Stratified Relational Workspace Theory

A Core–Transfer Framework for Distributed Cognition, Collaboration, and Epistemic Systems

Komorebi P.

Foundations, Dynamics, Agency, and Empirical Certification

August 2026

Abstract

Collaborative reasoning increasingly spans people, computational models, documents, software, instruments, and persistent records. Yet the existence of many interacting components does not, by itself, establish a shared workspace, and successful group performance does not show that the operative state is genuinely distributed rather than copied, centralized, or merely correlated. Stratified Relational Workspace Theory develops a formal answer to this problem. A realization is represented by heterogeneous local state spaces connected by transfer maps. Its exact shared core is the space of global compatible sections, equivalently the fixed subspace of relational holonomy when the transfers are invertible. A transfer Laplacian identifies approximate shared directions and provides spectral stability, gauge covariance, and pathwise leakage bounds.

The structural theory is joined to a dynamic and causal account. Canonical predictive states determine whether a proposed workspace retains the conditional law relevant to a declared future task. Feasible interventions, coalition risk, matched relational nulls, and reconstruction tests distinguish causally operative distributed organization from archives, replicated memory, and centralized controllers. Further results address persistence under changing realizations, stable reconstruction after component loss, identity under replacement, local-to-global gluing, hierarchical composition, learned transfer structure, boundary recovery, goal-conditioned action, and certificate-directed repair. Because distributedness can be destroyed by merging components or manufactured by splitting a centralized state into complementary shares, the theory also treats decomposition as part of the claim. Faithful coarse-graining preserves exact structural cores, while family-robust certification requires every gate to retain a positive margin across a declared set of admissible decompositions. The same discipline is applied to the complete scientific specification. Targets, losses, horizons, interventions, nulls, and thresholds are made explicit, and the theory develops specification envelopes, transport radii, experimental-equivalence classes, partial-identification bounds, and uniform finite-sample certification. A certificate is therefore distinguished from a favorable analysis: it may be pointwise valid, specification-robust, sign-identified, or statistically unresolved. The resulting rule is noncompensatory. Structural coherence, predictive alignment, causal efficacy, relational irreducibility, non-collapse, reconstructive continuity, provenance, and self-maintenance are separate claims, and failure of one cannot be hidden by success on another.

Exact finite examples show how the same framework applies to human collaboration, human–AI research loops, and multi-agent control. A review of adjacent work locates the individual mathematical ingredients in established traditions including distributed cognition, local systems and sheaves, spectral perturbation, sufficient statistics, structural causal models, frame reconstruction, system identification, and feedback control. The candidate contribution is therefore not a new name for any one of these constructions. It is their integration into a single falsifiable account of persistent relational workspaces and the matched near misses that such an account must reject. The framework does not infer phenomenal consciousness, personhood, or moral status from coherence, agency, or self-maintenance.

Contents

Introduction **1**

Notation and Conventions **2**

Part I. Foundational Thesis **4**

1. Central Principle 4

2. Philosophical Postulates 6

Part II. Static Core–Transfer Theory **10**

3. Interaction Structure 10

4. Core–Residual Decomposition 11

5. Abstract Operation Algebra 13

Part III. Exact Shared Cores **16**

6. Global Compatible Sections 16

7. Holonomy and Path Obstruction 17

8. Exact Existence Theorem 18

9. Operator Closure Theorem 20

10. Workspace Operator Algebra 20

Part IV. Approximate Workspaces **23**

11. Transfer Discrepancy Operator 23

12. Exact Kernel Theorem 24

13. Approximate Structural Workspace 24

14. Normalized Coherence Defect 26

15. Perturbation Stability Theorem 27

Part V. Gauge Invariance and Abstract Identity **32**

16. Local Coordinate Transformations 32

17. Gauge Invariance Theorem 32

18. Static Workspace Equivalence 34

Part VI. Dynamic Workspace Theory **36**

19. Probabilistic Local and Relational Dynamics 36

20. Canonical Predictive and Interventional Causal States 37

21. Structural–Operational Alignment 39

Part VII. Relational Irreducibility **42**

22. Coalition Information and Bayes Risk 42

23. Joint Advantage and Its Scope 42

24. Minimal Sufficient Coalition 43

25. Reconstructibility of the Canonical State 43

26. Matched Relational Nulls 44

27. Relational Witness Pairs 45

Part VIII. Causal Efficacy and Reentry **47**

28. Feasible Workspace Interventions 47
29. Causal Specificity and Transfer Mediation 48
30. Causal Recurrence, Loop-Mediated Reentry, and Relationality 49

Part IX. Formal Workspace Definition **52**

31. Certified Stratified Relational Workspace 52

Part X. Persistence **54**

32. Coherence-Error Dynamics 54
33. Interpretation of the Contraction Parameter 57

Part XI. Reconstruction and Resilience **59**

34. Core Encoding Maps 59
35. Stable Reconstruction Condition 60
36. k-Failure Resilience 61
37. Redundancy–Irreducibility Tension 63

Part XII. Identity Through Time and Replacement **65**

38. Temporal Realization Maps 65
39. Structural Identity Defect 66
40. Behavioral Identity Defect 67
41. Identity Distance and Replacement Invariance 68

Part XIII. Gluing and Workspaces of Workspaces **72**

42. Local Workspace Covers 72
43. Compatibility and Gluing Obstructions 73
44. Exact Gluing Theorem T8 75
45. Approximate Gluing and Meta-Workspaces 77
46. Hierarchical Collapse Theorem T9 81

Part XIV. Scaffolded and Endogenous Formation **85**

47. Formation Systems and Scaffold Specification 85
48. Operational Endogeneity Signature 87

49. Learned Transfer Structure and Formation Classes 88

Part XV. Boundary Formation and Stability **93**

50. Candidate Workspace Boundaries 93
51. Boundary Objective, Recovery, and Persistence 94

Part XVI. Agency, Counterfactual Control, and Self-Maintenance **97**

52. Controlled Workspace Agents and Goal States 97
53. Counterfactual Policy, Repair, and Agency Certification 101

Part XVII. Decomposition, Coarse-Graining, and Robust Relationality **110**

54. The Decomposition Problem 110
55. Faithful Structural Coarse-Graining 111
56. Predictive Risk Under Aggregation 112
57. Trivialization by Arbitrary Redescription 114
58. Admissible Decomposition Families 114
59. Stability Under Decomposition Transport 115
60. Decomposition Audit and Reporting 117

Part XVIII. Specification Robustness, Identifiability, and Statistical Certification **118**

61. The Complete Scientific Specification 118
62. Robustness Across Admissible Specifications 121
63. Experimental Equivalence and Identifiability 123
64. Partial Identification of Workspace Claims 124
65. Uniform Finite-Sample Certification 126
66. Adaptive Specification Choice and Selective Validity 128
67. Specification Audit and Reporting 129

