs an actual family, `coherentBranch`, and proves it approaches zero and satisfies the limit. That witness prevents the general statement from being useful only for a family that does not exist.

What it does not give. Asymptotics of the solutions of a *stationary* equation. It establishes no stability, no dynamical selection and no time course. Reading it as a square-root foot on a recovery curve requires two further assumptions, both stated in the manuscript as assumptions: adiabatic separation, and a threshold crossing at non-zero speed.

In Lean. `Phase8_CriticalExponent.lean`: `hasDerivAt_vmMoment`, `contDiff_two_vmMoment`, `vonMisesSRatio_second_order`, `coherent_solution_critical_exponent`.

23 From exponent to prediction

Two graphs are involved. The stationary graph puts coupling K on the horizontal axis. A recovery trace puts time t there. To connect them, one must know how coupling changes with time and whether the phase distribution can adjust quickly enough to stay near each stationary solution.

Relaxation time is the timescale on which a disturbed distribution approaches its settled form. *Adiabatic tracking* means the controls change little during that adjustment, so the actual distribution remains close to the corresponding stationary branch. Here "adiabatic" means slow tracking; in thermodynamics the same word can instead mean no heat exchange.

The statement. If the coupling $K(t)$ rises through K_c at time t_0 with non-zero speed $\dot{K}(t_0) > 0$, substituting $K(t) - K_c \approx \dot{K}(t_0)(t - t_0)$ into (3) gives the stationary benchmark at fixed D:

$$r_{\mathrm{ss}}(K(t)) \;\sim\; \sqrt{\frac{\dot{K}(t_0)}{D}}\,(t - t_0)^{1/2}, \qquad t > t_0. \tag{4}$$

This is the stationary branch evaluated along a moving control. Observed coherence follows it only where the phase distribution can track that branch. The mean-field crossing control starts near its initial stationary density and crosses symmetrically about K_c. None of its scanned rates meets the declared tracking tolerance under that initialization. Slowing the crossing therefore does not by itself justify using this benchmark as a recovery time course.

What it does not give. Four qualifications, each of which the manuscript raises before a referee could.

1. **No saturation follows.** Where the curve goes after onset requires a specified subsequent coupling trajectory. Monotone coupling alone does not determine a logistic or any unique sigmoid.

2. **Latency needs measured dynamics.** An independently known $K(t)$ locates the stationary threshold crossing. The time at which an observed coherence criterion is reached additionally depends on the dynamics and its initial state. Inferring coupling from that same coherence trace cannot supply an independent prediction of latency.

3. **Critical slowing down.** Near a continuous transition, relaxation times diverge, so the adiabatic (branch-tracking) assumption is worst exactly where the prediction is sharpest.

4. **Finite-size floor.** A finite population carries non-zero fluctuating order below threshold, of order $1/\sqrt{N}$. An observable onset therefore requires a *predeclared* threshold above that floor, and a finite-rate trajectory need not show an infinite-slope cusp at all.

Part IX. Finite populations, finite ramps, and what simulation adds

The stationary theory describes fixed-parameter densities of an idealized infinite population. Applying it to a changing coupling requires adiabatic tracking in addition. A cortical population is finite, and its speed of adjustment must be measured. The simulations probe these differences.

24 Euler–Maruyama integration

A simulation approximates continuous motion by short updates. The *time step* Δt is the duration of one update. The Euler method uses the current systematic velocity to predict the change over that step; Euler–Maruyama adds the appropriate random increment. The notation $\xi_i \sim \mathcal{N}(0,1)$ means draw a Gaussian random number of mean zero and standard deviation one, independently for each site and each step.

What it is. The numerical scheme for stochastic differential equations. The deterministic Euler step $\theta \mapsto \theta + f(\theta)\,\Delta t$ acquires a noise term scaling as $\sqrt{\Delta t}$ rather than Δt—because Brownian increments over a step of length Δt have standard deviation $\sqrt{\Delta t}$. Applied to (1):

$$\theta_i \;\mapsto\; \theta_i + \Big[\omega_i + \sum_j K_{ij}\sin(\theta_j - \theta_i)\Big]\Delta t + \sqrt{2D\,\Delta t}\,\xi_i, \qquad \xi_i \sim \mathcal{N}(0,1).$$

For illustration only, set $\Delta t = 0.01$, $D = 0.5$, and the current systematic velocity to 2 in chosen time units. The deterministic increment is 0.02 radians; the noise increment has standard deviation 0.1 radians. A draw $\xi_i = -1$ would give a total increment -0.08 radians. A single step can oppose the drift. Persistent attraction is a statistical effect over many steps and oscillators.

1. Choose the population size, frequencies, coupling rule, diffusion, starting phases and duration. Decide whether coupling stays fixed or changes.

2. Compute every site's drift from the current phases, then update all sites with independent random increments. Recompute the order parameter.

3. Save observations at a declared interval. Early adjustment is the *transient*; a claimed stationary average needs a specified later window.

4. Repeat with different random-number *seeds*. A seed makes a random sequence reproducible; each repeated run is a *replica*. Their ensemble mean averages over runs, whereas a time mean averages within one run.

5. Reduce the time step while keeping the physical duration fixed to check integration bias. Increase N separately to examine population-size effects. A parameter *sweep* repeats the experiment over declared parameter values.

Independent uniform arrows have $\mathbb{E}[r_N^2] = 1/N$: cross terms cancel in the expected squared sum, leaving N unit self-terms divided by N^2. Their root-mean-square order is therefore $1/\sqrt{N}$. This *finite-size floor* is a fluctuation scale, not a hard lower bound. Correlated near-critical populations can fluctuate above this independent reference.

What it does not give. Every simulation reported is a finite run at sampled parameters, with explicit seeds and saved compact summaries. Equal-duration half-step controls probe numerical sensitivity near transitions; they do not establish convergence of the integrator, and no run discharges a hypothesis of the formal composition.

25 Ramp speed sets the apparent onset exponent

25.1 What finite ramp speed means

A *ramp* is a prescribed change in a control parameter over time. Here it is the coupling strength, not an oscillation's frequency or amplitude. A linear upward ramp has

$$K(t) = K_{\text{start}} + vt, \qquad v = \frac{dK}{dt} > 0.$$

Finite ramp speed means making that change at an actual nonzero rate, giving the system only finite time to adjust. It is distinct from finite population size N and from finite numerical time step Δt.

For a hypothetical dimensional example, increasing K from 1 to $3\,\mathrm{s}^{-1}$ over 10 seconds gives $v = 0.2\,\mathrm{s}^{-2}$. If $D = 1\,\mathrm{s}^{-1}$, the static threshold is $K_c = 2\,\mathrm{s}^{-1}$, crossed after 5 seconds. That crossing does not promise that a measured coherence criterion is reached at 5 seconds. These numbers illustrate units; they are not a calibration of human recovery.

Suppose a distribution needs a relaxation time τ_{relax} to adjust. During that time the target coupling moves by roughly $v\tau_{\text{relax}}$. Tracking is plausible only if this is small on the coupling scale of interest. Near a continuous transition the relaxation time grows very large—*critical slowing down*—so even a slow ramp can outrun the response close to threshold.

25.2 Crossing the threshold and observing escape are different events

What it is. Drive coupling through the threshold and watch when coherence rises beyond a declared fluctuation criterion. Call the threshold-crossing time t_0, and the first qualifying escape time t_{escape}. For a linear ramp, the coupling delay is

$$\Delta K = K(t_{\text{escape}}) - K_c = v(t_{\text{escape}} - t_0).$$

Coupling delay and time delay are different quantities. A faster ramp can increase the former without increasing the latter. The simulations study delayed passage through a noisy bifurcation. The Kibble–Zurek idea likewise concerns the failure of slow tracking near a transition; it does not by itself determine every escape criterion or fitted exponent in this finite model.

A *leg* is one ramp rate in the sweep. Right-censored means the run ended before the event was observed; it does not mean escape never occurs. The fitted A sets the vertical scale, K_0 is an estimated horizontal offset, and β_{eff} is an *effective* exponent over the chosen window. It need not equal the limiting stationary exponent β. A log–log fit takes logarithms of both axes: a power law becomes a line whose slope is its exponent. The *foot* of a curve means its earliest rising portion.

The statement. With $N = 2000$, $D = 1$, 32 seeded replicas and $v \in \{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}\}$, escape defined as the first of three consecutive samples at $r \geq 0.2$ after threshold:

- The 2 fastest legs are right-censored—at the fastest, only 5/32 replicas escape before $K = 2.5$.
- The 6 uncensored legs of 8 give a log–log fit $\Delta K \propto v^{0.444}$, with 95% slope interval $[0.394, 0.494]$, which does not contain 1/2; the decade-spaced legs among them have mean coupling delays 0.1947, 0.0684 and 0.0253. Fitting the faster and slower halves of the span apart returns 0.492 and 0.447, which agree but are each too few legs to resolve the shortfall alone. The reference for that fit is a deterministic mean-field control seeded at the finite-population fluctuation floor: it escapes at all 8 speeds, so the censoring above is the ensemble's, and it returns 0.425 over the 6 legs the fit uses, inside the interval on the fit. Tightening the escape criterion to $r \geq 0.1$ gives 0.482 and to $r \geq 0.05$ gives 0.490. This shows sensitivity to the criterion without explaining the entire finite-population difference. The larger-population low-threshold fit also has substantial log residuals, and some trajectories reach the absolute criterion before K_c; its fitted exponent is a finite-window summary, not evidence for one scaling law across these settings.
- At fixed $v = 10^{-2}$ the delay grows with population size (0.0619, 0.1947, 0.2341 at $N = 500$, 2000, 8000). However, dividing the delays by $\sqrt{\log N}$ does not give a common constant. At the smallest size, fluctuations can already meet the absolute escape criterion near threshold. The size trend does not establish the proposed scaling law.
- **The apparent onset exponent**, fitted as $r = A(K - K_0)^{\beta_{\text{eff}}}$ over $0.1 \leq r \leq 0.4$, runs 0.964, 0.898, 0.471, 0.417 from fastest to slowest ramp. Applying the same fit to the exact stationary branch on the same coupling grids gives 0.443, 0.443, 0.443, 0.448. Those reference estimates are below 1/2 because the fitting window extends beyond the asymptotically small onset region. The two slowest trajectories fall on either side of their references. At the fastest ramp, the ensemble mean never reaches the declared window; the returned value is a local slope of a delayed trajectory, not a measured critical exponent.

Why this matters. This is a methodological result with reach beyond the framework. **The exponent recovered by fitting a driven transition is a joint function of the mechanism and the drive rate.** An onset exponent read off a recovery trace therefore says little about the underlying transition unless the ramp rate is independently known and the finite-size floor is declared—and that holds for *any* analysis reading a critical exponent off a driven transition.

The first thing it constrains is the framework's own prediction: the stationary square-root law cannot be imposed on an arbitrary recovery trace, and its absence is not by itself evidence against a coupling-gated mechanism. A linear apparent onset counts against the adiabatic branch interpretation only in a regime where that approximation is independently justified.

What it does not give. The (a, r) collapse is more robust than the temporal shape: independently estimated (a, r) traces stay close to $I_1(a)/I_0(a)$ even where the trajectory lags the branch, with post-critical RMS deviations 0.00245, 0.00599, 0.00689, 0.00705, none reaching twice the slowest run's floor ($2\text{Dev}_0 = 0.01410$). The specified deviation-based critical ramp

speed is therefore *not bracketed* by this sweep. Mapping these dimensionless rates to human recovery times needs an independently calibrated phase diffusion and coupling trajectory.

RMS means root mean square: square deviations, average, and take the square root. A *collapse* here means points from different runs lie near one common relation after choosing the stated axes; it is not a loss of coherence. "Not bracketed" means the sampled rates do not include cases on both sides of the specified deviation threshold, so the sweep cannot locate its crossing.

26 Spatial coupling and geometric frustration

A spatial network specifies who interacts with whom as a function of position. A *lattice* is a regular grid; periodic boundaries join opposite edges to remove an outer boundary. An exponential kernel makes influence fall by a fixed factor over each fixed increase in distance; its decay length sets that distance scale. Row normalization rescales the inputs to each site to keep their total fixed as interaction range changes.

In a signed network, positive and negative entries give opposing influences. *Dale's law* is modelled by requiring a source site's outgoing weights to share a sign; excitatory and inhibitory sources have positive and negative signs here. Row balance means each target receives cancelling positive and negative weight totals, not that the phase-dependent forces always cancel. An Erdős–Rényi network samples each permitted edge independently with a fixed probability. A self-edge is a connection from a site to itself.

What it is. Two things could break coherence in a spatially extended sheet: coupling that decays too fast with distance to hold the sheet together, and *frustration* —signed (excitatory and inhibitory) couplings that cannot all be satisfied at once, in the way that three mutually antagonistic spins on a triangle cannot.

The statement. *Spatial decay.* On a periodic 128×128 sheet spanning 2.0 mm, an exponential kernel row-normalized to fixed total coupling at every decay length: with steady $r > 0.2$ as the operational criterion, the smallest interpolated decay length above which every larger sampled length stayed coherent was 0.0140 mm. At the empirical controls 0.1, 0.2 and 0.3 mm, steady order was 0.9424, 0.9427 and 0.9427, with winding-defect density at most 2.4×10^{-6}. So this sweep shows no fragmentation in the empirical band—which sits inside a plateau where steady order spans only 0.00038 out to the longest range sampled. The boundary itself is 0.90 lattice spacings, below that grid's own resolution; halving the spacing—a 256×256 sheet over the same 2.0 mm—leaves it at 0.01413 mm, 1.01 times its coarse value and 1.81 spacings on the finer sheet. So the operational crossing barely moves under this one refinement— which is agreement between two interpolated thresholds, not convergence of the sheet near them: at 0.01467 mm, the shortest length from which both sheets stay coherent, steady order is 0.6176 on the coarse sheet and 0.2687 on the finer one.

Frustration. A row-balanced Dale-law network (no self-edges, Erdős–Rényi edge probability 0.2, 80/20 excitatory/inhibitory presynaptic fractions, each row's positive and negative weights summing to $\pm g$) shows that row balance *alone* does not guarantee a floor-level incoherent baseline: signed coupling can amplify finite-size fluctuations, and $N\langle r^2 \rangle$ stays above the independent-phase value of unity. In a separately specified weak-synapse regime, adding a uniform per-pair modulatory term ϵ (contributing mean-field strength $K_{\text{eff}} = N\epsilon$) produces an operational coherence crossing at $K_{\text{eff}}/D \in [1.96, 1.99]$ across seeds. The ϵ grid is geometric with ratio 1.136, refined linearly to steps of 0.025 inside the step those values interpolate across; the refined bracket $[1.958, 2.009]$ still contains $K_c/D = 2$.

What it does not give. Each is one finite parameter regime. Fixing total coupling while changing range tests *fragmentation*, not whether cortex supplies that total coupling. The crossing range describes finite-run variability, not a confidence interval or an infinite-size critical point.

And the ratio K_{eff}/D is dimensionless—carrying it over to cortex needs the rate calibration γ of Section 49, which neither sweep supplies. No clinical mechanism is inferred from any of this.

27 Squared-drift plasticity does not learn structure

The question is whether changing connection strengths to reduce motion also makes the network represent a target pattern. These are separate scores. Imagine grouping oscillators by natural frequency: a template could reward strong connections within each group. An update reducing differences in drift may instead strengthen connections *between* groups to pull unlike frequencies together. A lower objective can therefore accompany worse agreement with that template.

Each experimental *arm* is one update rule applied to comparable starting conditions. A frozen arm keeps the couplings fixed. A random arm moves them in random directions. A permuted arm relabels which sites receive the proposed changes. A hollow matrix has zero diagonal; symmetric weights act equally in both directions. The Frobenius norm is a matrix's size, calculated by summing the squares of all entries and taking the square root. Holding the sum of weights fixed does not hold this norm fixed: concentrating weight on fewer edges increases the sum of squares.

A latent signal is a shared underlying influence in a model; covariance describes how signals fluctuate together. The template is such a specified covariance pattern, used to score the result. A percentile against relabelled templates reports how the chosen template ranks among those alternatives. It is not the fraction of biological information the model has learned.

What it is. It is tempting to read the predictive-dissipation bound as saying that a system lowering its dissipation thereby comes to mirror the structure of its input. The bound licenses no such reading: it bounds *non-predictive* memory, and a memory carrying no information at all satisfies it with room to spare, so lowering dissipation presses toward *less* representation rather than toward better-aligned representation. This control exhibits a system in which the pressure is not merely toward less but away.

The statement. Phases evolve with noise and clustered frequencies (three nearly equal clusters at 0.5, 1, 1.5 per unit time, with non-zero grand-mean drive), while symmetric couplings update every 50 fast steps by descending the squared-drift objective of Section 12. Couplings stay non-negative, hollow, and normalized to fixed total resource. The symmetric-edge derivative of $\frac{1}{2}\sum_i v_i^2$ is $(v_i - v_j)\sin(\theta_j - \theta_i)$, so both endpoints of an edge contribute. Each seed runs four arms on one phase-noise stream: the gradient step, a symmetric random step at the Frobenius norm of that arm's own gradient, that arm's own gradient step delivered to randomly relabelled nodes, and a frozen kernel. Each arm computes its direction from its own phases and couplings, so the controls follow the gradient arm's recipe rather than receiving its step.

The first thing to measure is how much of the objective the update removes. Diagnostics are sampled at the update cadence, at the instant each update has just acted, so every objective figure below is a mean of that post-update readout rather than a time mean over the intervals between updates. Symmetric coupling cancels from the mean drift, which fixes $\sum_i v_i$ and so puts a floor of $N\langle\omega\rangle^2/D = 990$ under the objective: the update can equalize drift across nodes but never remove it. Across three seeds:

- The frozen arm sits at 1380.6063 to 1382.3448; the gradient arm reaches 1110.5659 to 1116.6277, removing 0.6773 to 0.6914 of the headroom between the two. **So the update really does lower the objective where it acts.** The random arm removes −0.0173 to −0.0103, so this belongs to the gradient direction and not to displacing a kernel at fixed resource. A step size set the same way is not a matched deformation, though: isotropic steps cancel across updates, leaving that arm at a kernel norm growth of 1.2879 to 1.3056 where the gradient arm reaches 12.4094 to 14.6958. The permuted arm closes that gap: a relabelling moves an arm's own gradient entries onto edges that gradient did not choose, so

the arm deforms the kernel to 11.4185 to 16.3296 and holds order at 0.9184 to 0.9189 while *raising* the objective by -0.3319 to -0.1607 of the headroom.

- Second-half mean order lies between 0.9167 and 0.9222. **So coherence is maintained throughout.**
- The ratio of mean within-cluster to mean between-cluster coupling—a measure unchanged by rescaling the kernel—falls from 0.9824–0.9948 to 0.3507–0.6135, while the random arm leaves it at 0.9965–0.9997 and the comparably deformed permuted arm leaves it at 0.7290 to 1.2935, above the gradient arm's range and containing unity. Against 2000 node relabellings the final kernel is further from the true template than from a fraction 0.9510 to 1.0000 of them, where the permuted arm reaches 0.2680 to 0.8410.
- Scored against 2000 partitions of the same group sizes drawn blind to frequency, at most a fraction 0.0495 give a lower within-over-between ratio than the frequency partition does, against 0.0755 to 0.3480 for the frozen arm. **So the update moves against the structure the environment carries, rather than against block structure of any kind.**

Most of the available objective is removed at the sampled instants, coherence is maintained, and within-cluster coupling falls to a fraction of between-cluster coupling. **Coherence is not a proxy for representational alignment in this regime, and descending a squared-drift functional does not produce representational learning.**

There is a further question underneath the first bullet, and it is the one an argument from dissipation to representation actually needs: does the objective *stay* low between updates? Averaged over complete update intervals rather than at the update cadence, it does not. A run at this rate and cadence sits at 0.998 and 1.003 of its frozen arm over the two halves of its second half—what the update takes off at the instant it acts, the interval that follows returns. A bounded study, its grid and criteria declared before the runs, sweeps 6 combinations of learning rate and update interval looking for one that holds a 5% reduction while keeping coherence and losing alignment. None does; the best manages 1.22% to 1.36%. So what this regime exhibits is a coupling update that loses environmental structure while staying coherent, and not one that sustains a lower objective. A bounded grid failing to find such a regime is not a proof that none exists.

Nor is the direction of the failure accidental, and this is the part worth carrying away. Because symmetric coupling conserves the mean drift, minimizing $\sum_i v_i^2$ can only *equalize* drift across nodes—and equalizing drift rewards coupling between clusters of *unlike* frequency, which is the complement of a template rewarding coupling within clusters of *like* frequency. Objective and template are opposed by construction. That is not a weakness of the demonstration but its content: nothing about a system's dissipation fixes which structure its slow variables acquire, and which functional they follow is what decides it.

What it does not give. The template is a covariance of shared latent cluster signals, used as a diagnostic—it is not an input to the update, nor a measured covariance of the constant frequencies. Two readouts are recorded but do not carry the result. Frobenius distance to the template grows, but the kernel norm grows by a factor 12.4094 to 14.6958 over the same interval and a fixed *total* does not hold the *norm* fixed, so most of that growth is mass concentrating onto fewer edges rather than misalignment; this is why the scale-free ratio is the statistic quoted above. The off-diagonal correlation with the template ends at -0.0362 to -0.0208, near where it started: at this sparsity edge-level variance swamps a block-mean contrast. The learning rate 0.2 is chosen by a sweep from 0.001 to 0.4 as the rate at which that post-update readout falls furthest—a declared selection, not a rate at which the update is known to converge. Resource constraints and evolving phases remove the fixed-phase monotonicity guarantee of Section 12, so this is not a counterexample to that theorem but a demonstration of its scope. A plateau in the objective establishes neither stationarity of the adaptive process nor a thermodynamic entropy-production rate—the latter would require probability-current information the calculation does not contain. And the non-zero grand-mean frequency drives the collective coordinate, since

symmetric interactions cancel in its mean drift.

Part X. From discrete to continuum

28 Building a coupling matrix from a field

Discrete means represented by separate sites; *continuum* means represented at every point of a space. Think of estimating the area under a curve by rectangles. Subdivide the interval, sample the height in each piece, multiply by its width, and add. Refining the mesh makes the pieces smaller. Convergence means the approximation approaches the integral as the largest piece shrinks, provided the pieces cover the intended region correctly.

A triangulation organizes vertices, edges and higher-dimensional pieces into a mesh; a face is one of those constituent pieces. Measurable regions are regions to which the chosen measure assigns a size. Pairwise disjoint regions do not overlap. Anchoring a sample to its region prevents sampling far away from the area it is meant to represent. A modulus of continuity bounds how much a field's value can vary across a small piece.

The physical input here is a stress–energy tensor, an object recording energy, momentum and stress components relative to directions. A tensor is a rule for combining directional inputs; a $(0, 2)$ tensor takes two vector inputs and returns a number, linearly in each. Turning it into one nonnegative magnitude at each location requires a choice, just as one can summarize a matrix by its trace (sum of diagonal entries) or its norm (size).

What it is. The framework wants the coupling matrix A_{ij} to be *derived* from a continuum physical field rather than posited. The construction: triangulate the substrate; assign to each edge (u, v) a region of the manifold; and define the coupling weight as the integral, over that region, of a scalar magnitude read off the stress–energy tensor. Symmetry and non-negativity of A are then theorems rather than assumptions.

The statement. The bridge back to the continuum is *mesh refinement*: as the regions shrink, the discrete energy $\frac{1}{2}\sum_{u,v}\mu(R_{uv})f(x_u)$ should converge to $\int_S f$.

Stated over the bare class of triangulations, that is **false**—nothing forces the regions to cover the manifold, empty regions have diameter zero, and the everywhere-empty triangulation refutes it. Stated correctly it is a theorem, quantified over sequences of triangulations whose regions are measurable, pairwise disjoint across distinct edges, covering, and anchored to the embedded vertices, with mesh size tending to zero. The fixed-mesh error is bounded by $\varepsilon\,\mu(S)$ with ε a modulus of continuity of the integrand, and the symmetric double sum is proved equal to the midpoint rule over unordered edges.

Carrying the regularity conditions as a structural field, rather than checking them case by case, buys three further statements:

- `weight_self`: a vertex has no edge to itself, so the coupling matrix has zero diagonal.
- `face_of_weight_ne_zero`: a non-zero weight forces $\{u, v\}$ to be an actual face of the complex—the discretization cannot invent couplings between vertices the triangulation does not join.
- `total_weight_eq_setIntegral`: $\frac{1}{2}\sum_u\sum_v w(u, v) = \int_S |T|$ —total coupling strength is exactly the scalar stress–energy of the region covered. Nothing counted twice (disjointness), nothing dropped (covering).

That last identity is the precise sense in which the coupling matrix is derived from the continuum rather than posited.

What it does not give. Three limits. (i) The theorem gives *convergence*, not the $O(1/N^2)$ rate the numerics report. (ii) One field is deliberately left free and marked as a modelling choice: `scalar_magnitude`, the scalar read off the stress–energy tensor, constrained only to be non-negative.

A $(0, 2)$ tensor has no canonical scalar—the trace, an observer's energy density and the tensor norm are different functions, and the structure carries neither a metric nor a frame—so imposing one would pin a choice rather than derive it. That this is a genuine physical input is stated as a *theorem*: two instances over the same triangulation, tensor and regions, differing only in that field, satisfy the partition identity while every edge weight differs (`gridThermo_magnitude_matters`). (iii) The witness is one-dimensional with a constant stress–energy tensor, so it establishes satisfiability, not a physically interesting discretization.

29 The negative result: coarse-graining produces no kernel

Knowing the total weight of all connections is different from knowing the weight of each connection. Similarly, a single limiting energy does not determine a function of two locations. *Coarse-graining* means replacing fine detail with summaries on larger regions. It can preserve a total while discarding which pairs contributed to it. The product space $M \times M$ consists of pairs of locations; that is where a two-site kernel belongs.

What it is. It would be natural to read the mesh-refinement theorem as constructing the continuum kernel $K(x, y)$ that the rest of the framework uses. It does not, and the development proves this rather than noting it.

The statement. Three obstructions, all recorded as results:

1. `mesh_refinement_convergence` converges a *scalar* coupling energy. It produces no kernel and no field dynamics.
2. The discrete side has no product structure to pass to a limit: `edge_region` is a subset of M, not of $M \times M$, so there is nothing kernel-shaped there to converge.
3. The obvious repair—a kernel supported on the vertices—does carry the weights, and yet contributes *exactly zero* to the continuum coupling energy on an atomless substrate.

Point (3) needs a word on *atomless* measures. A measure is atomless when no single point carries positive mass—Lebesgue measure on an interval is the example. A kernel supported on finitely many points is then integrated over a set of measure zero in each argument, so its double integral against $\mu \otimes \mu$ vanishes identically, for every phase field and every choice of weights.

What it does not give. The manuscript is scrupulous about the direction this cuts. It is the continuum/discrete comparison of Part XIV applied to the framework's own coarse-graining step—the same argument that would separate a finitely-sited architecture from a continuum one also blocks the framework's own route from mesh to kernel.

All three obstructions are to *this* architecture. The construction in §10.16 evades none of them and answers each separately: its kernel comes from declared hardware rather than from a scalar limit, its discrete data are cell *pairs* and so are product-shaped, and its cells carry positive mass, so nothing is sited on a null set. What that construction still does not supply is the hardware. The continuum kernel remains a *modelling input*, and its identification with the brain's endogenous electromagnetic field is a further empirical commitment. E45 or E45Active (the thermodynamic bound $\to$ a continuum energy limit) and E56 (continuum $\to$ field) are where those two steps live.

Part XI. Sheaves, gluing, and the unity of a state

30 What a sheaf is

30.1 Assemble two overlapping maps

Imagine a system with sites A, B, C. One local description covers A, B and assigns masses $0.2, 0.3$. A second covers B, C and assigns masses $0.3, 0.5$. They agree at their shared site B, so they assemble into the global profile $(0.2, 0.3, 0.5)$. If the second instead assigns 0.4 to B, no single profile can restrict to both. Agreement is required only on the overlap.

Local refers to a region; *global* to the whole chosen space. A section is an allowed description over a region. Restriction keeps the part of that description belonging to a smaller region. A cover is a collection of regions whose union is the whole space. "Open" specifies the regions the topology permits as neighbourhoods; in an ordinary plane an open region includes a small disk around each of its points. On a discrete space every subset is open, so the three-site example needs no geometric subtlety.

For spatial measures, restriction keeps the existing mass on the smaller region; it does not renormalize it to one. Thus the local masses above need not individually sum to one, even when the global profile does. Gluing arbitrary normalized probability distributions is not automatic: the permitted local objects and restriction maps must actually have the stated gluing property.

30.2 The vocabulary for that assembly rule

What it is. A *presheaf* assigns data to each open region of a space, together with *restriction maps* that cut data down to smaller regions. If $\mathcal{F}(U)$ is the set of "local descriptions over U" and $V \subseteq U$, there is a map $\mathcal{F}(U) \to \mathcal{F}(V)$ discarding what happens outside V. The formal construction uses spatial measures: weights assigned to locations in each patch. Interpreting those weights, or other suitable local data, as descriptions of represented content is a further modelling step.

A presheaf is a *sheaf* when local data that agree on overlaps assemble uniquely into global data. Formally: given an open cover $\{U_i\}$ of X and sections $P_i \in \mathcal{F}(U_i)$ such that P_i and P_j have the *same restriction* to every intersection $U_i \cap U_j$, there is exactly one global section $P_X \in \mathcal{F}(X)$ restricting to each P_i. This is the *gluing axiom*, and it is the whole reason sheaves appear here.

The auxiliary notion is a *germ*: what a local description says near one point, ignoring differences farther away. Two descriptions have the same germ if they agree on some smaller neighbourhood of that point. The *stalk* collects all possible germs there. Formally it is a *colimit*: one object assembled from a system of local descriptions by identifying the descriptions that become equivalent on smaller neighbourhoods.

Why the framework needs it. Because synchronization alone does not give unity, and the framework needs a mechanism that assembles local states into *one* state. Sheaf gluing is that mechanism, and it delivers uniqueness for free.

The statement.

> **Theorem (Sheaf-theoretic unity).** Given the specified sheaf $\mathcal{F}$ of spatial measures over a substrate X, if local states $P_U \in \mathcal{F}(U)$ agree on all intersections $U \cap V$, they glue uniquely into a global section $P_X \in \mathcal{F}(X)$.

Both existence and uniqueness are derived from the local data. The construction uses local data only—no global state is supplied as a premise— and the statement that every local section is

the restriction of the glued one is a *theorem* about the object produced, not a hypothesis.

There is a further question: is this sheaf section represented by an ordinary finite measure? The development proves that it is when the cover is finite and its local sections have finite-measure representatives agreeing as measures on every overlap. The construction assigns each location to one patch and sums the corresponding restricted measures, so overlaps are counted once. It then proves that the measure's sheaf image is the section already constructed. The substrate can be continuous; it is the number of patches and each local mass that are finite. An infinite family of singleton patches with unit mass at every natural number has no finite total mass. Sheafification by itself therefore must not be read as a general finite-measure representation theorem. The measures used here are unnormalised spatial masses.

31 Compatibility is not phase agreement

This is the conceptual hinge of the whole framework, so it is worth stating slowly.

What it is. Take two cortical regions whose receptive territories overlap. Each carries a probability distribution over what occupies the shared part. Their oscillators can be locked to the *same phase* at every instant while one region's distribution places an object left of fixation and the other places it right. The phases agree; the descriptions do not; and *no single distribution over the shared territory restricts to both.*

Why the framework needs it. A theory that identifies unity with synchrony cannot mark that difference, and so cannot separate a brain whose regions oscillate together from one whose regions are about the same world. If the unity of experience is a matter of local contents being about one world rather than merely co-timed, then compatibility, not coherence, is the property to look for—with coherence entering as the physical condition under which compatibility can be *maintained*, not as a substitute for it.

The statement. The development separates two inputs to the gluing theorem, and only one of them is discharged by dynamics.

Thermodynamic equilibrium (discharged).
The phase field sits at the minimum of the coupling potential. This *is* supplied by dynamics on a class of initial data—the convergence theorem of Section 17 builds the required structure, and the witness is a three-patch cover whose phase field is the limit of a trajectory started *unsynchronized.*

Overlap agreement (not discharged).
Patches at a common phase agree where they overlap. **No dynamics in the development derives this.** It is a class field of the cover the theorem is handed, so it is assumed alongside the eight edges rather than carried by one of them.

Overlap agreement has a concrete pointwise reading: on a discrete substrate, two patch sections have the same restriction to their overlap exactly when they assign the same mass to every site they share (`restrict_eq_iff_densityOn_eqOn`). And it genuinely excludes data the rest of the structure admits—the development carries a **counterexample** rather than merely noting the possibility:

> A two-patch cover of a three-site substrate, all patches at *one common phase*, whose local density profiles differ at the shared site. It satisfies every other requirement of the gluing theorem and fails compatibility alone (`overlap_agreement_fails`).

That object is what a prose table of assumptions cannot carry. It is also the target the compatibility observable of Part XVIII would have to separate from its compatible counterpart.

The cover is itself a choice, and overlap agreement is the only demand the class makes of it. Over an empty overlap there is nothing to compare—the presheaf's value on the empty region is a singleton—so a cover by pairwise *disjoint* patches is an instance for every assignment of local data and phases (`LocalSectionSynchronization.ofDisjointCover`), and the gluing theorem still returns a unique global section, which is the tuple it was handed (`glue_unique_of_disjoint`). Membership of the class then reports nothing about the system beyond the patches covering it. `HasNonemptyOverlaps` is the non-degeneracy this identifies, stated as a predicate rather than a field, so no result above assumes it.

What it does not give. The theorem does *not* require patches to agree away from their overlaps—the three-patch witness carries three pairwise-different profiles. And the whole construction is a formalized gluing *mechanism* under an explicit compatibility postulate, not a derivation of compatibility from synchronization. Neither field continuity nor synchronization alone guarantees a globally compatible representation.

In Lean. `Phase5_GlobalSection.lean`: `probability_glue_unique`, `ThermodynamicCover`, `LocalSectionSynchronization.section_agrees_of_phase_eq`, `sync_to_section_eq`, `subsingleton_section_of_eq_bot`, `section_agrees_of_disjoint`, `glue_unique_of_disjoint`, `LocalSectionSynchronization.HasNonemptyOverlaps`; the equilibrium discharge in `Phase5_EquilibriumBridge.lean` (`ThermodynamicCover.ofConvergentTrajectory`).

31.1 Agreement to within a tolerance

A biological description will not agree exactly. So ask what survives if the patches agree only to within some ε: at every shared site, the two profiles differ by at most ε in mass.

What it is. A *partition of unity* subordinate to the cover is a set of weights $w_i(x) \geq 0$, each vanishing off patch i, with $\sum_i w_i(x) = 1$ at every site. It averages the family into a single profile,

$$(\text{select } w\, s)(x) \;=\; \sum_i w_i(x)\, s_i(x),$$

which is a weighted vote among the patches that can see the site x.

The statement. Three facts, in the uniform distance (the largest mass difference at any site). The selected profile is within ε of each patch on that patch. Two partitions subordinate to the same cover select profiles at most ε apart — and that constant does not grow with how many patches cover a site, because the weights sum to one rather than to the number of patches voting. At $\varepsilon = 0$ the selection is the glued state, so the exact case is recovered and nothing is given up there.

The three-site substrate makes the second fact concrete. Take the three two-site patches, each blind to one site; the common profile is $(2, 1, 3)$, and let patch 0 report ε more mass at each of the two sites it sees. Assign every site to the lowest-indexed patch containing it and the selection is $(2,\ 1+\varepsilon,\ 3+\varepsilon)$; assign every site to the highest-indexed patch containing it and the selection is $(2, 1, 3)$. Those two selections are exactly ε apart, so the bound is attained and cannot be improved.

What it does not give. At positive tolerance an exact extension need not exist. For every positive ε that family has no global profile restricting exactly to all its patches: patches 0 and 1 both see the right-hand site and report $3+\varepsilon$ and 3 there. So selection is a *choice* among states rather than an extension of the data, and a different choice of weights gives a different unified state. Uniqueness is replaced by a selection rule with a diameter, which is a weaker thing to have.

Two further cautions. The ε must hold on *every* pair, not on a spanning set of overlaps: put a 2ε discrepancy at the middle site, which one patch cannot see, and the pairs $(0, 1)$ and $(1, 2)$ agree exactly while the two selections are 2ε apart. And a readout that is L-Lipschitz in the

profile moves by at most $L\varepsilon$ between selections — total mass, with $L = 3$ here, actually moves by 2ε. Reading the selected state as a conscious content at resolution δ would require exhibiting the decoder and arguing $L\varepsilon \leq \delta$. Nothing in this construction produces the approximate agreement, picks the cover or picks the weights.

In Lean. `Phase5_GlobalSection.lean`, namespace `ApproximateGluing`: `Compatible`, `IsPartition`, `select`, `select_close`, `select_dist_le`, `approximate_diameter_le`, `select_glue_unique`, `feature_dist_le`; the three-site witness and its regressions in `Examples/Phase5.lean` §21 (`apx_select_dist`, `apx_no_global_section`, `chainW_select_dist_gt`).

31.2 Covers of content rather than of space

What it is. The patches so far overlap where they share sites. Visual and motor territories that agree on where a cup is overlap in *what they describe*, not in where they lie. The manuscript proposes a cover of that second kind: take the content variables as the base, give each region the variables it can decode, and let two regions overlap on the variables both decode.

The statement. The gluing results carry over. A section over a set of variables is a list of values for them, forgetting a variable is restriction, and that presheaf is a sheaf, so compatible decodes glue to exactly one assignment of values to every variable. Weighted selection works the same way with signed values. One thing changes: compatibility stops being a hypothesis. If two regions each decode a variable to within δ of its true value, the triangle inequality puts their decodes within 2δ of each other, whatever else is true. Agreement up to that level is what two accurate decoders deliver anyway, which is why the manuscript's proposed experiment looks for agreement *beyond* each region's accuracy. A two-region example attains the 2δ exactly.

What it does not give. The decoder class has to be limited. A region that carries the whole relevant state can, with an unrestricted decoder, decode every variable exactly, and then every region overlaps every other on everything. The measure picture and the coherence bounds of the next subsection do not carry over: they bound disagreement between spatial patches of oscillators, and nothing here connects a region's phase order to what it decodes. The tolerance is a worst case, standing in for the statistical criterion a real experiment would use.

In Lean. `Phase5_ContentCover.lean`: `decode_glue_unique`, `decode_glue_value`, `decode_compatible`, `select_decode_near_value`, `exists_exact_decoder_of_injOn`, `not_decodes_of_read_eq`; `Examples/ContentCover.lean`, §37.

31.3 Where the agreement comes from

What it is. On spatial covers, everything above takes the agreement as given. The obvious question is what would produce it, and the answer is the coherence the rest of the framework is about — but only for a particular kind of content, and the restriction is the interesting part.

The statement. Start from an identity. Writing r for the order parameter of N oscillators,

$$N^2 r^2 = \sum_{i,j} \cos(\theta_i - \theta_j),$$

because r^2 is the squared length of the mean unit vector and expanding that square pairs every oscillator with every other. The sharper form reads the same fact about the mean. Write $z_k = e^{i\theta_k}$ for the unit arrows and m for their mean, so $|m| = r$; then the squared distances of the arrows from their mean add up to

$$\sum_k |z_k - m|^2 = N\big(1 - r^2\big),$$

the variance of points on the unit circle. Two arrows are each within their own distance of the mean, and those two distances are two terms of a sum of nonnegative things, so for *every* pair

$$1 - \cos(\theta_i - \theta_j) = \tfrac{1}{2}|z_i - z_j|^2 \;\le\; N\big(1 - r^2\big).$$

A single global number bounds every individual phase gap. The factor N comes from comparing two summands with the entire sum, and it cannot be removed: one oscillator in antiphase among N has a gap of 2 against every partner while $N(1 - r^2)$ is just under 4.