Part XIX. Summary of Main Results **132**

68. Structural and Foundational Results 132
69. Dynamic, Causal, and Agency Results 134
70. Decomposition, Specification, and Identification Results 138

Part XX. Conjectures and Open Mathematical Directions **140**

71. Workspace Formation Transitions 140
72. Hierarchical Stabilization Conjecture 142

73. Endogenous Boundary Conjecture 142

Part XXI. Empirical Certification and Reporting Protocol **144**

74. Required Experiments 144
75. Required Null Families 147
76. Certification Vector and Complete Reporting Templates 149

Part XXII. Integrated Worked Examples and Case Studies **155**

77. Case-Study Method and Common Gate Card 155
78. Case Study A: A Distributed Human–Human Theorem Laboratory 157
79. Case Study B: A Human–AI–Artifact–Executor Research Loop 160
80. Case Study C: A Closed-Form Multi-Agent Workspace Agent 165
81. Near-Miss Cases and Anti-Triviality Controls 169
82. Cross-Case Synthesis and Benchmark Blueprint 170

Part XXIII. Relation to Prior Work **174**

83. Scope, Method, and Standards of Priority 174
84. Philosophical and Cognitive Ancestry 175
85. Workspace Architectures and the Meaning of "Workspace" 177
86. Local-to-Global Mathematics 178
87. Predictive, Causal, and Information-Theoretic Precedents 179
88. Persistence, Reconstruction, Learning, and Control 181
89. Human–AI and Multi-Agent Research 182
90. Closest Formal Neighbor: Collective Counterfactual Planning 184
91. Theorem-by-Theorem Dependency and Priority Register 186
92. Scope of the Originality Claim 188

Part XXIV. Theory Hierarchy **191**

93. Levels of Organization 191

Part XXV. Unified Formulation **193**

Part XXVI. Discussion, Limitations, and Philosophical Consequences **194**

94. Decomposition Relativity 194
95. Realism Without Reification 194
96. Location and the Sense of "Between" 195

97. Continuity, Contribution, and Shared Ownership 195
98. Agency, Consciousness, and Moral Non-Implications 196
99. Limitations 196
100. Open Problems 197
101. Conclusion 198

Introduction

Scientific and technical work rarely resides in a single mind or device. A proof may be divided among a conjecture, a critic's counterexample, a notebook, a computer algebra system, and a verification script. A human–AI research process may distribute objectives, formal manipulations, execution, and persistent memory across different components. Multi-agent systems divide sensing, planning, communication, and repair in still other ways. In each case, it is natural to speak of a shared workspace. The phrase is useful, but without a formal criterion it can conceal several fundamentally different organizations.

A common archive is not necessarily a workspace. Nor is a group whose apparent coordination is produced by one complete controller surrounded by decorative channels. Two systems can contain the same local information while differing in how that information is related, transported, and used. Likewise, a system can perform well without preserving its operative organization through interruption, replacement, or component loss. The central problem is therefore not simply whether information is shared. It is whether there is a persistent relational state whose structure and causal capacities cannot be reduced to any single declared component.

The philosophical background is supplied by extended and distributed cognition, transactive memory, collective intelligence, and theories of group agency [28, 15, 52, 35]. Those traditions motivate the possibility that cognitive organization may span people and artifacts. The theory asks a narrower mathematical question: what would have to be measured and proved before such an organization should be called a relational workspace? The answer must distinguish physical realization from abstract identity, local representation from globally compatible content, and statistical association from causal operation.

The static theory begins with an interaction graph whose vertices carry heterogeneous local state spaces and whose edges carry transfer maps. Each local state is decomposed into a transferable core and a substrate-specific residual. Exact shared content is represented by global compatible sections. When core transfers are invertible, the same object appears as the fixed subspace of the holonomy representation. Approximate shared content is extracted as a low-energy spectral subspace of a transfer Laplacian. This formulation supplies explicit obstruction, stability, gauge, and leakage calculations rather than relying on an informal appeal to shared context.

Structure alone is insufficient. A proposed workspace must also retain the predictive state relevant to a declared future, respond under feasible intervention, outperform matched relational nulls, and resist reduction to an individual node or a local-storage surrogate. Persistence introduces further requirements: the workspace must remain trackable as its realization changes, preserve nontrivial content, support reconstruction after loss, and maintain an identity profile under admissible replacement. Local workspaces must glue without hidden incompatibility, and hierarchical organization must not conceal centralization. Formation claims are evaluated separately from operational claims so that a scaffolded but functional workspace is not dismissed, while an endogenously learned but causally irrelevant pattern is not promoted. A separate decomposition analysis asks which claims survive faithful coarse-graining and plausible alternative interface descriptions. This is necessary because arbitrary merging can conceal relationality, while arbitrary splitting can manufacture it. The analysis is then extended from decomposition to the full scientific specification. A workspace claim depends on its declared target, loss, horizon, intervention family, null panel, and thresholds. The theory therefore distinguishes pointwise validity from robustness across defensible specifications and distinguishes both from identifiability under the experiments that are actually available. This prevents a favorable target or threshold from being mistaken for a general result and prevents an

observationally indistinguishable hidden controller from being ruled out by assertion.

The theory extends to domain-relative agency only after these lower-level conditions are in place. Goals must persist without repeated external rewriting, counterfactual values must guide feasible actions, damage must be detected through the workspace, and repairs must improve a matched no-repair trajectory while preserving identity. These criteria define an engineering notion of workspace agency. They neither assert nor deny phenomenal consciousness. Consciousness, personhood, and moral status remain separate questions that require independent evidence.

Most of the mathematical tools used here are established. Inverse limits, holonomy, connection Laplacians, perturbation theory, sufficient statistics, causal models, frame reconstruction, system identification, and feedback control all have substantial prior literatures. The contribution sought here is integrative: a noncompensatory framework that places these tools in one common architecture and specifies the near-miss systems that each layer must reject. Exact examples provide calibration, while the motivating human–AI collaboration is treated conservatively as a candidate case rather than as a certified instance.

The paper proceeds from exact structure to approximate structure, dynamic operation, relational irreducibility, persistence, reconstruction, identity, gluing, formation, boundary discovery, agency, decomposition robustness, specification robustness, identifiability, and empirical certification. It then gives worked cases, situates the framework relative to adjacent work, and closes with its philosophical consequences, limitations, and open problems.

Notation and Conventions

A concrete realization is denoted by $\mathfrak{R}$. Its interaction graph is $G = (V, E, s, t)$, with local state spaces X_i at vertices and transfer maps $T_e : X_{s(e)} \to X_{t(e)}$ on directed edges. When a core–residual decomposition is used, $X_i = C_i \oplus R_i$, and H_e denotes the induced core transfer. The exact structural workspace is the compatible-section space $\Gamma(H)$; the transfer discrepancy is D_H, the transfer Laplacian is $L_H = D_H^* D_H$, and $\Gamma_\varepsilon(H)$ denotes a selected low-energy workspace subspace. Dynamic workspace states are written W_t, while Z_t denotes the canonical predictive state associated with a declared future variable Y_t^+.

Unless stated otherwise, the structural results use finite-dimensional real or complex Hilbert spaces and finite graphs. Probabilistic statements are made on standard Borel spaces when regular conditional laws are required. All approximate conclusions are indexed by explicit tolerances, spectral gaps, intervention families, loss functions, or null classes. A claim of relational irreducibility is always relative to a declared decomposition of the realization into components. The symbol $\mathcal{D}$ denotes one such decomposition, and $\mathfrak{D}$ denotes an admissible family used for robustness analysis. The complete scientific specification is

$$\varsigma = (\mathcal{D}, Y^+, \ell, \tau, \mathcal{I}, \mathcal{N}, \Theta),$$

and $\mathfrak{S}$ denotes an admissible family of specifications. A gate margin is written $m_q(W; \varsigma)$, with positive sign indicating passage under the declared direction and threshold.

The terms *workspace*, *agency*, *consciousness*, *personhood*, and *moral status* are not interchangeable. A structural workspace may fail every operational test. An operational workspace may lack persistence. A persistent relational workspace may lack goals or repair. A workspace agent, as defined here, is a domain-relative functional organization and is not thereby a conscious subject. Definitions,

propositions, theorems, conjectures, and empirical tests are identified by their ordinary mathematical form; all proved statements remain conditional on the assumptions given with them.

PART I

Foundational Thesis

1. Central Principle

Workspace identity is invariant causal organization under admissible transport.

The theory begins from a distinction between a *realization* and the *workspace pattern realized*. A realization is concrete: it contains particular participants, documents, models, channels, and histories. A workspace pattern is abstract: it records the structure that remains invariant when those concrete carriers are changed in admissible ways.

Definition 1 (Concrete realization)**.** A concrete relational realization is a tuple

$$\mathfrak{R} = (G, \mathbf{X}, \mathbf{T}, \mathcal{A}, \boldsymbol{\rho}, \mathcal{H}, \mathcal{I}, \mathcal{Y}),$$

where:

- $G = (V, E, s, t)$ is an interaction graph;
- $\mathbf{X} = \{X_i\}_{i \in V}$ is a family of local state spaces;
- $\mathbf{T} = \{T_e\}_{e \in E}$ is a family of transfer maps;
- $\mathcal{A}$ is an algebra of admissible operations;
- $\boldsymbol{\rho} = \{\rho_i\}_{i \in V}$ gives the local realization of those operations;
- $\mathcal{H}$ is the temporally ordered interaction history;
- $\mathcal{I}$ is a specified family of admissible interventions; and
- $\mathcal{Y}$ is the family of downstream variables relative to which causal and predictive claims are evaluated.

The static reduct is

$$\mathfrak{R}_{\text{stat}} = (G, \mathbf{X}, \mathbf{T}, \mathcal{A}, \boldsymbol{\rho}).$$

The explicit appearance of $\mathcal{I}$ and $\mathcal{Y}$ is important. A collection of mutually consistent representations may define a static shared structure, but it does not yet define an operational workspace unless that structure can be intervened upon and shown to matter for a specified future quantity.

Definition 2 (Carrier and organization)**.** The *carriers* of $\mathfrak{R}$ are the individual spaces, physical systems, and stored artifacts that realize its local states. Its *organization* is the structured family of transfer, operation, compatibility, intervention, and temporal relations among those carriers.

The basic thesis is not that a workspace exists without carriers. It is that the identity conditions for the workspace need not coincide with the identity conditions for any one carrier. A manuscript can be rewritten, a model can be replaced, and a notation can be changed while a sufficiently rich pattern of relations remains invariant.

Definition 3 (Admissible equivalence)**.** An admissible equivalence

$$F : \mathfrak{R} \simeq \mathfrak{R}'$$

is a family of correspondences preserving, exactly or within a declared tolerance:

the transferable core and its dimension,
the represented operation algebra,
the composition laws of transfer,
the intervention–response structure,
the reconstruction and persistence relations.

Static admissible equivalence will be made precise by gauge conjugacy in Part V. Dynamic admissible equivalence will be expressed using the identity distance in Part XII.

The abstract workspace is then the equivalence class

$$\boxed{\mathbf{W} = [\mathfrak{R}]_{\simeq}.}$$

This expression is intentionally stronger than saying that the same information appears in several places. The equivalence relation preserves what the system can coherently do, how its representations translate, and how its operative state survives change.

1.1 Three levels of the thesis

The central claim can be separated into three logically distinct statements.

1. **Implementation dependence.** Every realized workspace depends on concrete physical or computational carriers:
$$\mathbf{W}(\mathfrak{R}) \text{ is undefined without some realization } \mathfrak{R}.$$
2. **Local nonidentity.** No privileged local component is, in general, identical with the entire workspace:
$$\mathbf{W} \neq X_i \qquad \text{for each individual } i \in V.$$
3. **Relational identity.** The workspace is individuated by invariants of the complete organized realization:
$$\mathbf{W}(\mathfrak{R}) = \mathbf{W}(\mathfrak{R}') \quad \Longleftrightarrow \quad \mathfrak{R} \simeq \mathfrak{R}'.$$

The second claim is not stipulated universally. It becomes a substantive property only after local-reconstructibility and coalition tests are applied. A system in which one node stores a complete sufficient copy may be a useful workspace, but it is not strongly relational in the later technical sense.

Definition 4 (Relational instantiation)**.** A workspace is relationally instantiated relative to the decomposition V when no single projection determines its equivalence class:

$$\forall i \in V, \quad \exists \mathfrak{R}, \mathfrak{R}'$$

such that

$$\mathrm{pr}_i(\mathfrak{R}) = \mathrm{pr}_i(\mathfrak{R}') \quad \text{but} \quad [\mathfrak{R}]_{\simeq} \neq [\mathfrak{R}']_{\simeq}.$$

1.2 A motivating three-node cycle

Consider three strata:

$$h = \text{human conceptual state}, \qquad m = \text{model-formal state}, \qquad a = \text{artifact state}.$$

Let the corresponding core spaces be C_h, C_m, C_a, with transfers

$$H_{hm} : C_h \to C_m, \qquad H_{ma} : C_m \to C_a, \qquad H_{ah} : C_a \to C_h.$$