Now the content. A phase is an angle, so a content that is a function of phase has to be a function of the point $(\cos\theta, \sin\theta)$ on the circle; a model that reads θ as a real number is not 2π-periodic and so is not a function of the phase at all. In the chord distance on that circle the squared separation is exactly $2(1 - \cos\Delta)$ — the same quantity the inequality bounds. Let every patch read one assigned oscillator through an encoder that can depend on the site being described. Its output may be a vector, such as two coordinates. All encoders must have one common Lipschitz bound L, and patches meeting at a site must use the same encoder there. The resulting observations agree on every overlap in norm to within $L\sqrt{2N(1 - r^2)}$. At $r^2 = 1$ the residual is zero and the agreement is exact, so perfect locking is the endpoint of this scale rather than a separate theorem.

Retained vector contents can be mixed with these observations: $s_i \mapsto (1 - \eta)s_i + \eta o_i$, for $0 \le \eta \le 1$. The triangle inequality bounds the next overlap mismatch by $(1 - \eta)\varepsilon + \eta\delta$ when the current contents and observations agree to tolerances ε and δ. Repeated updates give the upper bound $(1 - \eta)^n\varepsilon + \delta$. A common nonexpansive linear summary preserves agreement, but agreement of one summary can conceal disagreement in another vector component.

What it does not give. This does not contradict the counterexample that opened the section, and seeing why is the point. The counterexample's two mass profiles are chosen *independently of the oscillators*; nothing makes them functions of phase, so equal phases place no constraint on them. The positive bound needs shared encoders on overlaps and a uniform Lipschitz condition. A formal witness has perfectly locked phases and disagreeing vector reports when the shared-encoder condition is dropped.

That turns the framework's central distinction into an empirical question instead of a modelling convenience. Which kind a neural description is — how much of its content is carried by phase and how much is free of it — is something a decoder study can ask. It must also test whether the local readings share the encoder required by the bound.

The encoder and its constant are declared hardware, in the same sense as anything else the model supplies. The agreement is derived rather than declared, which is the whole point of the section. The contents are normed vector fields identified with no neural variable, and the bound is attained on the smallest example: two antiphase oscillators give a residual of 2, which is the actual disagreement. The scalar case recovers the mass-profile setting only when nonnegative initial profiles and observations are supplied. Vector-valued agreement itself does not supply local coordinate changes or a probability-sheaf interpretation.

In Lean. `Phase4_KuramotoDynamics.lean`: `order_parameter_r_sq_eq_mean_cos`, `sum_normSq_sub_order_parameter`, `chord_sq_le_of_coherence`, `cos_gap_le_of_coherence`, `circlePoint`, `chord`, `chord_le_of_coherence`, `chord_eq_zero_of_phase_locked`. `Phase5_ContentDynamics.lean`: `SharedEncoder`, `UniformLipschitzEncoder`, `compatible_of_shared_coherence`, `compatible_of_shared_phase_locked`, `compatible_map`, `run_residual_floor`, `run_residual_of_shared_coherence`; vector witnesses in `Examples/VectorContent.lean`, §22–23; the antiphase witness, its computed $r^2 = 0$ and the disagreement control in `Examples/AgencyFoundations.lean`.

31.4 Three ways off the population count

What it is. The factor N above has exactly one source, and it is worth staring at: every term of $\sum_k |z_k - m|^2 = N(1 - r^2)$ is nonnegative, so the two terms for the pair in question are at

most the whole sum. That single step is what lets all the disorder in a population sit at the two sites you happen to be asking about, and it is why the estimate has to assume the worst. Three results decline it, in the three ways it can be declined: by asking for fewer pairs, by asking over shorter distances, or by not reading phase differences off r at all.

The statement. *Count instead of bound.* Ask not how large one gap can be but how *many* gaps can be large at once, and the pair identity becomes Markov's inequality. If every pair in a set S is separated by a chord of at least c, then $|S|/N^2 \le 2(1-r^2)/c^2$ — with the population size gone. At $r = 0.99$ and c one Lipschitz unit that is four per cent of pairs; at the patch-local order of the winding sheet, 0.9015, it is 0.37 of them, weak but scale-free, and improving like the square of what the uniform bound improves like. Reading the sum by rows instead of by terms gives the same thing for sites: a site with many distant partners spends its own row, so most sites agree with most sites.

Chain instead of compare. Give the cover a nerve — two patches are adjacent when they share a site — and use the fact that the chord is a metric. A bound β on the separation inside every patch then composes along a walk: k hops give $(k+1)\beta$. The $k+1$ is not slack. Both endpoint sites pay a step inside their own patch, while the shared sites of the walk are interior to the chain. Nothing in this needs a patch containing both endpoints, which is what makes it say something about two territories far enough apart to lie in no common patch — and that is the situation the whole compatibility question is about. At the winding sheet's nearest-neighbour value, $\beta = 0.0981\,L$, the chain stays inside the maximum chord for 19 hops: 20 patch diameters, or 0.312 mm of that sheet.

Do not read r at all. At a configuration turning rigidly at a common rate Ω, the phase differences are pinned by the dynamics themselves: $\omega_i - \Omega = \sum_j K_{ij} \sin(\theta_i - \theta_j)$ at every site, which is the Kuramoto equation at such a solution rather than an extra assumption. Pair that with the configuration, symmetrize, and apply Jordan's inequality termwise — keeping the sine and paying $2/\pi$ rather than linearizing — and the coupling graph's Dirichlet form appears. Its algebraic connectivity λ_2, the same quantity as the graph Laplacian's spectral gap on an unweighted graph, then closes the estimate: $2\lambda_2\|\theta - \bar{\theta}\| \le \pi\|\omega - \Omega\|$. There is no N in that, and the bound *improves* as the coupling graph becomes better connected — which is the dependence physical intuition expects and exactly the one an order parameter cannot express, r being blind to who is coupled to whom.

What it does not give. The first two buy their freedom with a quantifier and the third with a hypothesis, and neither trade is free.

A count names no pair and no site. The exceptional set it permits is unlocated, and a cover's overlaps are a set of sites fixed before the state is known, so reading a fraction as agreement on an overlap needs the exceptions to miss that overlap — a joint fact about the state and the cover that no coherence hypothesis supplies. The only way to get it is to choose the cover knowing the state, which is the move that empties a gluing claim. A guarantee about a named overlap still needs the uniform bound, at the grain the uniform bound is stated for.

The chain supplies no walk. Connectivity of the nerve, and its diameter, are properties of the declared cover; a cover whose nerve is disconnected says nothing between its components. Accumulation is linear and nothing here makes it sublinear, so the reach is some 20 patch diameters rather than a hemisphere, and the per-hop constant is the cover's own.

The spectral-gap route declares four things and discharges none. λ_2 is hardware data, in the same standing as the mode-price constant κ: Mathlib has the graph Laplacian and its positive semidefiniteness but no Fiedler value, so the Rayleigh inequality is carried as an obligation, and it is computed here only for the complete graph. A locked configuration is assumed to exist, which above the critical coupling is true and below it is false, and neither direction is proved. The quarter-turn confinement is assumed, and locking does not imply it; it is also what fixes where the route is worth running, since confinement to a quarter turn already puts every chord below $\sqrt{2}$ before any coupling datum is read, so the estimate says something new only where

$\lambda_2 > \pi\|\omega - \Omega\|/2$. And the detuning enters in the population's ℓ^2 norm, which is extensive: a population whose frequency spread grows with its size still pays for that growth, through the data rather than through the estimate. What the route removes is the unconditional factor, not the physics of heterogeneity.

In Lean. `Phase4_KuramotoDynamics.lean` §11: `sum_gap_eq`, `card_gap_le`, `card_chord_le`, `chord_fraction_le`, `card_site_gap_le`, `site_fraction_le`; §12: `patchNerve`, `patchNerve_shared`, `chord_le_of_patch_walk`, `chord_le_of_patch_walk_coherence`, with the metric lemmas `chord_triangle` and `chord_le_abs_sub` in §5. `Phase4_RotatingFrame.lean` §8: `couplingForm`, `couplingForm_eq_lapMatrix`, `SpectralGap`, `SpectralGap.scale`, `is_frequency_locked`, `is_frequency_locked_iff`, `sum_mul_coupling_sin`, `couplingForm_le_pairing`, `spread_le_of_frequency_locked`, `chord_le_of_frequency_locked`. `Phase5_ContentDynamics.lean`: `compatible_of_patch_nerve`, `compatible_of_frequency_locked`. Witnesses in `Examples/Phase4.lean`, §34–35.

31.5 How soon can a patch respond?

What it is. Agreement at one instant says nothing about how quickly a new value reaches a patch. In a synchronous network, every site updates from its own state and messages from specified neighbours. After T rounds it can depend only on initial states within its causal past: sites whose messages can reach it in at most T hops.

The statement. Two initial configurations identical throughout that past produce the same local state at round T. This is derived from the update rule. If a fresh intervention changes a value outside the past, the site cannot distinguish the possible values. No report read from its state can be accurate to ε for two values separated by more than 2ε. The reconstruction inequality of Section 37.1 makes the last step quantitative.

A three-site line takes two rounds to relay a value from one end to the other. The formal example fails after one round and reports exactly after two. Cutting the path blocks delivery at every round.

How long the wait is depends on how widely each site listens. If no site hears from more than ν others, then the causal past after T rounds holds at most $1 + \nu + \cdots + \nu^T$ sites, which is below $(T+1)\nu^T$. While that count stays under the number of sites there is always somewhere the world can be changed that the reading site has not heard about, so the obstruction fires without anyone having to point at a particular site: bounded fan-in produces one by counting. Reaching all N sites therefore takes about $\log_\nu N$ rounds. Read backwards, the same inequality prices the deadline in wiring instead of in time. A network that does reach everything in a single round must have a site listening to all but one of the others, which is a full crossbar. Whether the rounds or the wires are the scarce thing decides which reading is the useful one.

At the other extreme the bound has nothing to say at all. If every site hears from every site—which is how a mean-field coupling kernel is idealized—then after one round the causal past is already everything, and no site is left outside it for the obstruction to use. This is not an escape from delay. A physical field travels at a finite speed, so treating one this way assumes that transit across the substrate is quick compared with the phase dynamics, and that is a measurement rather than a theorem. The two extremes are the same count at its endpoints, and neither is a statement about substrates.

What it does not give. A constant report may happen to be right about one fixed world. What fails is a guarantee across fresh interventions. Static compatible descriptions remain possible. Applying the bound to a brain or computer requires its actual communication paths, including memory and sensory routes, and a justified deadline. The counting version adds nothing to that: the fan-in and the site count are declared numbers like the graph and the deadline, and what the wiring corollary bounds is the communication graph a guarantee would require, not what hardware can be built. Rounds are hops, so none of this is a claim about how fast anything computes.

In Lean. `Phase6_Locality.lean`: `Network.run_eq_of_agree_on_ball`, `not_reconstructs_of_outside_past`; `ball_rank_le` for an architecture that orders its sites, and `card_ball_le` with `card_ball_le_mul_pow` for the counting form, feeding `not_reconstructs_of_bounded_degree`. The wiring reading is `card_le_geomSum_of_reaches` and `le_degree_of_reaches_one`; the full-support extreme is `not_outside_past_of_isFullSupport`. `Examples/Locality.lean`, §25, supplies the delay, cut and snapshot controls and reaches the round-one rejection a second time from the fan-in alone; §31 is a causal mask, where `mask_ball_subset_le` keeps the past of a token position inside the positions at or before it at every depth.

32 Where uniqueness comes from

Return to the three-site example. Once both compatible local profiles are given, every global entry is fixed. That is uniqueness *for those inputs*. It does not choose the sites, decide which measurements define the profiles, or select this cover over another candidate cover. A theorem can settle the assembly question without settling which biological system to assemble.

What it is. Accounts of conscious unity need something that picks *one* experience from among the candidates a system could support. Integrated information theory postulates an *exclusion* principle for this purpose (Tononi et al., 2016).

The statement. Here uniqueness arrives twice as a *consequence* rather than a postulate: given a cover whose local sections agree, the sheaf condition gives exactly one global section; and given a contracting self-prediction map, Banach's theorem (Part XIII) gives exactly one fixed point.

What it does not give. Uniqueness is relative to a cover and a map. Choosing the regions, overlaps and encoding region remains part of identifying a candidate subject. E78 and E89 connect supplied objects; they do not select those objects from neural data. The theorems settle extension and fixed-point uniqueness within a construction, leaving selection between constructions to a biological model.

Part XII. Topological obstructions in the phase field

33 Twisted gluing, and why it fails to add content

Two maps can use different compass zeros while describing the same directions. If one map is rotated by an angle, their overlap disagrees by a constant offset. Adjusting each map's zero may remove all such differences. "Twisted" gluing asks when this kind of coordinate adjustment makes local data agree.

For three patches with absolute phase choices ϕ_1, ϕ_2, ϕ_3, the offsets around their loop add to

$$(\phi_1 - \phi_2) + (\phi_2 - \phi_3) + (\phi_3 - \phi_1) = 0.$$

Every term cancels. Calling these offsets a *coboundary* says they come from differences of choices on individual patches, so changing those choices removes them. An arbitrary collection of pairwise offsets need not have that property.

What it is. A natural weakening of exact overlap agreement: allow local states to differ by a *phase* on each overlap, and ask when the family still glues after one rotation per patch. In sheaf-theoretic language this is asking whether the family of offsets is a *coboundary*—whether the offset on $U_i \cap U_j$ can be written as $\phi_i - \phi_j$ for some per-patch phases.

The statement. Both directions are proved: coboundary offsets glue after one rotation per patch, and conversely, when the phase action is free on overlaps, a family that glues up to phase has coboundary offsets.

A phase action is a rule for rotating the local data. It is free if only the zero rotation leaves the relevant overlap data unchanged. A quotient groups objects treated as equivalent. Cochains are lists of values assigned to patches or overlaps; Čech cohomology organizes which such overlap data cannot be removed by making local choices.

And that is precisely why the construction adds nothing here. Any Kuramoto configuration—including a twisted or splay state—supplies absolute patch phases ϕ_i, so its offsets $\phi_i - \phi_j$ are *automatically* a coboundary, and the twisted family always glues. Phase frustration generates no obstruction by this route.

What it does not give. The formal quotient is deliberately *not* called Čech cohomology, because its total cochains still assign irrelevant values to patch pairs with empty overlap. Naming it after the real thing would overstate what is built.

In Lean. `Phase5_TwistedGluing.lean`: `gluesUpToPhase_of_isCoboundary`, `isCoboundary_of_gluesUpToPhase`, `isCoboundary_of_phaseField`, `PhaseObstruction`.

34 Phase singularities and winding numbers

An angle says where a hand points; an *unwrapped* angle also counts how many turns it has made. Choosing that real-number version continuously is called a *lift*. Walking round a loop while the phase makes one full turn brings the pointing direction back to its start, but increases the unwrapped angle by 2π. The winding number records that net turn count, with its sign recording direction.

For a signal pictured as an arrow, amplitude is arrow length and phase is its direction. A zero-length arrow has no direction. This is why a point of zero amplitude can be a phase singularity: the amplitude is defined but the phase is not. A spiral is one possible pattern around it; not every amplitude zero has nonzero winding. The analytic-signal construction used to extract such arrows from recordings is explained in Part XVI.

What it is. A narrower and genuinely positive answer comes from the data. The *analytic signal* $z = Ae^{i\phi}$ assigns an amplitude and a phase to a recording at each point. Around a point where the amplitude vanishes—a *phase singularity*, the centre of a spiral or rotating wave—the circle-valued phase need not admit a single real-valued *lift*. Local lifts exist on small patches, and on overlaps they differ by integer multiples of 2π. Summing those integer transitions around a closed loop of electrodes gives the *winding number*, an integer that counts how many times the phase wraps.

This is measured, not stipulated: Townsend and colleagues (2015) identify phase singularities in multielectrode primate LFP recordings by an integer winding number around electrode loops, with the associated complex waves modulating spiking; Xu and colleagues (2023) find task-dependent spiral centres and rotation directions in human fMRI phase maps.

The statement. The formalization treats the finite electrode loop: `loopWinding` sums its integer transitions, `loopWinding_eq_zero_of_hasGlobalLift` proves that transitions arising from a single global lift telescope to zero—so non-zero winding *obstructs* a global lift—and a winding-one witness on three electrodes shows the obstruction is realized.

The same integers are available from a state rather than from a recording. `ringTransition` is the number of whole turns the q-fold winding's affine local lifts drop between consecutive sites: zero inside the ring, q across the wrap where the residue resets (`winding_succ`). Summed around the loop they give q (`loopWinding_ringTransition`). What this buys is a pairing on one state at one coupling: the global resultant is exactly zero and the degree is q (`winding_degree_obstructs`). The resultant is an average over sites and moves when they are reweighted or the patches

redrawn; the degree telescopes, so no redrawing of the loop's intermediate transitions changes it. That is the sense in which the winding state is an invariant and not only a counterexample.

The following notation spells out where phases can be measured. The *punctured domain* is the surface with singularities removed. The *nerve* of a cover records which patches overlap: a node for each patch, an edge for each pairwise overlap, and higher pieces for shared multiple overlaps. Its loops describe closed routes through patches. The cover-level specification is explicit: cortical surface M, analytic signal $z = Ae^{i\phi}$, singularity set Z_t, punctured domain $M_t^* = M \setminus Z_t$, and measurement map $q_t = z/|z| : M_t^* \to S^1$. A trial-independent electrode or source-space cover of M_t^* supplies local real lifts; their differences divided by 2π are integer transitions on connected overlaps; and loops in the nerve of the cover are the inputs to `loopWinding`.

What it does not give. Three limits, stated firmly. First, this does *not* rescue the twisted-gluing quotient: that quotient uses circle-valued differences of absolute patch phases, while the measured obstruction uses integer transitions between local lifts. They are different objects. Second, **no theorem connects winding sectors to content.** A sector groups configurations by their winding numbers. A proposed test compares two classifiers, rules trained to predict a content label from measurements. One uses stimulus, task, behavior, eye movement and arousal (level of wakefulness); the other adds winding, estimated brain-source location of singularities, rotation direction and lifetime. They are nested models because the second contains the first model's inputs plus more. Preregistration fixes the analysis before seeing its outcomes. Held-out participants and sessions are reserved for evaluation rather than training; shuffling content labels within comparable groups supplies a control for chance prediction. The added information must improve prediction on these unseen data, replicate, and survive a changed or absent motor report, such as a button press. Source localization is inference about the brain locations generating recorded signals, with its own uncertainty. A positive result would be an empirical *association*, not an identification of a topological sector with phenomenal content; a null increment rejects the proposed role. In finite-resolution data the amplitude threshold defining Z_t, the loop geometry and the required cross-scale persistence are fixed using training or baseline data, never the content labels of test trials.

The construction and its test sit *outside* the conditional chain. A full nerve-indexed integer complex is deferred until a measurement protocol warrants it.

Third, the framework's own dynamics do carry an amplitude beside the phase—a Stuart–Landau layer whose phase equation is exactly the Kuramoto equation with each coupling weighted by a ratio of amplitudes (`amplitude_phase_is_kuramoto`), so the phase model is that layer with the ratios set to one. This does not put a spiral inside the framework, and the reason is not that the sites form a ring—they need not, the results holding for any finite abelian group, the periodic sheet included. It is that every site of these solutions has the same arrow length (`norm_charState`), so the phase winds without the amplitude ever vanishing, where a phase singularity is precisely a point at which it does vanish. Around every small square of the sheet the phase therefore returns exactly to where it started (`plaquetteCirculation_eq_zero_of_winding`); and on a closed sheet one defect cannot stand alone in any field whatever, exact or not, since the circulations over all squares cancel (`no_isolated_defect`). What the layer buys instead is a band: a winding of a given number exists only when one inequality between the coupling kernel and the distance from the bifurcation holds (`exists_windingState_iff`), where the phase model admits every winding number at every coupling. The two layers differ about the state space and not about a measurement: broadening the kernel takes the winding out of existence here, and leaves it in the phase model as a state that no longer survives integration. Three things the band is not. The distance from the bifurcation is measured nowhere here, so it locates nothing; it is a band of existence rather than of stability, no stability result being proved; and the arrow length of the model is a normal-form radius rather than the envelope of a recording, which is an identification the development does not make.