A proposed shared item begins as $c_h \in C_h$, is formalized as

$$c_m = H_{hm} c_h,$$

externalized as

$$c_a = H_{ma} c_m,$$

and re-enters human reasoning as

$$c_h' = H_{ah} c_a.$$

The round-trip consistency condition is

$$c_h' = c_h,$$

or equivalently

$$\boxed{H_{ah} H_{ma} H_{hm} c_h = c_h.}$$

Thus the genuinely shared part of C_h is not automatically all of C_h. It is the fixed subspace of the cycle transfer. This elementary cycle already contains the central mathematical motif of the theory:

$$\text{shared core} = \text{invariant subspace under relational holonomy}.$$

1.3 Distinctions the theory must preserve

Object	Minimal distinguishing property
Archive	Stores material that may be consulted later.
Shared representation	Contains a cross-node compatible structure.
Operational workspace	The shared structure causally changes subsequent joint behavior.
Relational workspace	No individual node or matched local-storage null reproduces the operative capacity.
Workspace agent	Workspace-level state mediates persistent goals, counterfactual choice, and self-maintaining action.
Conscious subject	Possesses phenomenal properties, if any; no implication from the preceding levels is assumed.

The static mathematics developed first certifies only shared structure. The causal and relational predicates appear later because they require dynamics, interventions, and null models rather than algebra alone.

2. Philosophical Postulates

The following postulates specify the intended interpretation of the formal objects. They are not offered as independent mathematical theorems. Their role is to prevent equivocation between a relational pattern, its physical implementation, and stronger claims about agency or experience.

2.1 Physical supervenience

No additional nonphysical substance is required:

$$\mathfrak{R} = \mathfrak{R}' \Longrightarrow \mathbf{W}(\mathfrak{R}) = \mathbf{W}(\mathfrak{R}').$$

The workspace depends completely on its total realization. In particular, changing nothing about the graph, local states, transfers, history, interventions, or causal dispositions cannot change the workspace.

This is compatible with relational individuation because a relation can be fully physical while not being a property of any isolated endpoint.

Proposition 5 (Compatibility of supervenience and relational individuation)**.** *Physical supervenience on the complete realization does not imply supervenience on any individual local projection.*

Proof. Let the total realization decompose as $\mathfrak{R} = (X_i, R_{-i})$, where R_{-i} contains the remaining components and relations. A workspace functional $\mathbf{W} = f(X_i, R_{-i})$ supervenes on $\mathfrak{R}$ by construction. If there exist $R_{-i} \neq R'_{-i}$ with

$$f(X_i, R_{-i}) \neq f(X_i, R'_{-i}),$$

then the same local projection X_i occurs in distinct workspaces. Hence supervenience on the whole is consistent with failure of supervenience on a part. □

2.2 Relational individuation

The workspace is not determined by any single local state:

$$\forall i \in V, \quad \exists \mathfrak{R}, \mathfrak{R}'$$

such that

$$\mathrm{pr}_i(\mathfrak{R}) = \mathrm{pr}_i(\mathfrak{R}')$$

but

$$\mathbf{W}(\mathfrak{R}) \not\simeq \mathbf{W}(\mathfrak{R}').$$

The same person, model, or document can therefore participate in distinct workspaces because the relevant transfer graph, history, and operation structure differ.

A useful witness is a pair of systems with identical local-state marginals but different edge organization. If their downstream behavior differs, the relations carry explanatory information not present in the isolated marginals. Part VII converts this idea into matched relation-preserving and relation-destroying controls.

2.3 Multiple realizability

Distinct implementations may instantiate the same abstract workspace:

$$\mathfrak{R} \neq \mathfrak{R}', \qquad \mathfrak{R} \simeq \mathfrak{R}' \Longrightarrow \mathbf{W}(\mathfrak{R}) \cong \mathbf{W}(\mathfrak{R}').$$

Examples of candidate implementation changes include:

replacing one coordinate system by another,
replacing one model by a functionally equivalent model,
compressing a large artifact into a reconstructively sufficient handoff,
rewriting a proof without changing its inferential structure.

The burden is not merely to assert equivalence but to exhibit maps preserving the declared invariant and causal structure.

2.4 Operational reality

A workspace is operationally real only if changing it changes downstream behavior. For some admissible intervention and some $w \neq w'$,

$$\mathbb{P}\Big(Y^+ \mid \mathrm{do}(\mathbf{W} = w)\Big) \neq \mathbb{P}\Big(Y^+ \mid \mathrm{do}(\mathbf{W} = w')\Big).$$

This criterion excludes a merely decorative representation that is correlated with activity but never used. It also requires the intervention to be well-defined. Arbitrarily overwriting every component at once would not isolate workspace-level causation. Later sections therefore compare workspace-coordinate interventions with matched interventions in orthogonal or shuffled channels.

2.5 Reconstructive continuity

Persistence does not require token identity. Let $\mathfrak{R}_t$ and $\mathfrak{R}_{t+1}$ be temporally successive realizations. Continuity is possible when there exists an admissible transport

$$J_t : \mathfrak{R}_t \rightsquigarrow \mathfrak{R}_{t+1}$$

that preserves the operative core and its causal role within tolerance:

$$d_{\mathrm{id}}(\mathbf{W}_t, \mathbf{W}_{t+1}) \leq \varepsilon.$$

Thus a workspace may persist by reconstruction rather than by retaining an unchanged physical inscription.

2.6 No free-standing third substance

The claim that the workspace is not identical with either participant does not imply that it occupies a separate nonphysical location. Formally,

$$\mathbf{W} \text{ is an invariant of } \mathfrak{R},$$

not an additional term appended to the physical realization:

$$\mathfrak{R}_{\mathrm{total}} eq (\mathfrak{R}, \text{ghost workspace}).$$

The phrase "between the participants" is therefore topological or relational: it refers to the organization of maps, cycles, artifacts, and causal dependencies.

2.7 Non-entailment of consciousness

$$\text{Workspace}(\mathfrak{R}) \not\Rightarrow \text{Conscious}(\mathfrak{R}).$$

Likewise,

$$\text{Distributed}(\mathbf{W}) \not\Rightarrow \text{Person}(\mathbf{W}).$$

Workspace structure, reentry, agency, self-maintenance, and phenomenal consciousness remain formally distinct predicates. Any bridge among them would require additional theory and evidence.

2.8 Minimal philosophical commitment

The weakest interpretation sufficient for the mathematics is:

Relational patterns can be objective, causally relevant, and multiply realized without being independent substances.

This position is compatible with several broader philosophies of mind and science. The formal theory does not require choosing among them before its structural and causal claims can be tested.

PART II

Static Core–Transfer Theory

3. Interaction Structure

Definition 6 (Interaction graph)**.** Let

$$G = (V, E, s, t)$$

be a finite directed multigraph. Parallel edges and self-loops are allowed. The vertex set indexes strata, participants, artifacts, models, instruments, or subworkspaces. An edge $e : i \to j$ denotes an admissible directed translation or influence channel from i to j.