In Lean. `Phase5_PhaseLifts.lean` (`ringTransition`, `winding_succ`, `loopWinding_ringTransition`, `not_hasGlobalLift_of_loopWinding_ne_zero`, `winding_degree_obstructs`), `Phase4_KuramotoDynamics.lean` §7.

Part XIII. Metric spaces, contractions, and the reflexive self

35 Banach's fixed-point theorem

Start with the update rule $f(x) = (x+2)/2$. From 0, repeated application gives $1, 1.5, 1.75, \ldots$, approaching 2. Starting somewhere else also approaches 2, because each update halves the distance to it. The number 2 is a *fixed point*: $f(2) = 2$. Iteration means applying the same rule repeatedly. It need not represent physical time unless a model explicitly gives it that meaning.

A *Cauchy sequence* has terms that become arbitrarily close to one another sufficiently far along. Completeness says their limiting point belongs to the space. If we removed 2 from the real line, the example sequence would approach a missing point; that space would not be complete. A distance must be nonnegative, symmetric, zero exactly for equal states, and obey the triangle inequality: going via a third point cannot shorten the direct distance.

What it is. A *metric space* is a set with a notion of distance. It is *complete* when every Cauchy sequence converges—no limits are missing. A map f is a *contraction* with rate $\Lambda < 1$ if it shrinks all distances by at least that factor: $d(f(x), f(y)) \le \Lambda\, d(x, y)$.

> **Banach's fixed-point theorem** (Banach, 1922). A contraction on a non-empty complete metric space has exactly one fixed point, and iterating the map from any starting point converges to it.

The proof is three lines: iterates form a Cauchy sequence because distances shrink geometrically; completeness supplies a limit; continuity makes the limit a fixed point; and two fixed points would have to be at distance $\le \Lambda\, d$ from each other, forcing $d = 0$.

Why the framework needs it. The framework proposes that a minimal self involves an accurate internal representation of the state whose contents form the episode. Its map encodes the global state locally and reconstructs it through a readout. If the map contracts, Banach gives exactly one fixed point. Identifying that point with the state glued on the chosen cover is an additional hypothesis; interpreting the reconstruction as a minimal self is a philosophical commitment. The map does not describe a temporal prediction without a further dynamical model.

The statement.

> **Theorem (Conditional auto-resonance).** On a nonempty complete metric space of global sections, a self-prediction map with Lipschitz rate
>
> $$\Lambda \;=\; e^{-(K-K_c)\tau/2} \;<\; 1$$
>
> has a unique fixed point.

Linearizing means approximating the dynamics of a small disturbance by its first-order terms. In the near-threshold amplitude approximation $\dot r = \frac{1}{2}(K-2D)r - cr^3$, $c > 0$ sets the saturation strength: the cubic term slows growth at larger amplitudes. The rate in the exponent, $(K - K_c)/2$, is the growth rate of the incoherent state's instability, so supercriticality $K > K_c$ makes the factor less than one. Relaxation about the *coherent* branch is twice as fast—linearizing

$\dot{r} = \frac{1}{2}(K - 2D)r - cr^3$ about its stable fixed point gives $(K - K_c)$—so Λ is the larger of the two factors and the Lipschitz obligation stated with it is the weaker. Both are below one on exactly the half-plane $K > K_c$. Here $\tau > 0$ is a specified interval, and a Lipschitz bound says output distances are at most Λ times input distances. Below threshold this particular formula gives $\Lambda \geq 1$, so it does not supply the strict factor below one required by the contraction theorem. Other maps can still be contractions at other rates; what fails is this rate's contraction condition, not the existence of every possible contraction.

What it does not give. Coherence supplies $K > K_c$; it does not supply the Lipschitz law. The bound is an obligation on each instance, not a consequence of field dynamics, and no field dynamics in the development produces the read-out map. E89 carries both the Lipschitz law and the identification of the glued section with the fixed point—Banach supplies uniqueness, not that identification.

35.1 An output's information ceiling

An emitted token can distinguish at most its vocabulary's worth of outcomes. The information statement is $I(X;Y) \leq H(Y) \leq \log_2 |V|$ bits for a finite joint law of internal state and emitted symbol. Over k tokens the ceiling is $k \log_2 |V|$, even when emissions depend on one another. No tokens convey zero information. The formalization checks both a binary copy carrying one bit and an independent binary output carrying none. This counts what is emitted; internal cache state or a timing channel is a different output variable.

35.2 The error of a linear bottleneck

A linear encoder can retain only as many independent directions as its code space has real coordinates. This gives a quantitative comparison when the target is a finite spatial coupling operator with orthonormal eigenmodes. Order their eigenvalues by magnitude. If the encoder and readout factor through d coordinates, their composite A satisfies

$$\sum_{i=1}^{N} \|Ke_i - Ae_i\|^2 \geq \sum_{i=d+1}^{N} \lambda_i^2.$$

Retaining the first d eigenmodes attains equality. With eigenvalues three and one, retaining one mode pays squared error one, and every other rank-one linear approximation pays at least one. The proof allows approximations that mix the modes; it does not assume they discard eigenmodes individually.

The restriction must describe the actual operator. A softmax attention matrix need not have the rank of its logits: logits $\left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ have rank one, whereas their row softmax has rank two. Consequently query/key dimension alone does not make this bound a restriction on ordinary softmax attention. The target spectrum also has to be supplied. These are finite-dimensional linear approximation results, not a universal distinction between digital machines and continuous media or a bound on the number of stored bits.

When task inputs have unequal or correlated second moments, the linear floor uses the singular values of the target after an input-covariance factor. On a shared synthetic task, saved linear decoders attain that training-law floor. They need not preserve their accuracy on a new family: a full-capacity decoder trained where two scene coordinates always agree fails when those coordinates disagree. The declared task map reconstructs the noiseless test states exactly; the restricted training family cannot identify that map uniquely. Repeating the noiseless linear fit at different seeds leaves the input arrays unchanged. Small attention and coupled-phase controls also fail under the declared training budget. A three-node phase chain gives a separate causal-past check: an intervention two graph edges from its encoding node cannot change the

code in fewer than two synchronous rounds. These are measurements of the specified codes and task, without hardware-energy or experience claims.

35.3 Approximate reconstruction

A state need not reconstruct itself exactly to be close to the fixed point. Suppose its reconstruction error is at most ε, measured with the same metric as the contraction. Since $F(s_*) = s_*$,

$$d(s, s_*) \le d(s, F(s)) + d(F(s), F(s_*)) \le \varepsilon + \Lambda d(s, s_*).$$

Moving the last term to the left gives

$$d(s, s_*) \le \frac{\varepsilon}{1 - \Lambda}.$$

The exact fixed point remains unique; several approximately reconstructed states may lie near it. For the three-site witness, the silent profile $(0, 0, 0)$ reconstructs as $(0, 1, 0)$, while the fixed profile is $(0, 2, 0)$. The reconstruction error is one and the distance to the fixed point is two. With $\Lambda = 1/2$, the bound is attained exactly.

The denominator measures how strongly successive predictions contract. At the stipulated contraction factor $\Lambda = e^{-(K-2D)\tau/2}$, the proved bounds are

$$\frac{2\varepsilon}{(K - 2D)\tau} \le \frac{\varepsilon}{1 - \Lambda} \le \frac{2\varepsilon}{(K - 2D)\tau} + \varepsilon \qquad (K > 2D,\ \tau > 0,\ \varepsilon \ge 0).$$

For fixed positive error tolerance and interval, the radius therefore grows without bound as coupling approaches the threshold from above. The estimate loses precision; it does not say that an actual state moves arbitrarily far away. At zero reconstruction error, its distance from the fixed point is zero at every supercritical coupling.

If the state is also displaced by at most δ, its new reconstruction error is at most $\varepsilon + (1 + \Lambda)\delta$, giving distance at most $[\varepsilon + (1 + \Lambda)\delta]/(1 - \Lambda)$. The errors add and the inverse contraction gap amplifies the bound. This calculation holds the map fixed and assumes the displacement bound. It supplies no mechanism for achieving agreement between local descriptions.

Strict contraction matters. On the same metric, identity encoding and readout through a whole-space avatar give $F(s) = s$ for every state. This map is nonexpansive, with factor one, and both silent and baseline profiles are fixed points. The result locates the failure of the uniqueness guarantee at the loss of strict contraction. Choosing a biological state space, metric and readout remains a modelling task.

36 The metric on global sections, and why the obvious ones fail

Consider two mass profiles $(0.2, 0.3, 0.5)$ and $(0.2, 0.4, 0.4)$. Their largest sitewise difference is 0.1. This *uniform distance* measures the largest local error in reconstructing a profile. A 0/1 distance instead calls any unequal profiles distance one, whether they differ slightly or completely. That discards the gradations an informative contraction needs.

Total variation compares how much two measures can disagree on a whole event or region, rather than one site. Lévy–Prokhorov distance also allows a small spatial enlargement of regions before comparing their probabilities. These are alternative ways to measure discrepancy, with different contraction and completeness properties; naming one does not establish the needed theorem.

What it is. To apply Banach one needs an actual complete metric on the space of global sections. Two natural candidates fail, and the development records both failures.

The statement. *The discrete (0/1) metric fails informatively.* Under it, every contraction is constant (`contracting_implies_const`)—because distances take only the values 0 and 1, and shrinking 1 by any factor < 1 forces it to 0. So that metric supports no informative fixed point.

The Lévy–Prokhorov metric—named in the file header—is not *what is constructed*, and the supplement says why. On a discrete substrate the ε-thickening of a set is the set itself for $\varepsilon < 1$, so the averaging map $\mu \mapsto \frac{1}{2}\mu + \frac{1}{2}\mu_0$ is not a Lévy–Prokhorov contraction there; and Mathlib carries no completeness result for that metric anyway.

What is constructed is the uniform distance between densities: the largest difference between the masses two sections' measures put on any single site. It is proved complete by exhibiting the limit section of a Cauchy sequence of densities, and distinct sections at distance 1/2 are exhibited so the space is not degenerate. On a finite substrate this is equivalent to total variation of the glued measures within a factor of the number of sites—but *not equal* to it, and no claim of equivalence with Lévy–Prokhorov is made.

36.1 The germ–measure dictionary

The abstract local descriptions must be translated into numbers before their distance can be calculated. On a discrete substrate, a point is itself its smallest neighbourhood. Its germ therefore determines a single site's mass. Reading those masses at all sites turns a section into a mass profile. Conversely, the permitted mass profile defines a measure and hence a section. The result `massEquivOn` proves this correspondence works in both directions without identifying distinct sections or missing profiles. That bijection is the dictionary making the uniform-distance construction possible.

37 The avatar: making the self-model run through a region

An *encoding* writes information about a global state into a chosen local region. A *readout* reconstructs a global state from that local record. The framework calls the region the avatar; the word does not imply a picture, a person-shaped representation or a digital character. The composite map means "write, then read":

$$s \longmapsto \text{write}(s) \longmapsto \text{readout}(\text{write}(s)).$$

At a fixed point, the reconstruction equals the original state. A predicate is a yes/no condition on an object. The *restriction-resonance* predicate requires the write to be the state's own restriction to the encoding region. A constant encoding writes the same record whatever the state; this is the mathematical meaning of the blinding control.

What it is. The self-prediction map could have been carried as free data sitting beside the structure. The development declines to, for a reason worth understanding.

The statement. If `predict` were a free field, then a *blinded* boundary—one whose self-encoding region is replaced by a constant—would retain the same predictive model and the same fixed point, since nothing would tie the map to the region. So the structure instead carries the region's *write* (`auto_resonance`) and its *read-out* (`readout`), and `predict` is their *composite* rather than a field. The self-model runs through the encoding region—the "avatar"—or it does not exist.

Five theorems fix the status of the resonance condition, which is a *predicate* (`IsRestrictionResonance`: the write is the presheaf restriction of the state to the avatar region) rather than a field:

- `eq_of_avatar_eq`: if the avatar regions of a family of resonant boundaries *cover* the substrate, two global states with the same reading at every avatar are equal. The readings are local; the global state is reconstructed from them by the same sheaf condition Part XI uses.

- `self_mem_range_readout`: a fixed point lies in the range of the read-out, so the Self is reconstructed from a single section over the avatar region. This proves reconstruction from a local section; it does not by itself establish the encoding's dimension or storage capacity.
- `self_eq_readout_restrict`: under resonance the Self satisfies $s = \text{readout}(s|_{\text{avatar}})$—the field is what its own localized self-encoding reconstructs. This is exactly the statement a structure carrying `predict` as free data could not make.
- `not_self_of_constResonance`: the negative half. The blinded boundary's map is constant, its only fixed point is named, and the original Self is *not* one. Note the asymmetry: the opposite statement—that blinding leaves the Self where it was—does not even typecheck, since a contraction hypothesis about one boundary's map says nothing about another's.
- `not_isRestrictionResonance_of_reconstructs`: what the region has to resolve. A boundary reconstructing two relevant states more than 2ε apart while its avatar region reads them identically is not resonant; read forwards, `restrictToAvatar_injOn_of_isRestrictionResonance` says a resonant boundary that reconstructs a pairwise separated family restricts injectively on it. Which states a region reads identically is fixed by where the region sits, before an encoding is chosen, so no read-out repairs it.

What it does not give. Three limitations are stated in the source and should be repeated here. (i) The read-out is a *modelling choice*: no field dynamics produces it, and the Lipschitz bound at the mean-field rate is an instance obligation. (ii) The witness substrate's coupling value is a choice of units, not a measurement. (iii) Nothing derives the resonance condition—that a physical avatar performs the restriction is an assumption on an instance, in the same class as the overlap-agreement obligation of Part XI. The criteria above refute it of a given region and construct a resonant avatar for none: a region that does resolve the relevant family is not thereby shown to carry one.

In Lean. `Phase6_ReflexiveTopology.lean`, with the three-site witness in `Examples` §10 exercising every clause—including `cortexBlind_self_ne` (the baseline is *not* a fixed point of the blinded map) and `cortexBlind_not_determined` (two distinct states with identical blinded readings).

37.1 Reconstructing more than one relevant state

What it is. A fixed point says the predictor gets one state exactly right. Sensitivity asks whether the same encoding and readout can also distinguish relevant changes. Choose a family of macrostates, a metric and an acceptable error ε, and require accurate reconstruction throughout that family.

The statement. If two states s, t receive the same code, their readout is the same state. The triangle inequality then gives

$$d(s,t) \le d(s, R(E(s))) + d(t, R(E(t))).$$

At least one error is half the separation or larger. Thus states more than 2ε apart need different codes if both are reconstructed to tolerance ε. A family of M such states needs at least M codes, and an alphabet of at most 2^b codes can serve only $M \le 2^b$.

This applies directly to the avatar: E is its encoding, R its readout, and $R \circ E$ the self-prediction map already used in the framework.

Read the code as the state's own restriction to a region and the criterion moves onto the region, where it constrains nothing by itself. Take the region to be the whole substrate: restriction is then the identity, so the boundary is resonant, reconstructs every declarable family exactly and restricts injectively on all of them, with nothing folded into any proper region (`topAvatar_satisfies_obligations`). `IsProperAvatar` is the condition this identifies, a predicate

rather than a field, and the two regions can be compared on one substrate: the proper one provably fails to reconstruct the family it cannot resolve while the improper one reconstructs it exactly, both being resonant (`properness_is_what_costs`).

The count also generalizes from an alphabet to a metric, which is the form a continuous family of states needs. With a metric on the codes and an L-Lipschitz readout, the triangle inequality that forced distinct codes forces *separated* ones: a family reconstructed within ε whose members lie pairwise more than $d_{\min}$ apart has δ-separated codes whenever $\delta L \le d_{\min} - 2\varepsilon$ (`isSeparated_image_encode`), so the family is no larger than the δ-packing number of the codes the encoder writes (`encard_le_packingNumber_range`), and under resonance of the region itself (`encard_le_packingNumber_restrictToAvatar`). At the whole substrate that bound is empty, the encoder's range being everything. The cortex witness reconstructs its two extreme states within tolerance one using a nonconstant, contracting map. Its blinded version fails at tolerance $1/2$ even though it has a unique fixed point. Another example needs three codes to distinguish three separated states.

One map cannot reconstruct much. If a single contraction by a factor $\Lambda < 1$ reconstructs every member of a family within ε, any two members lie within $2\varepsilon/(1-\Lambda)$ of each other, and at zero error the family is one state. A self that follows a changing scene applies a different map for each scene u, each with its own fixed point. If the map moves by at most L times the change in scene, two members lie within $(2\varepsilon + L\,d(u,v))/(1-\Lambda)$ and the fixed points within $L\,d(u,v)/(1-\Lambda)$, so the family can be as wide as the scenes are. That opens a shortcut: a readout that stores one state per scene and ignores the code reconstructs perfectly and encodes nothing. The code criterion therefore applies within one scene. A claim of *self*-reconstruction, as opposed to reconstruction of the scene, is a claim about several separated states per scene, and only there must the code tell them apart. A four-state example with two scenes and one bit shows all of this. No single contraction reconstructs it, each scene needs both codes, and no constant code works.

On its own the packing bound is a *ceiling*: it says what a code space cannot exceed, never that a code space supplies anything. So "a continuous medium has room for more states" stays a hand-wave until the other direction is available, and the other direction comes from measuring the code space. A maximal δ-separated family is also a δ-cover—that is what maximality means—so the region it packs is covered by that many balls of radius δ, and its measure is at most the number of them times the largest ball's. Read backwards that is a floor: a region of measure m whose δ-balls each measure at most v holds at least m/v mutually resolvable codes. In ordinary d-dimensional space a ball of radius δ has measure proportional to δ^d, so the floor is volume over the volume of one resolution cell.

With a ceiling and a floor on the same quantity the two kinds of code space can finally be compared at matched resource, because both are now counted in the same unit: how many states they can tell apart. An alphabet of 2^b codes distinguishes 2^b states. A region of volume m read at resolution δ distinguishes at least m/v. So the continuum wins exactly when its volume exceeds 2^b resolution cells, and loses when it does not—at a coarse enough resolution a region of any size holds a single code, which any alphabet matches. **What the comparison returns is a criterion and not a verdict.** What decides it is the volume, the resolution and the dimension, every one of them a measured quantity, and continuity decides nothing by itself: a continuous code space read at a *finite* resolution is a finite alphabet, and its capacity is a number of bits. Since every physical medium has a noise floor, that is the only resolution there is.

What it does not give. Accuracy across a family is an additional operational criterion, not a consequence of having a Self or UnifiedSelf. It asks for bounded error rather than multiple exact fixed points of one contraction. The relevant states and metric are declared; the theorem demands no recovery of every microscopic hardware state. The code alphabet is a bare type, so a physical interpretation of b as memory bits needs a hardware model. The packing form

declares its gain L and derives one for no physical readout, so a decoder of unbounded gain is constrained by neither form. The floor inherits that: it is a property of the code space and the resolution alone, so it does not say that an encoder writes those codes, that a readout separates them, or that the states they would encode exist. Ceiling and floor meet only where δ is a measured noise floor and L a measured gain, and neither is measured here—so the comparison is a criterion waiting on two numbers, not a result about hardware.

In Lean. `Phase6_Reconstruction.lean`: `dist_le_add_error`, `half_dist_le_max_error`, `Encoding.card_le_two_pow`, `ReflexiveBoundary.dist_le_prediction_residuals`, `Encoding.sub_two_mul_le_lipschitz_mul`, `isSeparated_image_encode`, `encard_le_packingNumber_range`, `encard_le_packingNumber_restrictToAvatar`; `measure_le_mul_packingNumber` and `measure_le_pow_mul_packingNumber` for the floor, and `two_pow_lt_packingNumber_of_lt_measure` with `Encoding.encard_le_two_pow_of_packingNumber_le` for the two directions of the comparison, both signs witnessed on the unit interval in `Examples/Phase6.lean`, §33. `topAvatar`, `topAvatar_satisfies_obligations`, `IsProperAvatar`. `Examples/Reconstruction.lean`, §24, checks the positive and blinded cases. `Phase6_ConditionedReconstruction.lean`: `dist_le_of_conditioned_contracting`, `dist_selfState_le`, `card_le_card_codes_fibre`, `dist_le_of_encode_const`; `Examples/ConditionedReconstruction.lean`, §36, is the two-scene example.

Part XIV. Coupling support, atomless measures, and hardware

Two features of a coupling representation bear on the substrate question: which pairs can interact, and whether components occupy points or regions under the measure used to evaluate them.

38 Resource-matched coupling reallocation

Imagine a wiring budget of R units to distribute among allowed pairs. For fixed phases, the score of pair (i, j) is $\cos(\theta_j - \theta_i)$: one for alignment, minus one for opposition. With nonnegative weights, the weighted sum cannot exceed the budget times the best allowed pair's score. If an unavailable pair has a higher score, allowing a wire there can improve the optimum at the same budget. This is the comparison made below.

Support specifies where a coupling may be nonzero. Resource matching means both alternatives receive the same total weight. Optimization means finding the best score among permitted choices; suboptimal means some permitted comparison achieves a strictly higher score.

What it is. *Rigidity* here means a constraint on *which pairs can be wired at all*—weights tunable, wiring fixed—not a constraint on weight values. The predicate `RealizableIn` $S\,R\,A$ says A is a valid coupling meeting resource budget R and vanishing on every pair the architecture has no wire for.

The statement. If phase correlation over the *wired* pairs is bounded by M, then no realizable coupling achieves total correlation above $R \cdot M$, and that bound is attained. Consequently, if some pair the wiring *misses* has correlation exceeding M, every realizable coupling is strictly suboptimal—and the quantitative gap is at least $R\,(\cos(\theta_l - \theta_k) - M) > 0$ at equal budget. Note that this *derives* the suboptimality hypothesis that the reallocation lemma would otherwise have to assume: it is a consequence of the wiring pattern, not an extra premise.

What it does not give. The result is resource-matched and not question-begging, but it turns on *fixed wiring support*, not on continuity. **An architecture wired to the best pair is not beaten.** It is a statement about constrained optimization, not about biology.

39 Point support and finite regions

Ordinary area gives a single point zero area, and any finite list of points still has zero area. A finite patch can have positive area. An *atomless* measure has this no-point-mass property in the spaces considered here; length and area are examples. Integrating a function nonzero only at finitely many points therefore gives zero. A different measure, such as counting measure, would see those points. The distinction is in what is being measured.

The statement. `fieldCorrelation` integrates a kernel against $\mu \otimes \mu$. A finitely-sited architecture is a finite set of points; on an *atomless* substrate such a kernel contributes exactly zero to the coupling energy, for every phase field and every choice of weights, while a patch-supported kernel gives $c\,\mu(U)^2$. So a continuum observable cannot register a finitely-sited architecture at all.

Finite components can instead occupy regions with positive mass. Partition the substrate into cells with masses $m_i > 0$ and spread a finite matrix entry A_{ij} over its cell pair at density $A_{ij}/(m_i m_j)$. The two masses cancel when integrating. With phases constant on each cell,

$$\iint K = \sum_{i,j} A_{ij}, \qquad \iint K(x,y)\cos(\theta_y - \theta_x) = \sum_{i,j} A_{ij}\cos(\theta_j - \theta_i).$$

The embedding preserves both total weight and phase correlation exactly. Two intervals of masses one and three witness twelve units of weight and twelve units of correlation. The point-supported construction gives zero because of its measure-zero support; having finitely many components is insufficient to force that result.

There is also a control for the wiring comparison. If the architecture wires both directions of a distinct, perfectly correlated pair, allocating all weight there attains the upper bound and no valid matrix of the same weight beats it. The formal maximum hypothesis includes diagonal pairs, whose correlation is one; merely being best among distinct pairs is not sufficient for that theorem as stated.

What it does not give. The point-supported comparison is **not resource-matched**—a measure-zero substrate carries zero resource in the $\mu \otimes \mu$ sense as well. What it establishes is a claim about the *model's observable*, not about computational power. A resource-matched comparison between the two kinds of substrate does exist, but it is about *capacity* rather than coupling energy, and it lives in Section 37.1: how many states a code space can tell apart, counted the same way on both sides. Its answer is a criterion decided by volume, resolution and dimension, and it favours continuity nowhere. And as Section 29 noted, it cuts against the framework's own coarse-graining step as much as against anything else.

These results establish no biological-versus-artificial distinction in computational power or consciousness. The cell embedding matches coupling weight, not joules; it provides no continuous kernel dynamics or refinement limit. The manuscript's identification is, as stated, multiply realizable: nothing in the sheaf or fixed-point conditions is specific to cortex or to electromagnetic coupling, so the substrate claim is carried entirely by E56 and E67 and is empirical.

In Lean. `Phase7_Rigidity.lean`: `totalCorrelation_le_of_realizable`, `exists_realizable_pair`, `rigid_is_strictly_suboptimal`, `rigid_gap`, `exists_better_coupling_allocation`; `fieldCorrelation_sited_eq_zero`, `fieldCorrelation_patchKernel`, `sited_architecture_below_field_optimum`. `Phase7_FiniteRegion.lean`: `kernelIntegral_cellKernel`, `fieldCorrelation_cellKernel`, `cellKernel_not_sitedOn`, `no_forced_gap_of_best_wired`; `Examples/FiniteRegion.lean`, §26.

Part XV. Formal method: why the development declares no axioms

This part contains no physics. It is here because the manuscript's fourth contribution is a claim about method, and that claim rests on one technical point.

40 What formalization is claimed to add

A formal proof spells out an argument in a language whose rules software can check. A *type* specifies what kind of object is allowed: a number, a function, a probability measure, or evidence for a proposition. Type-checking checks that every expression supplies the kind of object the next step needs. A proof *term* is the explicit object encoding a derivation. The checker can verify that a conclusion follows from assumptions; it cannot measure whether a brain satisfies them or decide whether the interpretation is apt.

For example, a proof about a density does not itself supply a time-dependent path. To use it in a theorem requiring a trajectory, one must construct that trajectory and prove the relationship, or state that relationship as an assumption. This is how a missing physical bridge becomes visible in code.

What it is. A prose table of assumptions would list the same eight edges. What the Lean development adds is that it makes the list *non-negotiable*.

The statement. The characteristic failure in an argument of this shape is not a false step but a *silent change of subject*:

- a theorem about a stationary density read as a claim about a trajectory;
- an inequality about a register read as a claim about learning;
- agreement of phases read as compatibility of representations.

Type-checking forbids each of those substitutions. An assumption needed to make a step go through appears as a hypothesis *in the term*, rather than as a word in a sentence. The last substitution is not merely forbidden but *refuted by a constructed object*—the incompatible-at-equal-phase cover of Section 31—and that object is what a prose table cannot carry.

What it does not give. Formalization does not make the assumptions true. It makes them impossible to lose.

41 Why a global axiom is the wrong shape for a physical postulate

An *axiom* is accepted without proof. A *class* or structure is a record containing named quantities and conditions; its entries are called *fields*, a programming term distinct from a physical field. An *instance* is one concrete record satisfying those conditions. Putting a heat law in the record says "this result concerns systems obeying this law". Declaring it separately for every record can instead assert it even for a record with heat values chosen to violate it.

A variable is bound when a statement specifies what it ranges over. An implicit argument is still an input to a theorem even when Lean fills it in automatically. *Inconsistent* means the assumptions yield a contradiction. In classical logic a contradiction permits any conclusion; a successful build alone is therefore not evidence that the physical assumptions were expressed consistently.

What it is. Formalization is only as trustworthy as its axioms, and an *inconsistent* axiom silently makes every theorem downstream of it vacuous while leaving the build green. The failure mode is specific and easy to reproduce.

The statement. The failure has one shape: **an axiom that constrains a symbol it does not itself bind**. Such a declaration is not a postulate about one system; it is a constraint on *every* model of the class the symbol belongs to, including models built to violate it—and a class with such a model is thereby inconsistent. Three concrete cases, each of which proves `False` on its own:

1. A symmetry-breaking statement taking the potential-energy functional as an *implicit* argument occurring only in its hypothesis. Instantiating that functional with the constant map discharges the hypothesis by `rfl` and yields the conclusion for an arbitrary field. With $V(v) = v^2$, $v_0 = 0$ and $\phi \equiv 1$ this proves $1 = 0$.

2. A Landauer heat statement equating a free class field with a bath-entropy difference *for every instance*. An instance with zero heat dissipation and a bath gaining one bit gives $0 = \log 2$.

3. A KL bound asserting $\sigma \geq D_{KL}(P\|Q)/\Delta t$ for *all* P, Q, while σ is a single real fixed by a class field. Since D_{KL} is unbounded above, a point mass against a uniform distribution refutes it.

Two further declarations of the same shape are recorded because their refutation is different in kind: as global assertions pinning free fields, a phase-periodicity condition and a synchronization-to-section equation are refutable as soon as the probability presheaf admits two distinct global sections. That refutation needs no arithmetic at all, only that the sheaf is non-trivial.

The rule that prevents all of this is uniform:

> *A physical postulate that mentions a class field must be a **field of that class**, never a standalone axiom quantified over all instances.*

A postulate is then an *obligation each model discharges*, not a global claim about every model. Under that rule the development declares no axioms at all: `#printaxioms` on any result reports only Lean's own foundational three—`propext`, `Classical.choice` and `Quot.sound`.

Those foundations support treating logically equivalent propositions as equal, choosing elements whose existence is known, and identifying equivalent representatives in quotient constructions, respectively. "No axioms" in this repository means no additional declared axioms beyond those foundations. The default Lean build audits every library declaration and rejects any other axiom, including the placeholder `sorryAx` introduced by an unfinished proof. A separate local check also catches unfinished anonymous examples.

42 Witnesses, and what inhabitation does and does not prove

"Every object satisfying conditions A has property B" is true if no object satisfies A. This is *vacuous truth*. A witness supplies an actual object satisfying A, showing the result has at least one case to which it applies. Inhabited means at least one such object exists. A toy witness tests mathematical consistency, not biological plausibility. A discrete topology allows every set of sites as an open region; its Borel σ-algebra is the collection of measurable sets generated by those regions, which here includes every subset. A σ-algebra is a collection closed under complements and countable unions, providing the events to which a measure assigns sizes or probabilities.

What it is. A conditional theorem about an *uninhabited* class is vacuously true and worthless. So every class carrying a physical postulate gets an explicit witness under `Examples/`: a one-bit erasure system with a genuine bath reproducing $\Delta Q = k_B T \ln 2$; a perfectly resonant system discharging the KL bound; a scalar field on a one-point spacetime discharging the action principles; and—the most demanding case—a thermodynamic cover over a three-site substrate with

the discrete topology and its Borel σ-algebra, covered by two overlapping open patches meeting at a single shared site.

What it does not give. What the witnesses establish is that the classes are inhabitable, so the theorems about them are not vacuous. What they do not establish is that a cortex is such an instance. In the cover above the phase field is already constant—every patch locked at $\theta = 0$—which is precisely what makes the equilibrium obligation dischargeable. Getting a physical system into that configuration is the content of the informal dynamical argument, not of the witness.

43 The passive and active chain theorems

Composition means connecting conclusions to the hypotheses of later results. If A implies B and B implies C, a proof of A gives C through those two steps. The repository's chain has more objects and supplied conditions, but this is the logical operation. *Joint satisfiability* asks whether all its assumptions can hold at once, rather than each in an unrelated example. A regression test preserves a demonstrated property when source changes; here counterexamples also guard against weakened hypotheses.

The statement. `Chain.lean` provides two branches. The theorem `chain` uses passive predictive memory, with the eight named hypotheses E12–E89. The theorem `chain_active` replaces E34 and E45 with E34Active and E45Active for a specified finite feedback process. Both take a `ThermodynamicCover` and a `ReflexiveBoundary` whose class fields specify the local data and encoding, and conclude `UnifiedSelf`: the particular section glued by that cover *is* the unique fixed point of self-prediction. The cover includes compatibility. The separate restriction-resonance predicate identifies encoding with restriction to the avatar region; neither chain theorem consumes that predicate. `chain_hypotheses_jointly_satisfiable` discharges all eight at once on a witness assembled from a double well, a register, a uniform-grid mesh and a three-site cortex, which remain distinct toy systems rather than one mechanism. `chain_active_hypotheses_jointly_satisfiable` uses the noisy actuator of Section 10 together with the register, changing mesh and specified field/cover. Its controller has the register's state type and its heat fits the register's budget. The active bound therefore admits a process whose signed passive information difference is negative.

The following requirements tie premises to specified objects:

- **E34** asks for a predictive structure carrying *the register's own* temperature and Landauer heat, with non-zero waste—not merely for some predictive structure.
- **E34Active** supplies positive path probabilities, local detailed balance at the register's temperature, and an upper heat allocation for the named controlled substep. The entropy bound is then derived. Landauer's lower bound on erasure heat does not supply this upper allocation or identify the eraser with the actuator.
- **E45Active** consumes the bound for that same process, heat observable, temperature and budget, and requires the specified coupling-energy sequence to converge to its named limit. Section 10.15 discharges this for a scalar constitutive law and regular meshes, retaining the gain and profile as supplied inputs; Section 10.16 discharges it again for a kernel built from declared response modes, retaining their profiles and prices.
- **E78** asks for a cover coupled at least at K whose equilibrium is *the limit of a relaxation on that coupling* (`ThermodynamicCover.IsReachedByRelaxation`)—not merely for some cover.

The active witness rejects a zero budget because its actual mean heat is positive. It also rejects an incompatible energy limit: a constant-zero sequence cannot converge to three, even though the actuator satisfies its thermodynamic bound. The positive mesh witness supplies convergence independently of that bound. This demonstrates why the connection to coarse-graining needs a modelling assumption. The actuated witness makes that assumption explicit

as a density law: its energy depends on the same step's joint entropy reduction. Its spatial convergence follows from mesh regularity, and the budget bounds its limit only through the supplied gain and profile. That construction supplies no kernel. The microscopic witness supplies one, from occupancies the same step changes, and pays the installation work; its convergence is again spatial, and its hardware is declared. The passive edge is discharged on that same hardware from a declared predictive system's own joint law, and two such systems differing by their whole dissipated work give the identical edge.

For priced hardware with nonnegative occupancies, Section 10.17 bounds that kernel's mean-field coupling by its installed energy. E56 and E67 together require $U > 2D/\kappa$ for this arrangement (`installedEnergy_required_by_chain`); neither edge follows from meeting the energy condition.

After either branch reaches n5, `chain_from_coarseGrains` supplies the common proof through E56–E89. The field, cover and fixed section are preserved through that passage. Introducing feedback changes the thermodynamic premise; it does not remove the supplied compatibility, contraction or fixed-state identification.

What it does not give. The joint witnesses prove the hypotheses *simultaneously consistent*. They do not prove that one mechanism realizes them, and they do not establish the empirical truth of any premise. E12 remains equivalent to its own conclusion, which is why n1 and n2 are described as independent roots rather than as a derivation.

The active chain theorem concerns an autonomous substep with the controller fixed, finite states and strictly positive probabilities. The complete finite update theorem in Section 10.1 matches the intermediate law and accounts for both substeps at a common thermal scale. Its heat bound does not supply E34Active's register-to-actuator allocation, and no heat bound supplies the hardware or the partitions that E45Active's spatial convergence rests on. A claim about learning requires specified update dynamics and a criterion for improvement, as in the policy-register model of Section 10.4. That model's convergence is over learning updates; n5 concerns refinement of a spatial mesh. Convergence over learning time alone does not establish that limit.

43.1 A chosen physical model and its empirical interpretation

A physical model specifies variables with proposed physical meanings and rules for their evolution. For the two-bit model, these are the two states, their interaction energy, the thermal transition probabilities and the reservoir. Choosing those ingredients does not require first measuring a brain. One can prove a conditional entropy balance, construct a counterexample or derive a convergence theorem entirely from explicitly chosen mathematical assumptions.

An empirical interpretation is a further claim: that particular neural quantities obey those assumptions closely enough for the result to explain an observation. That claim needs evidence. Lean can check whether a specified dynamics implies a stated property; it cannot establish which dynamics cortex implements. Progress in the formal model and validation of its biological interpretation are therefore distinct tasks.

Part XVI. Circular statistics and the empirical test

44 Doing statistics on angles

An electrode channel records a voltage difference over time. The analysis extracts an angle from each selected oscillatory signal and then asks how those angles are distributed. A *sample*

is one measured value; a distribution is the probability pattern those values are used to estimate. An *estimator* is the rule turning samples into an estimate of a quantity. The *estimand* is the underlying quantity it aims to measure. Changing which sites or times are pooled can change the estimand even if the formula stays the same.

What it is. Phases live on a circle, where ordinary statistics break. The mean of 1° and 359° is 180° if you average the numbers and 0° if you average the directions; only the second is right. Circular statistics works with the unit vectors $e^{i\theta_j}$ instead:

Resultant length r.
$r = \left|\frac{1}{N}\sum_j e^{i\theta_j}\right|$. Zero for a uniform spread, one for perfect alignment. This is the order parameter.

Circular mean ψ.
The direction of that same resultant vector.

Concentration a.
The von Mises parameter, governing how tightly the distribution is peaked.

Why the framework needs it. Because the manuscript's protocol turns on estimating a and r *independently*, and that is the constraint most easily violated in practice: linear statistics on wrapped phases estimate neither quantity.

45 Getting the phases: the analytic signal

For a pure cosine $x(t) = \cos(\omega t)$, adding the quarter-cycle partner $\sin(\omega t)$ gives the arrow $z(t) = \cos(\omega t) + \mathrm{i}\sin(\omega t)$. Its direction is phase and its length is amplitude. The Hilbert transform constructs a corresponding quadrature partner for a more general signal. A *narrowband* signal concentrates its oscillatory activity in a limited frequency band, making a single phase easier to interpret. A mixture of unrelated rhythms can still yield a numerical angle, but not necessarily one meaningful oscillator phase.

Detrending removes a fitted slow baseline. A bandpass filter retains a chosen frequency range and suppresses others. Filter order controls its frequency response; a Butterworth filter is designed for a smooth, flat response in its passing band. Second-order sections implement it as a series of small filtering operations for numerical reliability. Filtering forward and backward removes net phase delay, but uses future as well as past samples and can distort short transients. Padding supplies extra samples beyond the interval of interest to reduce distortions at its edges.

EEG means electroencephalography, electrical recordings at the scalp. EOG records eye-related electrical activity and EMG muscle activity. Hertz (Hz) means cycles per second; milliseconds (ms) are thousandths of a second.

What it is. The *Hilbert transform* turns a real oscillatory signal $x(t)$ into a complex *analytic signal* $z(t) = x(t) + i\,\mathcal{H}[x](t) = A(t)e^{i\phi(t)}$, from which an instantaneous amplitude and an instantaneous phase can be read. It is meaningful only for reasonably narrowband signals, which is why the pipeline bandpass-filters first.

The statement. The implemented pipeline: detrend each channel; bandpass with a 4th-order Butterworth filter in the 4–40 Hz band using second-order sections (`sosfiltfilt`, which is zero-phase); take the Hilbert transform for $\theta_j(t)$; use only the 62 scalp EEG channels, discarding EOG and EMG. Three seconds of padding on each side is downloaded and trimmed after filtering to suppress edge artifacts.

46 The parameter-free prediction, and the trap in testing it

Suppose a curve claims one measurement predicts another. A meaningful check must calculate both quantities without forcing either to obey that curve. Here r comes from averaging phase arrows, while a is estimated from the shape of their histogram. "Independent" means distinct estimating equations; it does not mean the estimates, made from the same data, are statistically independent.

Maximum likelihood chooses model parameters that make the observed sample most probable. A *sufficient statistic* retains all the sample information that this fitting calculation uses about a parameter. A score equation sets the derivative of the log likelihood to zero to find a candidate optimum. In the centered von Mises fit that equation already enforces the very concentration–coherence relation being tested.