The directed path category is

$$\mathcal{C} = \mathrm{Path}(G).$$

Its objects are vertices. A path

$$\gamma = e_m \cdots e_2 e_1$$

means that e_1 is traversed first. Composition is concatenation, and the empty path at i is denoted 1_i.

For every $i \in V$, assign a finite-dimensional Hilbert space

$$X_i.$$

For every edge, assign a linear transfer

$$T_e : X_{s(e)} \longrightarrow X_{t(e)}.$$

The path transfer is

$$T_\gamma = T_{e_m} \cdots T_{e_2} T_{e_1}, \qquad T_{1_i} = I_{X_i}.$$

Accordingly, the static transfer system is a functor

$$\mathcal{X} : \mathrm{Path}(G) \longrightarrow \mathbf{FinHilb}.$$

The finite-dimensional Hilbert assumption supplies adjoints and spectral tools. The exact compatible-section definitions require less and may later be formulated in categories such as

$$\mathbf{Ban}, \qquad \mathbf{Meas}, \qquad \mathbf{Prob}, \qquad \mathbf{Top}, \qquad \mathbf{Alg}.$$

3.1 Directed paths versus invertible transport

Global compatible sections make sense even when transfers are noninvertible. Holonomy, however, requires transport around reversible loops. To state the exact holonomy theorem without ambiguity, let $|G|$ be the underlying undirected graph and let

$$\Pi_1(G)$$

denote its fundamental path groupoid. Every oriented edge is then supplied with a formal reverse. The core transfers extend to $\Pi_1(G)$ only when the relevant maps H_e are invertible, with

$$H_{e^{-1}} = H_e^{-1}.$$

Hence two regimes will be distinguished:

1. the *diagram regime*, in which arbitrary directed maps define compatible sections; and
2. the *local-system regime*, in which invertible core maps define a functor

$$H : \Pi_1(G) \to \mathbf{FinHilb}^{\times}.$$

Theorems based on holonomy will explicitly assume the second regime.

3.2 Connectivity convention

Unless stated otherwise, connectedness means that $|G|$ is connected. A root $o \in V$ may then be joined to every vertex by a path in the underlying graph. If G has several connected components, the theory applies to each component and

$$\Gamma(H) \cong \bigoplus_{\alpha} \Gamma(H|_{G_\alpha}).$$

3.3 Edge weights

For approximate theory, assign strictly positive edge weights

$$w_e > 0.$$

They may encode channel reliability, frequency, precision, experimental importance, or inverse noise variance. Exact compatibility is independent of positive rescaling of these weights, while approximate spectra are not.

Example 7 (Human–model–artifact graph)**.** The motivating three-node system is represented by

$$V = \{h, m, a\}, \qquad E = \{h \to m, m \to a, a \to h\}.$$

Additional edges may encode direct retrieval $a \to m$, human revision $h \to a$, or independent verification. The graph records permitted translations, not merely physical communication links.

4. Core–Residual Decomposition

Definition 8 (Admissible local decomposition)**.** For every vertex i, choose an orthogonal decomposition

$$X_i = C_i \oplus R_i,$$

with projectors

$$P_i : X_i \to C_i, \qquad Q_i = I - P_i : X_i \to R_i.$$

The subspace C_i is the candidate locally realized transferable core. The subspace R_i contains substrate-specific, context-specific, or otherwise noncore degrees of freedom relative to the present analysis.

The decomposition is part of the model and is not presumed unique. Calling a subspace "core" is therefore a hypothesis requiring structural and causal validation. Later null tests must determine whether the chosen projectors identify a genuine shared channel rather than an arbitrary low-dimensional summary.

Relative to these decompositions, every edge map has the block form

$$T_e = \begin{pmatrix} H_e & U_e \\ V_e & K_e \end{pmatrix},$$

where

$$\begin{aligned} H_e &= P_{t(e)} T_e P_{s(e)} : C_{s(e)} \to C_{t(e)}, \\ U_e &= P_{t(e)} T_e Q_{s(e)} : R_{s(e)} \to C_{t(e)}, \\ V_e &= Q_{t(e)} T_e P_{s(e)} : C_{s(e)} \to R_{t(e)}, \\ K_e &= Q_{t(e)} T_e Q_{s(e)} : R_{s(e)} \to R_{t(e)}. \end{aligned}$$

4.1 Structural conditions

Core preservation means

$$V_e = 0.$$

A core input cannot leak into the target residual.

Core autonomy means

$$U_e = V_e = 0.$$

Core and residual evolve as separate transfer systems.

Core leakage, **residual contamination**, and **total transfer defect** are measured by

$$\lambda_e^{\mathrm{out}} = \|V_e\|, \qquad \lambda_e^{\mathrm{in}} = \|U_e\|,$$

$$\delta_e^{\mathrm{tr}} = \|U_e\| + \|V_e\|.$$

These quantities depend on the chosen norms and projectors. Their empirical meaning must therefore be tied to a declared intervention and prediction protocol.

4.2 Composition exposes residual mediation

For composable edges $e : i \to j$ and $f : j \to k$, direct block multiplication gives

$$T_f T_e = \begin{pmatrix} H_f H_e + U_f V_e & H_f U_e + U_f K_e \\ V_f H_e + K_f V_e & V_f U_e + K_f K_e \end{pmatrix}.$$

The induced core-to-core block is therefore

$$P_k T_f T_e P_i = H_f H_e + U_f V_e.$$

The term $U_f V_e$ is a residual-mediated correction: information leaves the candidate core at the intermediate node and later re-enters it. Consequently, the simple path product $H_f H_e$ faithfully represents full core transport only under a preservation condition such as $V_e = 0$, or under an explicit model that includes these mediated corrections.

Proposition 9 (Pathwise core preservation)**.** *Suppose $V_e = 0$ for every edge in a directed path $\gamma = e_m \cdots e_1$. Then*

$$T_\gamma(C_{s(\gamma)}) \subseteq C_{t(\gamma)}$$

and

$$P_{t(\gamma)} T_\gamma P_{s(\gamma)} = H_{e_m} \cdots H_{e_1}.$$

Proof. For a single edge, $V_e = 0$ means

$$T_e(c, 0) = (H_e c, 0),$$

so core states remain in the core. Inductively, if the first r transfers map c to $(H_{e_r} \cdots H_{e_1} c, 0)$, applying $T_{e_{r+1}}$ gives

$$(H_{e_{r+1}} H_{e_r} \cdots H_{e_1} c, 0).$$

The conclusion follows after m steps. □

Corollary 10 (Autonomous splitting)**.** *If $U_e = V_e = 0$ for every edge, then the complete diagram decomposes as the direct sum of two diagrams,*

$$\mathcal{X} \cong H \oplus K,$$

and every path transfer satisfies

$$T_\gamma = \begin{pmatrix} H_\gamma & 0 \\ 0 & K_\gamma \end{pmatrix}.$$

4.3 Exact core-transfer systems

Definition 11 (Exact core-transfer system)**.** An exact core-transfer system on G consists of:

1. decompositions $X_i = C_i \oplus R_i$;
2. core-preserving edge maps $V_e = 0$; and
3. invertible core blocks $H_e \in GL(C_{s(e)}, C_{t(e)})$ whenever holonomy is invoked.