What it is. The stationary model predicts that concentration and coherence lie on one curve:

$$r \;=\; \frac{I_1(a)}{I_0(a)},$$

with *no fitted shape parameter*—the same curve for every subject and every state.

The statement. The estimator must not be a maximum-likelihood one, and this is the sharpest methodological point in the empirical section. The von Mises family is a one-parameter exponential family whose sufficient statistic is $\sum_j \cos\theta_j$—that is, r itself. So *any* ML estimate of the concentration solves its own score equation $I_1(\hat{a})/I_0(\hat{a}) = r$ and returns $\hat{a} = A^{-1}(r)$ identically. The pair $(\hat{a}, r)$ would then lie on the predicted curve *for any input whatsoever*, including phases drawn from a distribution that is not von Mises, and the collapse would be an algebraic identity rather than a test.

The implemented estimator therefore uses the log-density *shape*: phases are counted in intervals called bins. A pseudocount adds a small positive count so an empty bin does not require taking $\log 0$. Regression fits a line; ordinary least squares chooses its slope and intercept to minimize the sum of squared vertical errors. Unweighted means each bin contributes equally. Specifically, phases are centred on their circular mean, binned into 40 bins on $[-\pi, \pi]$, and the log of the binned density (with a 0.5 pseudocount) is regressed on $\cos\theta$ by *unweighted* ordinary least squares. Under the von Mises model $\log\rho(\theta) = a\cos\theta - \log(2\pi I_0(a))$, so the slope is a. Other weighting rules are different estimators and are not assumed equivalent to maximum likelihood.

The final residual $r - R(\hat{a})$ compares measured coherence with the curve's prediction using the separately estimated shape; the hat denotes an estimate. The artificial controls use other shapes. A wrapped Cauchy wraps a Cauchy distribution, which has heavier tails than a Gaussian, from the line onto the circle. A top-hat arc has equal density on a chosen arc and zero elsewhere; a mixture draws from one of two distributions with specified proportions. These controls ask whether the procedure can reject other density shapes.

Discriminating power was verified directly. At the analysis's sample size, a wrapped Cauchy ($\rho = 0.6$), a top-hat arc and a 50/50 von Mises–uniform mixture give residuals $r - I_1(\hat{a})/I_0(\hat{a})$ of $+0.088$, -0.117 and $+0.075$, against a worst case of 0.0024 on genuine von Mises samples—a separation of 31.7×. A uniform distribution passes, correctly, being von Mises at $a = 0$. The pseudocount biases the estimator by ≤ 0.007 for $a \leq 3$, reaching -0.375 only at $a = 5$ where the troughs empty out.

What it does not give. Even a perfect result here identifies neither ephaptic coupling nor this framework: the relation is common to the whole von Mises family, so agreement with it does not distinguish this framework from alternative synchronization models. Worse, within the same ansatz $a/r = K/D$ for $r > 0$, so the limiting threshold ratio *follows from* the density relation instead of testing it—and ratios near the finite-size floor are especially unstable.

47 What the EEG analysis found

Propofol is an anesthetic drug; these are recordings during drug-induced sedation, not ordinary awakening. OpenNeuro is the repository hosting the dataset. A *montage* specifies which voltage differences become channels. A common-reference montage subtracts one shared reference from each electrode; a bipolar montage takes differences between electrode pairs. Shared reference signals can therefore enter many channels at once.

SEM, the standard error of a mean, estimates uncertainty in that mean under the stated sampling assumptions. It differs from the spread of individual observations. RMSE is root-mean-square error and reports a typical residual size. Theta, alpha, beta and gamma name conventional EEG frequency bands; these names concern different ranges of rhythm speed. "Broadband" combines the whole selected range. A bootstrap repeatedly resamples observations to estimate sampling uncertainty; percentile intervals use quantiles (cutoffs at specified fractions) of those estimates. Nearby time bins can be dependent, so treating them as independent in resampling can misstate uncertainty.

The statement. On public propofol-sedation EEG (OpenNeuro ds005620, Bajwa et al., 2025), 21 subjects at 65 channels and 5 kHz, the pipeline forms 61 adjacent bipolar pairs from the 62 scalp channels, band-passes and Hilbert-transforms them, and pools phases over all pairs and non-overlapping 100 ms bins.

- **Cross-subject (8 subjects), bipolar montage.** Pooled a spans 0.087–0.142 (mean 0.104), pooled r averages 0.052, and the mean residual from I_1/I_0 is -0.0002 (RMSE 0.0027).
- **Narrowband.** Theta, alpha, beta, gamma and broadband mean residuals are -0.0018, +0.0001, +0.0001, +0.0000 and +0.0000.
- **The same pipeline on signals of known coherence.** A perfectly phase-locked 10 Hz rotation, whose instantaneous resultant is one, comes out at $r = 0.0001$ and $a = 0.0004$, squarely on the curve. A field with raw-channel resultant 0.89 has bipolar resultant 0.06 and pooled resultant 0.007.

The third item decides how to read the first two. Within 100 ms the common phase of an alpha rhythm turns through a whole cycle, and pooling reads that turn as dispersion; subtracting adjacent electrodes adds a second loss, because the phase of a difference of two offset rhythms depends on their amplitudes as much as on the shared phase. Any nearly uniform pooled distribution has $r \approx a/2$, and the Bessel ratio is $a/2$ to within 0.0002 over the observed range. Sitting on the curve here is therefore what a perfectly synchronized signal would also do.

What it does not give. Four limits, and the first is the one that matters most.

1. **The estimand is not what the protocol asks for.** Pooling over pairs and time without removing the common rotation measures neither the spatial concentration nor the coherence of the model; centring each sample on its instantaneous mean phase would. The bipolar montage changes spatial phase too, so it needs its own observation model.
2. **Referencing is a material analysis choice**, not a preprocessing detail.
3. **The range cannot discriminate the curve.** Separating I_1/I_0 from a straight line needs concentrations near $a \gtrsim 1.5$, where the two differ by 0.154, and an estimator calibrated at the actual spatial sample count.
4. **Bootstrap intervals are exploratory.** Resampling time bins does not preserve dependence between neighbouring bins.

And the purpose was never to test the awakening hypothesis, for which these are the wrong data: they concern propofol sedation, not recovery from sleep. The exercise shows what a pooled bipolar estimand cannot measure.

48 The awakening experiment

A *protocol* fixes how measurements and decisions will be made before the relevant outcomes are examined. A proxy is an indirect measurement of a quantity; calibration establishes its relationship to that quantity in a specified setting. A behavioral endpoint is a separately defined outcome, such as performance on a task. Predeclaring it prevents redefining recovery after seeing the coherence curve.

Diffusion MRI is magnetic resonance imaging sensitive to water movement; it is not a direct measure of the model's phase diffusion D. Astrocytes are supporting brain cells whose signals might supply other indirect measurements of extracellular conditions. These proxies require their own validation. Unblinding means revealing outcomes or group labels hidden during analysis design. Observation noise is measurement error, distinct from random motion in the underlying model.

A sigmoid is an S-shaped rising curve; an exponential recovery has a fixed characteristic rate of approach. They are alternative shapes to compare under the same filtering and measurement process. *Hysteresis* means the response depends on the path by which a control value was reached, such as different coherence on increasing and decreasing coupling. A stationary fold is a branch turning back in the control parameter; it can permit multiple stationary states. Dynamical lag can produce a loop even without such a fold.

The statement. The manuscript specifies six requirements, and the design—not any single measurement—is where the content of the hypothesis lives.

1. **Coupling and diffusion in the same units.** K and D estimated from the same cortical regime, in inverse-time units, separating phase diffusion from fixed natural-frequency differences and from observation noise. Without this, "$K > 2D$" is not an evaluable statement about cortex. Watching a sequence of equilibrated phase distributions will not do it: the stationary density is von Mises with concentration Kr/D, so it fixes the *ratio* and nothing else, and a coupling increase, a diffusion decrease and a common drive of matched strength are then identical window for window. The estimate has to reach the phase increments, which still carry the rate, and telling an endogenous coupling from a common drive needs a measured drive or a controlled perturbation on top of that.
2. **Geometry calibrated to coupling.** A diffusion-MRI observable or astrocytic signal indicates the extracellular volume fraction α; its mapping to effective coupling must be established separately, with its own uncertainty. The conductivity argument below supplies the expected sign of that mapping and nothing more.
3. **Calibrated spatial resolution.** Estimate r by resultant length and a by unweighted log-density regression. At 100 sites the tested 40-bin estimator fails the bias-floor criterion at every scanned concentration. At the declared design concentration $a = 0.542$, the scanned designs require 3000 independent sites, rising to 31000 under the declared clustering model. These counts are specific to the estimator and dependence model; the recording design needs its own calibration. Pooling over time changes the spatial estimand.
4. **An independent recovery endpoint.** Behavioural recovery measured separately from the phase trace, so latency is not defined by the signal whose latency is at issue.
5. **A predeclared dynamical onset criterion**, fixed above the finite-size floor before unblinding and evaluated under the measured coupling trajectory, including ramp speed or fluctuations.
6. **Model comparison under matched observation.** Against exponential, flexible sigmoidal and other slow-recovery models under identical temporal filtering and observation noise. Latency alone cannot distinguish the mechanism, because many processes create delays.

The statement. The design fails informatively in three ways: failure of the stationary phase-

density relation over a sufficiently broad concentration range; absence of the proposed coupling–geometry association; or a recovery trace better explained by an alternative after finite-rate and observation effects are accounted for.

What it does not give. Two apparent failures are not failures. A linear apparent onset is evidence against the adiabatic approximation only where that approximation is independently justified—see Section 25—and does not falsify the field hypothesis under arbitrary ramp speeds. And the absence of a fold in the stationary equation excludes stationary coexistence of two positive branches in this model, but does not exclude finite-rate hysteresis or adjudicate theories of ignition.

Part XVII. Dimensional analysis and the calibration problem

49 Why the units do not match

Dimensional analysis checks whether quantities have compatible physical units. Distance divided by time is speed; distance alone cannot be compared with a speed threshold. Likewise, a phase shift and a phase-change rate are different quantities. A Fermi estimate is a rough estimate assembled from plausible factors to establish an approximate scale. Correct arithmetic cannot repair incompatible units.

What it is. The manuscript calls this "the sharpest open limitation of the framework", and it is a pure dimensional-analysis problem—the kind that is embarrassing precisely because it cannot be argued away.

The statement. A natural Fermi estimate of field-mediated coupling multiplies a population size N, a spike-time shift per unit field $\mathcal{T}$, a field amplitude E, and an oscillation frequency f. Track the units:

- s is a time per unit field, so sE is a *time*;
- multiplying by a frequency f gives a *pure number*—the fractional phase advance a cell receives per cycle from the population field;
- multiplying by a population size N leaves it a pure number.

But a coupling in the sense of the stochastic model (1) is an *inverse time*, comparable with D. So the estimate cannot be compared with D at all. The conversion

$$K_{\text{eff}} = \gamma\, N \mathcal{T} E f$$

puts the entire gap into a rate calibration γ.

50 Why γ is not recoverable

A unit conversion changes how the same quantity is expressed: seconds into milliseconds, for example. Calibration determines an unknown physical relationship. A single timing shift does not tell us whether it persists, decays, or accumulates with later shifts. The rate γ encodes that missing relationship, so choosing a time unit cannot measure it.

The statement. The substantive point is that γ is a *hypothesis about persistence*, not a unit conversion to be fixed by convention. It states how a per-cycle phase advance accumulates into a drift rate, which depends on how long a field-driven timing shift survives. Three candidates are available and they are not interchangeable:

- $\gamma = f$: each cycle's advance is retained in full and independently;
- $\gamma = 1/\tau$ for a phase relaxation time τ: it decays as the medium relaxes;
- $\gamma = D$: it accumulates at the rate phase itself diffuses.

These need not agree, in cortex or anywhere else, and a field measurement chooses between them only if the persistence of induced timing shifts is measured alongside it.

The third must be declined for an additional reason, and the reason is instructive. Under $\gamma = D$ the threshold test $K_{\text{eff}}/D > 2$ reduces to $NsEf > 2$—a comparison of the dimensionless field estimate against the number 2, *with the diffusion parameter cancelled out of a test whose entire content was the ratio of coupling to diffusion.* The economy of that substitution is what makes it unsafe: it converts an unmeasured rate into an apparent confirmation while removing the quantity the confirmation was about. The same objection applies to any γ recovered by requiring the model to cross its own threshold, or read off the Bessel collapse used to test the stationary density.

What it does not give. Every saved field-strength value in the repository is therefore *conditional* on a calibration convention. Under the numerical convention $\gamma = 1$ in the simulation's inverse-time unit, the Fermi arithmetic gives required field ranges that fall *below* the 1–5 mV/mm comparison band, and they scale inversely with γ. The simulation does not measure the calibration, so it cannot independently validate physical-unit consistency or a cortical field threshold. Identification requires coupling and diffusion estimated in one cortical regime by a route *independent* of the field arithmetic—which is item 1 of the protocol above.

51 The empirical case for the field, and its critics

Endogenous means generated within the system. Transmembrane current crosses a cell membrane; extracellular current flows through the spaces outside cells. *Ephaptic coupling* is an influence of the surrounding electric field on a cell's membrane voltage and timing, without requiring a direct synaptic connection. Entrainment means an influence brings a rhythm's timing into a sustained relationship with it. Modulation changes ongoing activity without necessarily initiating it.

The cerebral cortex is the outer neural sheet of the brain. An action potential, or spike, is a brief electrical pulse of a neuron. Synapses are specialized junctions through which cells influence one another. Gap junctions are channels directly connecting cells; a syncytium here means a connected cellular network functioning collectively. Representational drift means the neural activity associated with a measured content changes over time. Superposition means adding contributions to a field.

Conductivity measures how readily current flows for a given electric field; its unit is siemens per metre (S/m). A field in volts per metre (V/m) is a voltage change per distance, and $1\,\text{mV/mm} = 1\,\text{V/m}$. Resistive conduction means current approximately follows the applied field without a substantial frequency-dependent phase offset in the stated range. An effective-medium model replaces microscopic detail by an averaged material description. Volume fraction is the proportion of tissue volume occupied by a component; diffusion tortuosity measures how tissue obstacles hinder water or solute diffusion compared with free medium. Sleep inertia is the temporary impairment after waking.

The statement. The identification of the continuum kernel with cortex's endogenous electromagnetic field is supported as *modulatory and integrative*, not as primary. Transmembrane currents generate extracellular fields that shift neighbouring membrane potentials—ephaptic coupling. Endogenous fields of 1–5 mV/mm entrain spike timing (Fröhlich & McCormick, 2010; Anastassiou & Koch, 2015); stable electric fields carry information that survives representational drift in the underlying spiking population (Pinotsis & Miller, 2022, 2023a,b); the field

is spatially continuous and superposes contributions from large populations, which is the integrative role CEMI field theory assigns it (McFadden, 2002, 2020; Pockett, 2012), physically supported by astrocytic syncytia joined by gap junctions (Verkhratsky & Nedergaard, 2018). Sleep- associated changes in extracellular space and clearance (Xie et al., 2013) motivate the possible slow control of coupling.

What it does not give. The formal predicate that E56 carries is explicit and modest: a jointly continuous kernel, a normalized substrate, a coupling constant equal to that kernel's double average, no single site contributing more than half of it, and a noise D that is the field's own. **No object in the development denotes cortex**, so whether cortex satisfies these is not a Lean question. Toy witnesses satisfy the conditions; none is a model calibrated to cortex.

The longest-standing objection is that the fields are too weak to matter, pressed from *within* the field-theoretic camp by Pockett (2012, 2002). Vöröslakos et al. (2018) found that applied current is shunted through scalp, delivering fields at or below 1 V/m in cortex. That bounds *applied* rather than endogenous fields, but it constrains how much causal work a field of that magnitude can do—and is the reason the framework does not claim the field drives spiking. Changes in phase organization under propofol (Bastos et al., 2021) likewise do not measure a crossing of the model's coupling threshold.

Extracellular geometry reaches the coupling through the field amplitude and not through a further mechanism. The extracellular space occupies a volume fraction $\alpha \approx 0.2$ of cortical tissue at a diffusion tortuosity $\lambda \approx 1.6$ (Syková & Nicholson, 2008), and natural sleep raises α from roughly 0.14 to 0.23 (Xie et al., 2013). Cortical tissue conducts approximately resistively below 10 kHz, at 0.3–0.6 S/m, an order of magnitude set by the conductivity of saline diluted into that volume fraction (Barbour, 2017). At a fixed transmembrane current density, then, a larger α means a smaller extracellular field: the space contracts on awakening and K should rise. That is an effective-medium argument, so it fixes a sign and not a magnitude. How a heterogeneous field aggregates into a single K is answered separately, by the row-sum rule the manuscript's spatial control measures: over decay lengths from 0.1 to 0.3 mm the reduced threshold sits within 0.003 of what a uniform control returns on the same sheet, and what displaces it is frequency heterogeneity rather than spatial reach. Neither the monotonic relation between extracellular geometry and coupling, nor its role in human sleep inertia, follows from the cited physiology; the model requires both to be measured.

The identification would be challenged if independently calibrated field-mediated coupling were insufficient for the proposed regime, or if a selective perturbation of that interaction left the proposed neural and behavioural endpoints unchanged. Such a perturbation needs controls for direct effects on transmembrane currents and other coupling pathways; selectivity is not assumed experimentally available.

It is worth being precise about what such a failure would take down, because most of it is not electromagnetic. Of the six requirements the manuscript's protocol lists, only the second—the calibrated relation between extracellular geometry and effective coupling—discriminates an electromagnetic field from any other continuous coupling. It is the only one that runs through conductivity and volume fraction, which is the single place a field equation enters this work. The rest ask for coupling measured against diffusion, a calibrated phase measurement, an independent recovery endpoint, a predeclared onset criterion and a matched model comparison, and extracellular ion diffusion, gap junctions treated as a density, astrocytic calcium or a mechanical medium would satisfy or fail them in exactly the same way. So a null result on the second requirement counts against the electromagnetic identification, and a null result on any of the others counts against the coupling-gated account whatever turns out to carry the coupling.

Part XVIII. What the mathematics is asked to mean

The composition terminates in a mathematical object, and the manuscript takes a position on what that object is. This part states the position, its motivations, the choices of cover and self-encoding it requires, and the measurements needed to test it.

52 The commitment

A *formal correlate* is a mathematical feature proposed to correspond to an aspect of experience. The correspondence requires an interpretation; a theorem about the feature alone cannot establish it. A conscious episode is a period considered as one experienced situation. *Unity* concerns the experience being one whole, and a *minimal self* the perceiver's place or perspective within that situation, without requiring a life story or an explicit verbal judgment about oneself.

The statement. Two clauses, neither a theorem:

Unity.
The unique global section glued from compatible local states is the formal correlate of the unity of a conscious episode—the fact that a subject undergoes *one* experience rather than a collection of co-occurring local ones.

Selfhood.
Accurate internal reconstruction of that same section is the proposed formal correlate of a minimal self. Assigning it this meaning requires evidence that the encoding represents the system's own condition.

53 Motivations and the choices they require

A representation carries information interpreted as being about something, its *content*. A decoder is a specified rule inferring that content from measured activity; it does not follow automatically from coherence. Reflexivity means the representation concerns the representing system itself. An idealization deliberately makes a feature exact, such as reconstruction without error, to define a tractable model. Testing it requires deciding how much approximation is acceptable in real measurements.

Compatibility gives a precise way for distinct local descriptions to contribute to one global state. Only shared domains must agree. A content model must therefore distinguish inconsistent descriptions from consistent descriptions of an uncertain situation or a conflict between judgments. Spatial mass profiles provide the formal example; extending it to decoded perceptual contents requires a suitable model of global extension.

The manuscript illustrates this with reaching for a cup. A description of the cup's appearance and one of the hand's relation to it can agree on cup location after conversion to common coordinates, while contributing different details. This is a modelling example, not an implemented neural decoder or an instance of the spatial-measure theorem. A content model must establish its own extension property.