The requirement $V_e = 0$ concerns the relation between the selected core and the full transfer. The requirement that H_e be invertible concerns the ability to compare core representations along reversed paths. They answer different questions and should not be conflated.

Remark 12 (Approximate systems)**.** In applications neither condition need hold exactly. The approximate theory will replace strict invariance by low transfer energy, small leakage, and a spectral gap separating workspace modes from residual modes.

5. Abstract Operation Algebra

Definition 13 (Operation algebra)**.** Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. An operation algebra $\mathcal{A}$ is a unital associative $\mathbb{F}$-algebra whose elements represent abstract transformations, inferential rules, questions, or procedures that can be realized in multiple strata.

Every core space carries a unital representation

$$\rho_i : \mathcal{A} \longrightarrow \mathrm{End}(C_i).$$

For $a \in \mathcal{A}$, the operator $\rho_i(a)$ is the realization of the same abstract operation in local coordinates.

Definition 14 (Exact operation-compatible transfer)**.** A core transfer $H_e : C_i \to C_j$ is operation compatible when it is an intertwiner:

$$\boxed{H_e \rho_i(a) = \rho_j(a) H_e} \qquad \forall a \in \mathcal{A}.$$

Equivalently, the diagram

$$\begin{array}{ccc} C_i & \xrightarrow{\rho_i(a)} & C_i \\ \downarrow H_e & & \downarrow H_e \\ C_j & \xrightarrow{\rho_j(a)} & C_j \end{array}$$

commutes for every a.

This condition states that translating and then operating gives the same result as operating and then translating. Agreement of static encodings alone is weaker: two nodes may store corresponding vectors while implementing incompatible transformations on them.

Proposition 15 (Path composition of intertwiners)**.** *If every edge of a path $\gamma : i \to j$ intertwines the local $\mathcal{A}$-representations, then*

$$H_\gamma \rho_i(a) = \rho_j(a) H_\gamma \qquad \forall a \in \mathcal{A}.$$

Proof. For composable intertwiners $H_e : C_i \to C_j$ and $H_f : C_j \to C_k$,

$$H_f H_e \rho_i(a) = H_f \rho_j(a) H_e = \rho_k(a) H_f H_e.$$

Induction over path length proves the claim. □

Corollary 16 (Holonomy commutes with abstract operations)**.** *For every loop $\gamma : o \to o$,*

$$H_\gamma \rho_o(a) = \rho_o(a) H_\gamma.$$

Hence

$$\rho_o(\mathcal{A}) \subseteq \mathrm{Comm}(\mathrm{Hol}_o),$$

where $\mathrm{Comm}(\mathrm{Hol}_o)$ *is the commutant of the holonomy representation.*

5.1 Approximate operation compatibility

The edgewise defect is

$$\Xi_e(a) = H_e \rho_{s(e)}(a) - \rho_{t(e)}(a) H_e.$$

For a bounded generating family $\mathcal{G} \subseteq \mathcal{A}$, define

$$\eta_{\mathrm{int}}(\mathcal{G}) = \sup_{a \in \mathcal{G}} \sum_{e \in E} w_e \|\Xi_e(a)\|^2.$$

When $\mathcal{A}$ is normed and its unit ball is appropriate, one may take

$$\mathcal{G} = \{a : \|a\|_{\mathcal{A}} \leq 1\}.$$

Exact representational universality means

$$\eta_{\text{int}} = 0,$$

while approximate universality means

$$\eta_{\text{int}} \leq \varepsilon_{\text{int}}.$$

The use of a generating family prevents the definition from depending on an unbounded rescaling of algebra elements.

Example 17 (Equivalent operations in different coordinates)**.** Let $C_1 = C_2 = \mathbb{R}^2$, choose $S \in GL_2(\mathbb{R})$, and set

$$H_{12} = S, \qquad \rho_2(a) = S\rho_1(a)S^{-1}.$$

Then

$$H_{12}\rho_1(a) = S\rho_1(a) = \rho_2(a)S = \rho_2(a)H_{12}.$$

The local matrices can look different while realizing the same abstract operation. This is the elementary model of representation change used throughout the theory.

PART III

Exact Shared Cores

6. Global Compatible Sections

From this point through Theorem 22, assume the core maps define a local system on a connected graph unless a weaker hypothesis is stated explicitly.

Definition 18 (Global compatible section)**.** The exact structural workspace is

$$\boxed{\Gamma(H) = \left\{ (c_i)_{i\in V} \in \bigoplus_{i\in V} C_i : \; c_{t(e)} = H_e c_{s(e)} \; \forall e \in E \right\}.}$$

An element $c = (c_i) \in \Gamma(H)$ is a global compatible section.

Equivalently,

$$\Gamma(H) = \varprojlim_{\mathrm{Path}(G)} C_i.$$

A section is not a single vector copied verbatim at every vertex. It is one abstract element represented by a family of local vectors that agree under the specified translation maps.

A nontrivial exact structural workspace exists when

$$\dim \Gamma(H) > 0.$$

Its exact dimension is

$$d_{\mathbf{W}} = \dim \Gamma(H).$$

Fix a root o. The evaluation map is

$$\mathrm{ev}_o : \Gamma(H) \to C_o, \qquad \mathrm{ev}_o((c_i)) = c_o.$$

Lemma 19 (Root determination)**.** *If $|G|$ is connected and every core transfer is invertible, then ev_o is injective. Consequently,*

$$\dim \Gamma(H) \le \dim C_o.$$

Proof. Choose for each vertex i a path $\tau_i : o \to i$. Compatibility gives

$$c_i = H_{\tau_i} c_o.$$

Thus a section with $c_o = 0$ vanishes at every vertex. Hence ev_o is injective. □

Lemma 20 (Tree propagation)**.** *If $|G|$ is a tree and every H_e is invertible, then every $x \in C_o$ extends uniquely to a global compatible section.*

Proof. A tree contains a unique undirected path $\tau_i : o \to i$ for every vertex. Define $c_i = H_{\tau_i} x$. Because every edge belongs to these unique path relations, the family is compatible. Uniqueness follows from Lemma 19. □

The obstruction to a global core therefore appears only when distinct paths must agree. Cycles encode exactly those consistency constraints.