The reflexive map adds an internal representation of that same state. Under restriction-resonance, it restricts the global state to an encoding region and reads the encoding back out. A fixed point is accurately reconstructed at the level of its modelled variables. Approximate reconstruction has explicit residual and code-separation bounds. Choosing a fidelity criterion appropriate to changing neural states remains an empirical question.

The self interpretation concerns the perceiver's place in the represented situation. Reconstructing a scene alone would not demonstrate a perspective. A candidate encoding of hand or viewpoint should have a measurable role in how the system represents its own relation to the cup. Interventions on that encoding should produce changes predicted by its proposed readout. This functional criterion helps assign meaning to the model's variables; the fixed-point theorem does not establish it or identify it with experience.

Both constructions require independent choices. Local descriptions must carry content attributable to their domains, and the encoding and readout must have a demonstrated relation to the global state across a specified range of states. Contraction and a fixed point alone would also hold for a constant readout returning a predetermined state. Their psychological significance depends on the proposed encoding relation.

The full identification proposes that unity and reflexivity occur together in conscious episodes. Gluing itself supplies no necessity for reflexivity. Unified experience without the specified self-encoding would challenge that requirement while leaving the compatibility clause assessable separately.

54 The costs

The *hard problem* is the question why a physical process should have any felt quality at all. *Multiple realizability* means the same abstract conditions could be implemented by different physical systems. Both terms concern the interpretation's reach, rather than an algebraic step in a proof.

What it does not give. The manuscript states these rather than leaving them to a referee.

- **It does not address the hard problem.** The commitment is a proposal about *which* physical state is the conscious one, not about why states of that kind feel like anything. It is therefore open to the objection that a system could satisfy every condition—compatible local states, a supercritical field, an accurate self-model—while nothing is experienced. *No reply to that objection is offered or claimed.*

- **It is multiply realizable as stated.** Nothing in the sheaf or fixed-point conditions is specific to cortex or to electromagnetic coupling, so the substrate claim rests entirely on the two empirical edges E56 and E67. No conclusion about artificial systems follows from the formal results.

- **It is independent of the physical tests.** The predictions of Parts VIII, IX and XVI concern coupling, coherence and recovery. They can fail while the identification stands, or succeed while it is wrong.

55 What would count against it, and the observable that is missing

An observable is a quantity that a specified measurement can estimate. *Sufficiency* means meeting the conditions guarantees the outcome; *necessity* means the outcome cannot occur without them. A dissociation test seeks cases where proposed linked properties come apart, such as coherent timing with incompatible decoded contents. Separately measuring each property is essential to such a test.

The statement. A full test must evaluate coupling, compatibility and the proposed self-encoding separately, alongside an independent experience endpoint. A state satisfying all the conditions with evidence against a unified episode would challenge sufficiency. Unified experience with incompatible local descriptions would challenge the unity clause; unified experience

without the specified self-encoding would challenge the reflexive requirement. Absence of report alone cannot establish absence of experience.

The compatibility observable is specified and exercised on constructed data, where it separates built-in incompatibility from built-in compatibility at AUC = 0.986 and its two silent failures are caught by per-patch diagnostics rather than by the statistic itself. It has no neural instantiation, and the neural self-encoding has yet to be identified. The supplement details what a compatibility measurement needs and what the implementation separates; the manuscript explains how that measurement would enter a dissociation test.

55.1 Three components

A *null model* specifies what a measurement would look like if the tested effect were absent. Here the null is compatibility with noisy decoders, so even agreeing underlying contents can give a positive estimated distance. Validation means checking the method on cases with known answers, including both matching and mismatching local descriptions.

A restriction map.
Two regions with overlapping receptive territories each carry a distribution over what occupies the shared part, decoded from that region's activity alone. Compatibility is a relation between the two *restrictions*, so the map taking a region's decoded distribution to a distribution over the overlap must be specified before any statistic is computed. For the spatial measures in the formalization, restriction retains the masses at shared sites. For a decoded distribution over content variables, marginalization instead sums over unwanted variables. The content model must specify which operation its shared domain requires.

A statistic with a null.
A distance between the two restricted distributions is not yet a measurement, because decoding error alone makes that distance positive. The statistic needs the distribution of the distance *under compatibility*—what two decoders of the same underlying content produce given their own noise—so that incompatibility is a departure from a stated null rather than a non-zero number. Overlapping territories supply dependent samples, and the uncertainty must be computed accordingly.

Validation where the answer is built rather than inferred.
The estimator must be exercised on data in which compatible and incompatible overlaps are constructed by design. The formalization supplies the configuration to target: local states at a single common phase whose densities disagree where the patches meet (Section 31). An estimator that cannot separate that configuration from its compatible counterpart at realistic decoding quality is not yet a measurement of the property the framework needs.

55.2 Three ways it fails silently

A prior is a probability distribution supplied before the current evidence; a posterior is the distribution after incorporating that evidence. A decoder can bias results by pulling different inputs toward the same prior. Bias is a systematic displacement of estimates, not just random variation.

1. A decoder that shrinks both regional posteriors toward a shared prior reports *compatibility* whatever the regions encode—having replaced two local descriptions with one global one.

2. A decoder carrying independent per-region bias reports *incompatibility* whatever the regions encode.

3. **Circularity.** A compatibility statistic computed from phases measures phase agreement—and the distinction between phase agreement and overlap compatibility is the entire content of E78. An observable that read compatibility off synchrony would confirm the framework's central distinction by assuming its negation.

The first two are opposite failures of the same component, neither visible in the statistic itself, which is why the constructed-answer validation is a requirement and not a refinement.

What it does not give. Such an observable would *not* establish E78, which asserts that coherence drives a cover the theorem is given to its relaxed equilibrium, nor the cover's own compatibility field. It would make the consequent measurable and leave the implication a hypothesis. It bears on neither the recovery prediction nor the identification of the glued section with experience. What it supplies is the missing term in the dissociation test. A null outcome—no reachable data separating compatible from incompatible overlaps at achievable decoding quality—is itself informative, and would place the interpretive commitment outside experimental reach rather than merely outside the article.

56 Relation to other accounts

Integrated information theories associate consciousness with irreducible cause-and-effect organization of a system. Global workspace accounts emphasize information becoming broadly available to other processes, often called broadcast. Predictive processing describes perception and action through models that predict inputs and respond to discrepancies. A generative model specifies how hidden conditions could produce observations. These are families of accounts, so the comparison concerns these broad emphases.

The exclusion principle in integrated information theory selects a definite experience and system among overlapping candidates. Ignition in workspace theories refers to a marked transition to broad availability. The combination problem asks how purportedly separate experiential constituents could form one experience. None of those questions is settled simply by making a physical field continuous.

The statement. The proposal differs in emphasis from integrated information (Tononi et al., 2016), global workspace (Baars, 1997; Dehaene et al., 2014), and predictive processing (Friston, 2010). It specifies a candidate interaction medium, a mathematical compatibility condition and a local self-encoding. Uniqueness follows within a chosen cover and contracting map; the empirical choice of those objects and their relation to experience remain open.

What it does not give. It does not derive the exclusion principle of integrated information, identify every broadcast with a field transition, or derive a generative model from a dissipation bound (Section 11). And continuity alone does not settle the combination problem: continuous fields can carry incompatible local descriptions, and discrete models can admit sheaf constructions. Compatibility is what this account offers in place of continuity. These mathematical differences do not settle which systems have conscious experience.

Appendices

A Master table: concept, role, and status

Read the status column strictly. "Theorem" means proved under stated hypotheses; it does not mean physically established. "Instance postulate" means a physical assumption carried as an obligation each model discharges, with an explicit witness showing the class is inhabitable.

Concept	Role in the argument	Status
Spontaneous symmetry breaking (§3)	Degenerate minima make a choice necessary	Theorem (holds a.e.; the a.e. is necessary)
Topological defects, π_0 (§4)	Forces departure from separated minima	Theorem; identifying an agent's boundary needs further assumptions
Finite capacity	Bounded register cannot absorb its history	Theorem; finiteness assumed and fenced
Landauer's principle (§6)	Erasure costs heat	Theorem + instance postulate ($\Delta Q = T\Delta S_{\text{bath}}$)
Norton–Shenker distinction	Logical $\neq$ physical erasure	Built into the definitions; survives as a theorem
Data processing inequality	Splits memory into predictive and not	Theorem; no-back-action is load-bearing
Still et al. bound	Nonpredictive memory is costly	Instance postulate; consequences derived
Landauer $\rightarrow$ Still bridge	One heat, two vocabularies	Theorem under an identification (= E34)
Agency and feedback (§9)	Explicit perception–action channels; separates two conditional information terms	Theorem; finite information, passive case recovered
Finite feedback thermodynamics (§10, §10.1)	Elementary and complete two-substep heat budgets; total first law	Theorem under positivity and local detailed balance; common thermal scale and energy for the cycle
Finite policy selection (§10.3)	Unique feasible optimum improves a declared lamp task	Theorem for specified deterministic policies and an actual-update heat budget; learning and switching excluded
Policy adaptation (§10.4)	Expected reward improves toward a stationary mixture; cumulative register heat is finite	Theorem for supplied values, thermal channel and initial preparation; task execution and readout installation excluded
Continuing agent (§10.7)	An act–learn–reset sequence retains above-chance reward at positive recurring cost	Theorem for a supplied prior, channels and work store; budgets bound expectations; battery dynamics excluded and the measurement's register carried separately
Store on the trajectory (§10.8)	A reading carried in the state is solvent on every path a funded process has, and an expected allowance is not a battery	Theorem without positivity; a supply covering each draw sustains any horizon and a uniformly positive net draw stops the process; the supply is declared
Finite source (§10.9)	Internal replenishment cancels when source and store are counted together	Conditional finite-resource bound; reversible gates conserve energy but need not deliver it; preparation and gate control are inputs
Sensor memory (§10.10)	Clearing a measurement register costs the entropy removed plus the relative entropy of what it held from the state it is driven to	Theorem for an erasure landing on one declared law; positive at every cycle from channel masses alone; a blind measurement pays the same and a matched memory nothing

Concept	Role in the argument	Status
Funded clearing (§10.11)	A charge carried in the agent's own world and spent once per cycle funds $m \le b/c$ clearings	Theorem for a stage-dependent cost; the erasure cannot be conditioned on the charge, so an exhausted run continues unfunded, and full support admits no net draw at all
Preparing the prior (§10.12)	A constant preparation channel has the same heat identity as erasure; a uniform target has zero mean heat	Theorem for the declared channel; it preserves the parameter marginal and creates no parameter-hardware correlation. Biased bit targets have unbounded heat; exact preparation and implementation costs remain separate
Learning from outcomes (§10.5)	Expected reward of the executed action rises from 1/2 toward 3/4; the work spent is irreducible	Theorem for an unknown parameter, a declared prior and observation-driven updates; preparation, power supply and readout fabrication excluded
Phase sensor (§10.6)	Symbol probabilities follow from oscillator counts; equal coherence can carry different task information	Theorem for supplied phase law, sampling and readout; no phase dynamics
Register's bath (§10.14)	A reversible register–bath gate delivers the erasure heat, and a map-indexed budget needs a prepared bath and a work supply	Theorem (negative); positivity, local detailed balance and the ledger identity do not follow from reversibility
Installed energy (§10.17)	$K \le \kappa U$ excludes the scalar coherent branch when $\kappa U \le 2D$	Conditional theorem; hardware sets κ, installed energy is not heat, and exceeding the line does not guarantee coherence
Learning installs coupling (§10.18)	Reward, kernel and installation work use the same executed law	Exact witness; clears a necessary energy bound. A separate finite phase/register model closes feedback under supplied sensor and price rules; its full sensing and reset heat protocol remains unspecified
KL bound; free-energy critique (§11)	Blocks "dissipation ⇒ true model"	Instance postulate; the limit does *not* follow
Squared-drift gradient (§12)	A descendable objective for plasticity	Theorem at fixed phases; *not* entropy production
Kuramoto SDE	The physical core	Model
Lyapunov descent (§14)	Motion is gradient flow on a potential	Theorem; minimum characterization needs positivity
Rotating frame	Repairs the unbounded potential	Theorem; exact only for identical frequencies
Picard–Lindelöf, global	Trajectories exist and are unique	Theorem on all of $\mathbb{R}$
Barbalat's lemma	Motion stops	Proved here (absent from Mathlib)
Łojasiewicz estimate (§17)	Locates the limit	Theorem, conditional on initial data
Propagation of chaos (§18)	Would link finite N to mean field	**Open**; not closed by any theorem
von Mises density (§19)	The stationary phase law	Theorem classifying positive normalized stationary densities for the specified cosine drift

Concept	Role in the argument	Status
Threshold $K_c = 2D$ (§20)	Coupling gate for coherence	Theorem about a stationary density
Fold-and-cross uniqueness (§21)	Exactly one coherent branch	Theorem; no differentiation used
Critical exponent $\beta = 1/2$ (§22)	Square-root onset	Theorem; filter-indexed, with witness
Ramp-speed dependence (§25)	Apparent exponent tracks the drive	Simulation; a result about method
Squared-drift plasticity (§27)	Dissipation descent $\neq$ learning	Simulation; **negative result**
Mesh refinement	Discrete energy $\to$ continuum energy	Theorem (scalar energy only)
No kernel from coarse-graining (§29)	Continuum kernel is an input	Theorem (negative)
Sheaf gluing (§30)	Local $\to$ unique global state	Theorem + instance postulate
Contents driven by observation	Vector-valued observations preserve agreement and bound retained mismatch by $(1-\eta)^n\varepsilon + \delta$	Theorem; an upper bound, with scalar mass profiles a restricted instance
Coherence bounds disagreement	Phase order bounds vector-content mismatch under encoders shared on overlaps	Conditional theorem; encoders and uniform Lipschitz constant are declared
Thermodynamics off finite state	Path divergence in $[0,\infty]$ on any measurable space: chain rule, data processing, second law; control over a compact action set	Theorem; the Shannon-entropy form does not generalize
Compatibility $\neq$ phase agreement (§31)	The framework's central distinction	Theorem, with explicit counterexample
Twisted gluing	Weakening to per-patch phases	Theorem; adds nothing (offsets are coboundaries)
Winding numbers, phase lifts	Measured topological obstruction	Theorem + empirical candidate; no link to content
Banach fixed point	The reflexive self	Conditional theorem; rate is an obligation
Avatar / restriction resonance	Self-model runs through a region	Theorems, positive and negative halves
Reconstruction over a family (§37.1)	Separated states require distinct codes; $M \le 2^b$ for an alphabet, and the δ-packing number of the codes for a metric	Theorem; declared states, metric and tolerance, without a physical memory model, and a declared readout gain
Locality and latency (§31.5)	Interventions outside a causal past cannot be reported accurately by its deadline	Theorem for local synchronous updates; snapshot compatibility remains possible
Rigid wiring suboptimality	Support constraints cost correlation	Theorem; *not* about biology
Point support and finite regions (§39)	Atomless integrals ignore finite points; positive-mass cells preserve finite matrix weight and correlation	Theorems about support representation; no device exclusion
No-axioms methodology (§41)	Postulates as instance obligations	Methodological; three refutations recorded
Passive and active chains (§43)	Eight hypotheses per branch $\Rightarrow$ the same glued fixed state	Conditional theorems; both jointly satisfiable, budget-to-coupling connection assumed
(a, r) collapse (§46)	Parameter-free distributional test	Compatible with EEG; range too narrow to discriminate
Sleep-inertia recovery	Conditional prediction requiring dynamics and mechanism identification	**Untested**
Rate calibration γ (§49)	Converts field estimate to coupling	**Not derivable**; the sharpest open limitation

Concept	Role in the argument	Status
Compatibility observable	Needed for dissociation tests	**Validated on constructed data**; no neural instantiation
Unity / selfhood identification	What the object is taken to be	**Philosophical commitment**, argued not proved

B Glossary

These are short reminders of explanations developed in the text. The five terms central to the Kuramoto walkthrough are *order parameter*, *rotating frame*, *ansatz*, *stationary branch* and *ramp speed*.

Adiabatic tracking
Remaining close to the fixed-control solution while the control changes slowly. Requires adjustment faster than the target moves; critical slowing down limits it.

Agency model
Specified policy, world, observation and memory-update channels. Permits action to change future inputs; supplies no learned objective.

Amplitude
Size of an oscillation; length of its signal arrow.

Analytic signal
The complex signal $Ae^{i\phi}$ built from a real signal by the Hilbert transform, supplying instantaneous amplitude and phase.

Ansatz A proposed form of solution, with parameters to determine and conditions to check. The stationary ansatz here is a von Mises density (Section 19).

Atomless measure
A measure giving zero mass to every single point. Lebesgue measure is the example; finitely-sited kernels are invisible to integrals against it.

Branch A continuous family of solutions as a control parameter varies.

Barbalat's lemma
If a function is well-behaved and its integral converges, the function tends to zero. Not in Mathlib; proved in the development.

Bessel ratio $R(a)$
$I_1(a)/I_0(a)$, the mean of $\cos\theta$ under a von Mises law at concentration a. Strictly increasing.

Coboundary
A family of offsets on overlaps expressible as differences $\phi_i - \phi_j$ of per-patch values. Kuramoto phases always give one, which is why twisted gluing adds nothing.

Compatibility
Agreement of local descriptions on their shared domains.

Complete space
A space containing the limit of every Cauchy sequence, whose terms eventually become arbitrarily close to one another.

Concentration a
: The von Mises peakedness parameter, estimated here from the slope of the log-density on $\cos\theta$.

Conditional mutual information
: Dependence between two variables remaining when a third is known, averaged over that third variable's values.

Contraction
: A map shrinking all distances by a factor $\Lambda < 1$.

Coupling
: An influence of one oscillator's phase on another's speed; K controls its strength in the uniform model.

Cover
: A collection of regions together covering the chosen whole space.

Critical exponent β
: The power law $r \sim (K - K_c)^\beta$ near threshold. Mean-field value $1/2$.

Critical slowing down
: Divergence of relaxation time at a continuous transition; what limits adiabatic tracking.

Data processing inequality
: Processing cannot create information: $I(X;Z) \le I(X;Y)$ for a Markov chain $X \to Y \to Z$.

Dale's law
: The constraint that a neuron is either excitatory or inhibitory for all its targets.

Density
: A function whose integral over a region gives its probability or mass. Its value at one point is not that point's probability.

Drift
: Systematic change per unit time, before random increments are added.

Equilibrium
: A state that can remain unchanged under fixed conditions. For noisy dynamics, the stationary object is a probability distribution.

Estimator / estimand
: The rule calculating an estimate / the underlying quantity it aims to measure.

Euler–Maruyama
: The basic numerical scheme for SDEs; noise scales as $\sqrt{\Delta t}$.

Feedback
: Influence of the internal state, through action, on the future environmental state that the system observes.

Finite-size floor
: Fluctuating sample order even for independent uniform phases; its root-mean-square size is $1/\sqrt{N}$. Not a hard minimum.

Fixed point
: A state unchanged by a specified map: $f(x) = x$.

Fokker–Planck equation
A partial differential equation (PDE), tracking changes in a probability density across position and time. The model's zero-current stationary equation gives the von Mises form.

Frustration
Interaction preferences that cannot all be satisfied at once. It can permit competing patterns; it does not guarantee useful memories.

Germ, stalk
Data "arbitrarily near" a point, obtained as a colimit over its neighbourhoods.

Gluing axiom
The sheaf condition: compatible local sections assemble uniquely into one global section.

Global section
An element of $\mathcal{F}(X)$ over the whole space, restricting to the given local data.

Kernel A function $K(x, y)$ giving interaction strength for two locations.

Kibble–Zurek mechanism
Failure to follow changing equilibrium conditions near a transition because the adjustment timescale grows. Particular scaling predictions require a specified model and driving protocol.

Landauer's principle
Under its thermodynamic assumptions, resetting one unbiased bit costs at least $k_B T \ln 2$ in heat. The general bound depends on the entropy actually erased.

Lipschitz
Bounded rate of change: $|f(x) - f(y)| \leq L|x - y|$.