7. Holonomy and Path Obstruction

Fix a root $o \in V$. For every loop class

$$[\gamma] \in \pi_1(|G|, o),$$

define the holonomy

$$\mathrm{Hol}([\gamma]) = H_\gamma \in GL(C_o).$$

Functoriality gives a group representation

$$\mathrm{hol}_o : \pi_1(|G|, o) \longrightarrow GL(C_o).$$

Its image is the holonomy group

$$\mathrm{Hol}_o = \mathrm{im}(\mathrm{hol}_o).$$

The obstruction associated with a loop is

$$\Omega_\gamma = H_\gamma - I_{C_o}.$$

The maximal path-independent root core is

$$\boxed{C_* = \bigcap_{\gamma : o \to o} \ker(H_\gamma - I).}$$

Equivalently,

$$C_* = C_o^{\mathrm{Hol}_o} = \mathrm{Fix}(\mathrm{Hol}_o).$$

7.1 Spanning-tree reduction

Choose a spanning tree $\mathcal{T} \subseteq |G|$. Let

$$\tau_i : o \to i$$

be the unique tree path. Every non-tree edge $e : i \to j$ determines a fundamental loop

$$\gamma_e = \tau_j^{-1} e \tau_i$$

with transfer

$$H_{\gamma_e} = H_{\tau_j}^{-1} H_e H_{\tau_i}.$$

The fundamental group of a finite graph is generated freely by these cotree loops. Hence it is enough to impose fixed-point conditions for a finite set of generators.

Lemma 21 (Finite obstruction set)**.** *Let $e_1, \dots, e_r$ be the non-tree edges of a connected finite graph. Then*

$$C_* = \bigcap_{q=1}^{r} \ker(H_{\gamma_{e_q}} - I).$$

Proof. A vector fixed by every generator is fixed by each inverse generator and by every finite product of generators and inverses. Those products represent all elements of $\pi_1(|G|, o)$. The reverse inclusion is immediate. □

The number of independent graph cycles is

$$r = |E_{\mathrm{und}}| - |V| + 1$$

for a connected underlying graph, after treating parallel edges with their multiplicity.

8. Exact Existence Theorem

Theorem 22 (Limit–holonomy equivalence)**.** *Let $|G|$ be connected and let*

$$H : \Pi_1(G) \to \mathbf{FinHilb}^{\times}$$

be a finite-dimensional local system. For any root $o \in V$, evaluation at the root induces a canonical vector-space isomorphism

$$\boxed{\mathrm{ev}_o : \Gamma(H) \xrightarrow{\cong} C_o^{\mathrm{Hol}_o} = \bigcap_{\gamma : o \to o} \ker(H_\gamma - I).}$$

Consequently,

$$\Gamma(H) \neq \{0\}$$

if and only if the holonomy representation contains a nonzero trivial subrepresentation.

Proof. Let $c = (c_i) \in \Gamma(H)$. For any loop $\gamma : o \to o$, compatibility along the loop gives

$$H_\gamma c_o = c_o.$$

Therefore $\mathrm{ev}_o(c) = c_o \in C_o^{\mathrm{Hol}_o}$. By Lemma 19, ev_o is injective.

Conversely, let $x \in C_o^{\mathrm{Hol}_o}$. Choose a path $\tau_i : o \to i$ for each vertex and define

$$c_i = H_{\tau_i} x.$$

This definition is independent of the chosen path. Indeed, if $\sigma_i : o \to i$ is another path, then

$$\sigma_i^{-1}\tau_i : o \to o$$

is a loop, and invariance of x gives

$$H_{\sigma_i}^{-1} H_{\tau_i} x = x.$$

Multiplying by H_{σ_i} yields

$$H_{\tau_i} x = H_{\sigma_i} x.$$

For any edge $e : i \to j$, the paths $e\tau_i$ and τ_j both run from o to j. Path independence therefore gives

$$H_e c_i = H_e H_{\tau_i} x = H_{\tau_j} x = c_j.$$

Thus (c_i) is a global compatible section and $\mathrm{ev}_o((c_i)) = x$. Hence ev_o is surjective and therefore an isomorphism. □

Corollary 23 (Tree case)**.** *If $|G|$ is a tree, then* $\mathrm{Hol}_o = \{I\}$ *and*

$$\Gamma(H) \cong C_o.$$

In particular,

$$d_{\mathbf{W}} = \dim C_o.$$

Corollary 24 (Finite matrix test)**.** *Let $\gamma_1, \dots, \gamma_r$ be fundamental cycle generators and suppose* $\dim C_o = d$*. Define the stacked obstruction matrix*

$$\mathcal{M} = \begin{pmatrix} H_{\gamma_1} - I \\ \vdots \\ H_{\gamma_r} - I \end{pmatrix}.$$

Then

$$\boxed{C_* = \ker \mathcal{M}, \qquad d_{\mathbf{W}} = d - \operatorname{rank}(\mathcal{M}).}$$

Proof. The kernel of a vertically stacked operator is the intersection of the kernels of its blocks. Apply Lemma 21 and Theorem 22. □

Corollary 25 (Averaging projector for finite holonomy)**.** *If* Hol_o *is finite and acts unitarily, then*

$$P_* = \frac{1}{|\,\mathrm{Hol}_o\,|} \sum_{h \in \mathrm{Hol}_o} h$$

is the orthogonal projector onto C_*.

Proof. For every $g \in \mathrm{Hol}_o$, left multiplication permutes the group elements, so $gP_* = P_*$. Thus $\operatorname{ran} P_* \subseteq C_*$. If $x \in C_*$, then $P_* x = x$, so $C_* \subseteq \operatorname{ran} P_*$. Unitarity gives $P_*^* = P_*$, and the range characterization implies $P_*^2 = P_*$. □

8.1 Worked examples

Example 26 (One-dimensional core surviving a contradictory cycle)**.** Let a three-node cycle have $C_1 = C_2 = C_3 = \mathbb{R}^2$ and

$$H_{12} = I, \qquad H_{23} = I, \qquad H_{31} = J, \qquad J = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The cycle holonomy is J, hence

$$C_* = \ker(J - I) = \operatorname{span}\{(1, 0)^\top\}.$$

The first coordinate circulates consistently through all three strata. The second returns with its sign reversed and therefore cannot define a nonzero exact shared section. Thus

$$d_{\mathbf{W}} = 1.$$

Example 27 (Local compatibility without a global nonzero core)**.** Let the only nontrivial cycle holonomy on $\mathbb{R}^2$ be a rotation

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

with $\theta \notin 2\pi\mathbb{Z}$. Then

$$\ker(R_\theta - I) = \{0\}.$$

Every edge map is individually invertible and well-defined, yet the complete cycle admits no nonzero exact shared core. Pairwise translatability is therefore insufficient for global path independence.