Local detailed balance
The physical relation between a transition's forward/reverse probability ratio and reservoir heat at a specified temperature.

Łojasiewicz estimate
An inequality bounding a functional's excess above its minimum by its current rate of decrease. Yields exponential convergence without compactness.

Lyapunov function
A quantity that never increases along trajectories.

Memory budget
A bound on how many states an evaluation may leave its results in. It constrains where answers go and not how they are computed, and it is what forces a Landauer charge without any assumption that the implementation chose to erase (Section 6.2).

Mean field
The population's aggregate influence. Its finite-sample formula is exact for uniform coupling; passing to a deterministic limiting density requires a separate justification.

Measure
A rule assigning weights, such as length, area or probability, to sets. An integral totals values using those weights.

Nonidentifiability
Different model mechanisms producing the same chosen observations, so those observations alone cannot distinguish them.

Order parameter r
Circular resultant length, $|\frac{1}{N}\sum_j e^{i\theta_j}|$: the length of the average phase arrow. One means perfect alignment; zero means cancellation, which can also occur in structured opposing groups (Section 13).

Path law
A probability distribution over transition records. The finite thermodynamic model compares normalized forward and reverse path laws.

Phase Position within a repeating cycle, expressed as an angle.

Phase diffusion D
Random timing spread: noise alone contributes increment variance $2D\Delta t$ over duration Δt.

Phase locking
Constant relative phases. Perfect in-phase locking is the special case in which all phases coincide.

Pitchfork bifurcation
A transition where a branch of solutions emerges continuously from zero as a parameter crosses threshold, the old solution persisting.

Policy A rule assigning action probabilities to internal states. It may be specified directly; the word does not imply learning or optimality.

Preconnected / connected
Topological notions of "in one piece"; continuous images of connected sets are connected—the whole proof of the domain-wall theorem.

Presheaf / sheaf
Data assigned to open sets with restriction maps; a sheaf additionally satisfies the gluing axiom.

Propagation of chaos
Asymptotic independence of particles in an interacting system as $N \to \infty$.

Ramp speed
Rate of change of a control parameter. Here $v = dK/dt$; a finite ramp gives the population finite time to adjust (Section 25).

Reduction
A computation reading many values and returning one summary of them. The mean-field coupling is one, so every implementation of the model performs a reduction; what it costs depends on the memory budget it keeps, not on the substrate.

Relaxation time
Timescale for returning toward a settled state after a disturbance.

Replica
One repeat of a stochastic simulation with its own random seed.

Resultant length
: See order parameter.

Rotating frame
: Coordinates turning at a chosen angular speed; $\phi_i = \theta_i - \Omega t$ removes a shared natural frequency while preserving relative phases and coherence (Section 15).

Self-consistency
: The density generated by a proposed mean field returns the same mean field when its phases are averaged.

Sliced Wasserstein distance
: A tractable distance between high-dimensional distributions, computed by averaging one-dimensional optimal transport over random projections. Used—and shown to have a finite-sample floor—in the propagation-of-chaos diagnostics.

Stability
: Whether small disturbances grow or decay under the dynamics. Existence of a stationary solution alone does not establish it.

Stationary branch
: Stationary solutions collected across choices of a fixed control parameter. It is a graph of solutions versus coupling, not a time course (Section 21).

Stationary density
: A probability pattern unchanged over time at fixed controls in the chosen frame, even while individual oscillators move.

Sufficient statistic
: A function of the data carrying all the information about a parameter. For von Mises it is $\sum_j \cos\theta_j$, i.e. r—which is why an ML concentration estimate would make the collapse test circular.

Trajectory
: The path a system follows through its states over time.

Vacuum manifold
: The set of lowest-energy field configurations. Its topology classifies the defects.

von Mises distribution
: The circular analogue of the Gaussian, $\rho(\theta) \propto e^{a\cos\theta}$.

Winding number
: The integer count of how many times a circle-valued phase wraps around a loop; nonzero winding obstructs a global real-valued lift.

C Where to look next

The manuscript.
: `main.tex` / `main.pdf`. The coherence section gives the physical model; the awakening section develops identifiability, finite ramps and the protocol; the compatibility section connects gluing to the plasticity control; the self-representation section develops the philosophical commitment; and the thermodynamic section gives the conditional composition.

Supplemental Material.
`supplementary.tex` / `supplementary.pdf`. The claim-by-claim map from statements to Lean identifiers is Table S1; the soundness argument, the fold-and-cross proof, the compatibility-observable specification and the full EEG methods are separate sections there.

The Lean development.
`PhysicsOfConsciousness/`, organized by phase. `Chain.lean` is the composition; `Examples/` carries every witness and every fencing counterexample, one file per phase, and is the most useful place to see what the hypotheses actually demand.

Simulations.
`simulations/`, with saved compact summaries and the generator that turns them into the numerical macros used by the manuscript, the Supplemental Material and this primer.

References

Pinotsis, D. A., & Miller, E. K. (2022). Beyond dimension reduction: Stable electric fields emerge from and allow representational drift. *NeuroImage*, 253, 119058.

Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory: from consciousness to its physical substrate. *Nature Reviews Neuroscience*, 17(7), 450-461.

Singer, W., & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neuroscience*, 18, 555-586.

Fries, P. (2015). Rhythms for cognition: communication through coherence. *Neuron*, 88(1), 220-235.

Shadlen, M. N., & Movshon, J. A. (1999). Synchrony unbound: a critical evaluation of the temporal binding hypothesis. *Neuron*, 24(1), 67-77.

Tsuchiya, N., Wilke, M., Frässle, S., & Lamme, V. A. F. (2015). No-report paradigms: extracting the true neural correlates of consciousness. *Trends in Cognitive Sciences*, 19(12), 757-770.

Metzinger, T. (2003). *Being No One: The Self-Model Theory of Subjectivity*. Cambridge, MA: MIT Press.

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, 17(11), 565-573.

Butlin, P., Long, R., Elmoznino, E., Bengio, Y., Birch, J., et al. (2023). Consciousness in artificial intelligence: insights from the science of consciousness. *arXiv preprint* arXiv:2308.08708.

Bobadilla-Suarez, S., Ahlheim, C., Mehrotra, A., Panos, A., & Love, B. C. (2020). Measures of neural similarity. *Computational Brain & Behavior*, 3, 369-383.

DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences*, 11(8), 333-341.

Kriegeskorte, N., & Diedrichsen, J. (2019). Peeling the onion of brain representations. *Annual Review of Neuroscience*, 42, 407-432.

Gallagher, S. (2000). Philosophical conceptions of the self: implications for cognitive science. *Trends in Cognitive Sciences*, 4(1), 14-21.

Baars, B. J. (1997). In the Theater of Consciousness: The Workspace of the Mind. *Oxford University Press*.

Dehaene, S., Charles, L., King, J. R., & Marti, S. (2014). Toward a computational theory of conscious processing. *Current Opinion in Neurobiology*, 25, 76-84.

Friston, K. (2010). The free-energy principle: a unified brain theory?. *Nature Reviews Neuroscience*, 11(2), 127-138.

Milinkovic, B., & Aru, J. (2026). On biological and artificial consciousness: A case for biological computationalism. *Neuroscience & Biobehavioral Reviews*, 181, 106524. doi:10.1016/j.neubiorev.2025.106524.

Aru, J., Larkum, M. E., & Shine, J. M. (2023). The feasibility of artificial consciousness through the lens of neuroscience. *Trends in Neurosciences*, 46(12), 1008–1017. doi:10.1016/j.tins.2023.09.009.

Seth, A. K. (2026). Conscious artificial intelligence and biological naturalism. *Behavioral and Brain Sciences*, 49, e315. doi:10.1017/S0140525X25000032.

Seth, A. K. (2026). The stuff matters: Consciousness, computation, and biology. *Behavioral and Brain Sciences*, 49, e366. doi:10.1017/S0140525X26106906.

Gurnee, W., et al. (2026). Verbalizable Representations Form a Global Workspace in Language Models. *Transformer Circuits Thread*. arXiv:2607.15495.

Aguilera, M., Millidge, B., Tschantz, A., & Buckley, C. L. (2022). How particular is the physics of the free energy principle?. *Physics of Life Reviews*, 40, 24-50.

Biehl, M., Pollock, F. A., & Kanai, R. (2021). A technical critique of some parts of the free energy principle. *Entropy*, 23(3), 293.

Bruineberg, J., Dolega, K., Dewhurst, J., & Baltieri, M. (2022). The Emperor's New Markov Blankets. *Behavioral and Brain Sciences*, 45, e183.

Pinotsis, D. A., & Miller, E. K. (2023). In vivo ephaptic coupling allows memory network formation. *Cerebral Cortex*, 33(17), 9877-9895.

Pinotsis, D. A., Fridman, G., & Miller, E. K. (2023). Cytoelectric coupling: Electric fields sculpt neural activity and "tune" the brain's infrastructure. *Progress in Neurobiology*, 226, 102465.

McFadden, J. (2002). The conscious electromagnetic information (cemi) field theory: the hard problem made easy?. *Journal of Consciousness Studies*, 9(8), 45-60.

Pockett, S. (2012). The electromagnetic field theory of consciousness: a testable hypothesis about the characteristics of conscious as opposed to non-conscious fields. *Journal of Consciousness Studies*, 19(11-12), 191-223.

McFadden, J. (2020). Integrating information in the brain's EM field: the cemi field theory of consciousness. *Neuroscience of Consciousness*, 2020(1), niaa016.

Fröhlich, F., & McCormick, D. A. (2010). Endogenous electric fields may guide neocortical network activity. *Neuron*, 67(1), 129-143.

Anastassiou, C. A., & Koch, C. (2015). Ephaptic coupling to endogenous electric field activity: why bother?. *Current Opinion in Neurobiology*, 31, 95-103.

Pockett, S. (2002). Difficulties with the electromagnetic field theory of consciousness. *Journal of Consciousness Studies*, 9(4), 51-56.

Vöröslakos, M., et al. (2018). Direct effects of transcranial electric stimulation on brain circuits in rats and humans. *Nature Communications*, 9, 483.

Verkhratsky, A., & Nedergaard, M. (2018). Physiology of Astroglia. *Physiological Reviews*, 98(1), 239-389.

Mountcastle, V. B. (1997). The columnar organization of the neocortex. *Brain*, 120(4), 701-722. doi:10.1093/brain/120.4.701.

Horton, J. C., & Adams, D. L. (2005). The cortical column: a structure without a function. *Philosophical Transactions of the Royal Society B*, 360(1456), 837-862. doi:10.1098/rstb.2005.1623.

Rakic, P. (2008). Confusing cortical columns. *Proceedings of the National Academy of Sciences*, 105(34), 12099-12100. doi:10.1073/pnas.0807271105.

Bastos, A. M., Donoghue, J. A., Brincat, S. L., Mahnke, M., Yanar, J., Correa, J., ... & Miller, E. K. (2021). Neural effects of propofol-induced unconsciousness and its reversal using thalamic stimulation. *eLife*, 10, e60824.

Bajwa, I. J., Nilsen, A. S., Skukies, R., Aamodt, A., Ernst, G., Storm, J. F., & Juel, B. E. (2025). A repeated awakening study exploring the capacity of complexity measures to capture dreaming during propofol sedation. *Scientific Reports*, 15, 32746. doi:10.1038/s41598-025-12695-z.

Bajwa, I. J., Nilsen, A. S., Skukies, R., Aamodt, A., Ernst, G., Storm, J. F., & Juel, B. E. (2024). A repeated awakening study exploring the capacity of complexity measures to capture dreaming during propofol sedation (Version 1.0.0) [Data set]. *OpenNeuro*. doi:10.18112/openneuro.ds005620.v1.0.0.

Xie, L., Kang, H., Xu, Q., Chen, M. J., Liao, Y., Thiyagarajan, M., ... & Nedergaard, M. (2013). Sleep drives metabolite clearance from the adult brain. *Science*, 342(6156), 373-377.

Syková, E., & Nicholson, C. (2008). Diffusion in brain extracellular space. *Physiological Reviews*, 88(4), 1277-1340.

Barbour, B. (2017). Analysis of claims that the brain extracellular impedance is high and non-resistive. *Biophysical Journal*, 113(7), 1636-1638.

Logothetis, N. K., Kayser, C., & Oeltermann, A. (2007). In vivo measurement of cortical impedance spectrum in monkeys: implications for signal propagation. *Neuron*, 55(5), 809–823. doi:10.1016/j.neuron.2007.07.027.

Gabriel, S., Lau, R. W., & Gabriel, C. (1996). The dielectric properties of biological tissues: II. Measurements in the frequency range 10 Hz to 20 GHz. *Physics in Medicine and Biology*, 41(11), 2251–2269. doi:10.1088/0031-9155/41/11/002.

Norton, J. D. (2005). Eaters of the lotus: Landauer's principle and the return of Maxwell's demon. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 36(2), 375-411.

Landauer, R. (1961). Irreversibility and heat generation in the computing process. *IBM Journal of Research and Development*, 5(3), 183-191.

Kibble, T. W. B. (1976). Topology of cosmic domains and strings. *Journal of Physics A: Mathematical and General*, 9(8), 1387-1398.

Sakaguchi, H. (1988). Cooperative phenomena in coupled oscillator systems under external fields. *Progress of Theoretical Physics*, 79(1), 39-46.

Strogatz, S. H., & Mirollo, R. E. (1991). Stability of incoherence in a population of coupled oscillators. *Journal of Statistical Physics*, 63, 613–635.

Acebrón, J. A., Bonilla, L. L., Pérez Vicente, C. J., Ritort, F., & Spigler, R. (2005). The Kuramoto model: A simple paradigm for synchronization phenomena. *Reviews of Modern Physics*, 77, 137–185. doi:10.1103/RevModPhys.77.137.

Ott, E., & Antonsen, T. M. (2008). Low dimensional behavior of large systems of globally coupled oscillators. *Chaos*, 18, 037113. doi:10.1063/1.2930766.

Dörfler, F., & Bullo, F. (2014). Synchronization in complex networks of phase oscillators: A survey. *Automatica*, 50, 1539–1564. doi:10.1016/j.automatica.2014.04.012.

Breakspear, M. (2017). Dynamic models of large-scale brain activity. *Nature Neuroscience*, 20, 340–352. doi:10.1038/nn.4497.

Still, S., Sivak, D. A., Bell, A. J., & Crooks, G. E. (2012). Thermodynamics of prediction. *Physical Review Letters*, 109(12), 120604.

Kawai, R., Parrondo, J. M. R., & Van den Broeck, C. (2007). Dissipation: The phase-space perspective. *Physical Review Letters*, 98(8), 080602.

Horowitz, J. M., & Esposito, M. (2014). Thermodynamics with continuous information flow. *Physical Review X*, 4, 031015. doi:10.1103/PhysRevX.4.031015.

Ito, S., & Sagawa, T. (2013). Information thermodynamics on causal networks. *Physical Review Letters*, 111, 180603. doi:10.1103/PhysRevLett.111.180603.

Corberi, F., dello Russo, S., Messuti, G., Scarpetta, S., & Smaldone, L. (2026). Thermodynamic learning. *arXiv preprint*. arXiv:2609.04732.

Banach, S. (1922). Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales. *Fundamenta Mathematicae*, 3(1), 133-181.

Shenker, O. R. (2000). Logic and entropy. *PhilSci-Archive preprint*. philsci-archive.pitt.edu/115.

Rosas, F. E., Mediano, P. A. M., Luppi, A. I., Boyd, A. B., Berta, M., Jensen, H. J., Gastpar, M., & Carroll, S. M. (2026). Four faces of information in natural systems. *Nature Reviews Physics*. doi:10.1038/s42254-026-00971-4.

Sznitman, A.-S. (1991). Topics in propagation of chaos. In *École d'Été de Probabilités de Saint-Flour XIX—1989*, Lecture Notes in Mathematics, vol. 1464, pp. 165–251. Springer. doi:10.1007/BFb0085169.

Dai Pra, P., & den Hollander, F. (1996). McKean–Vlasov limit for interacting random processes in random media. *Journal of Statistical Physics*, 84, 735–772. doi:10.1007/BF02179656.

Bertini, L., Giacomin, G., & Pakdaman, K. (2010). Dynamical aspects of mean field plane rotators and the Kuramoto model. *Journal of Statistical Physics*, 138, 270–290. doi:10.1007/s10955-009-9908-9.

Berglund, N., & Gentz, B. (2002). Pathwise description of dynamic pitchfork bifurcations with additive noise. *Probability Theory and Related Fields*, 122, 341–388. doi:10.1007/s004400100174.

Abramsky, S., & Brandenburger, A. (2011). The sheaf-theoretic structure of non-locality and contextuality. *New Journal of Physics*, 13, 113036. doi:10.1088/1367-2630/13/11/113036.

Townsend, R. G., Solomon, S. S., Chen, S. C., Pietersen, A. N. J., Martin, P. R., Solomon, S. G., & Gong, P. (2015). Emergence of complex wave patterns in primate cerebral cortex. *The Journal of Neuroscience*, 35(11), 4657-4662. doi:10.1523/JNEUROSCI.4509-14.2015.

Xu, Y., Long, X., Feng, J., & Gong, P. (2023). Interacting spiral wave patterns underlie complex brain dynamics and are related to cognitive processing. *Nature Human Behaviour*, 7(7), 1196-1215. doi:10.1038/s41562-023-01626-5.

Nicoletti, G., Di Terlizzi, I., & Busiello, D. M. (2026). Balancing information and dissipation with partially observed fluctuating signals. *Physical Review Letters*, 137(12), 127101. doi:10.1103/8gx2-rbns.

Kiyooka, D., Oomoto, I., Kitazono, J., Saito, Y., Kobayashi, M., Matsubara, C., Kobayashi, K., Murayama, M., & Oizumi, M. (2026). Single-cell resolution functional networks during unconsciousness are segregated into spatially intermixed modules. *Cell Reports*, 45(2), 116902. doi:10.1016/j.celrep.2025.116902.

Oomoto, I., Kiyooka, D., Oizumi, M., & Murayama, M. (2026). Multi-area single-cell calcium imaging dataset of the mouse cortex across wakefulness, sleep, and anesthesia. *Scientific Data*. doi:10.1038/s41597-026-08202-2.

Attwell, D., & Laughlin, S. B. (2001). An energy budget for signaling in the grey matter of the brain. *Journal of Cerebral Blood Flow & Metabolism*, 21(10), 1133-1145. doi:10.1097/00004647-200110000-00001.

Harris, J. J., Jolivet, R., & Attwell, D. (2012). Synaptic energy use and supply. *Neuron*, 75(5), 762-777. doi:10.1016/j.neuron.2012.08.019.

Nicholls, D. G., & Ferguson, S. J. (2013). *Bioenergetics 4*. Academic Press.

Baer, S. M., Erneux, T., & Rinzel, J. (1989). The slow passage through a Hopf bifurcation: delay, memory effects, and resonance. *SIAM Journal on Applied Mathematics*, 49(1), 55–71. doi:10.1137/0149003.

Friedman, E. B., Sun, Y., Moore, J. T., Hung, H.-T., Meng, Q. C., Perera, P., Joiner, W. J., Thomas, S. A., Eckenhoff, R. G., Sehgal, A., & Kelz, M. B. (2010). A conserved behavioral state barrier impedes transitions between anesthetic-induced unconsciousness and wakefulness: evidence for neural inertia. *PLoS ONE*, 5(7), e11903. doi:10.1371/journal.pone.0011903.

Hudson, A. E., Calderon, D. P., Pfaff, D. W., & Proekt, A. (2014). Recovery of consciousness is mediated by a network of discrete metastable activity states. *Proceedings of the National Academy of Sciences*, 111(25), 9283–9288. doi:10.1073/pnas.1408296111.

Proekt, A., & Hudson, A. E. (2018). A stochastic basis for neural inertia in emergence from general anaesthesia. *British Journal of Anaesthesia*, 121(1), 86–94. doi:10.1016/j.bja.2018.02.035.

Bobadilla-Suarez, S. (2026). Unity requires physically enforced agreement. *PsyArXiv*. doi:10.31234/osf.io/dkqep_v1.