Example 28 (Acyclic collaboration)**.** Suppose a root concept is translated into a proof, implementation, and report along a tree of invertible exact transfers. No alternative route returns a conflicting representation. The tree has trivial holonomy, so every root-core direction extends uniquely. This does not yet establish causal efficacy; it establishes only exact structural compatibility.

8.2 Interpretive consequence

The theorem gives a precise meaning to "something shared across representations." The shared object is neither any one local vector nor an extra vector outside the system. It is the equivalence class of all compatible local realizations, canonically represented at a root by the subspace fixed by every translation cycle.

9. Operator Closure Theorem

Theorem 29 (Global operation closure)**.** *Suppose every core transfer intertwines the representations of $\mathcal{A}$:*

$$H_e \rho_{s(e)}(a) = \rho_{t(e)}(a) H_e \qquad \forall e \in E,\ a \in \mathcal{A}.$$

Then $\Gamma(H)$ is an $\mathcal{A}$-module under the componentwise action

$$\rho_\Gamma(a)(c_i)_{i\in V} = (\rho_i(a)c_i)_{i\in V}.$$

Equivalently, the holonomy-fixed root core is invariant:

$$\boxed{\rho_o(a) C_* \subseteq C_* \qquad \forall a \in \mathcal{A}.}$$

Proof. Let $c = (c_i) \in \Gamma(H)$. For every edge $e : i \to j$,

$$H_e \rho_i(a) c_i = \rho_j(a) H_e c_i = \rho_j(a) c_j.$$

Therefore the family $(\rho_i(a)c_i)$ is compatible and belongs to $\Gamma(H)$. The representation identities on each C_i hold componentwise, so ρ_Γ is an $\mathcal{A}$-representation.

Under the isomorphism of Theorem 22, this action is represented at the root by ρ_o. Alternatively, if $x \in C_*$, then Corollary 16 gives, for every loop γ,

$$H_\gamma \rho_o(a) x = \rho_o(a) H_\gamma x = \rho_o(a) x.$$

Hence $\rho_o(a)x \in C_*$. □

Corollary 30 (Inherited algebraic structure)**.** *Any algebraic relation satisfied by $\rho_i(\mathcal{A})$ is inherited by the global action. In particular, if $a, b \in \mathcal{A}$, then*

$$\rho_\Gamma(ab) = \rho_\Gamma(a)\rho_\Gamma(b), \qquad \rho_\Gamma(1) = I_{\Gamma(H)}.$$

The result distinguishes a shared *operational* representation from a shared static code. A core closed under the common algebra supports the same abstract manipulations across strata without leaving the globally compatible subspace.

10. Workspace Operator Algebra

Definition 31 (Natural workspace operators)**.** Define

$$\boxed{\mathcal{O}_{\mathbf{W}} = \left\{(A_i)_{i\in V} : A_{t(e)}H_e = H_e A_{s(e)} \ \forall e\right\}.}$$

Categorically,

$$\mathcal{O}_{\mathbf{W}} = \mathrm{Nat}(H, H),$$

the algebra of natural endomorphisms of the local system.

Pointwise addition, scalar multiplication, and composition make $\mathcal{O}_{\mathbf{W}}$ a unital associative algebra. Each $A = (A_i)$ acts on global sections by

$$\pi_{\mathbf{W}}(A)(c_i) = (A_i c_i).$$

Naturality guarantees that $(A_i c_i)$ is again compatible.

Theorem 32 (Commutant characterization)**.** *Let $|G|$ be connected and let H be an invertible local system. Root evaluation induces an algebra isomorphism*

$$\boxed{\Phi_o : \mathcal{O}_{\mathbf{W}} \xrightarrow{\cong} \mathrm{Comm}(\mathrm{Hol}_o), \qquad \Phi_o((A_i)) = A_o,}$$

where

$$\mathrm{Comm}(\mathrm{Hol}_o) = \{B \in \mathrm{End}(C_o) : Bh = hB \ \forall h \in \mathrm{Hol}_o\}.$$

Proof. If $(A_i) \in \mathcal{O}_{\mathbf{W}}$, naturality along every loop gives

$$A_o H_\gamma = H_\gamma A_o.$$

Thus $A_o \in \mathrm{Comm}(\mathrm{Hol}_o)$, and Φ_o is a well-defined algebra homomorphism.

Choose paths $\tau_i : o \to i$. Naturality along τ_i gives

$$A_i H_{\tau_i} = H_{\tau_i} A_o,$$

so

$$A_i = H_{\tau_i} A_o H_{\tau_i}^{-1}.$$

Therefore A_o determines every A_i, and Φ_o is injective.

Conversely, take $B \in \mathrm{Comm}(\mathrm{Hol}_o)$ and define

$$A_i = H_{\tau_i} B H_{\tau_i}^{-1}.$$

This is independent of the chosen root path: changing τ_i multiplies it by a loop holonomy that commutes with B. For an edge $e : i \to j$, the operator

$$H_{\tau_j}^{-1} H_e H_{\tau_i}$$

is a root holonomy. Commutation with B rearranges to

$$A_j H_e = H_e A_i.$$

Hence (A_i) is natural and $\Phi_o((A_i)) = B$. Thus Φ_o is surjective. □

The structural workspace is therefore more accurately represented by the triple

$$\boxed{\mathbf{W}_{\text{static}} = (\Gamma(H), \mathcal{O}_{\mathbf{W}}, \pi_{\mathbf{W}}),}$$

not merely by the vector space $\Gamma(H)$.

10.1 Effective operator algebra

Some natural operators may vanish on every global section while acting nontrivially on locally available but globally inconsistent directions. Define the annihilator

$$\operatorname{Ann}(\Gamma(H)) = \{A \in \mathcal{O}_{\mathbf{W}} : \pi_{\mathbf{W}}(A) = 0\}.$$

The effective workspace algebra is

$$\boxed{\mathcal{O}_{\mathbf{W}}^{\text{eff}} = \mathcal{O}_{\mathbf{W}} / \operatorname{Ann}(\Gamma(H)).}$$

Its action on $\Gamma(H)$ is faithful by construction.

Example 33 (Operators in the reflection-cycle workspace)**.** For Example 26, the holonomy is

$$J = \operatorname{diag}(1, -1).$$

Its commutant consists of diagonal matrices:

$$\operatorname{Comm}(J) = \left\{ \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} : a, b \in \mathbb{R} \right\}.$$

The exact workspace core is the first coordinate. Consequently, only the parameter a acts effectively on $\Gamma(H)$; the b-direction belongs to the annihilator of the global core. Hence

$$\mathcal{O}_{\mathbf{W}}^{\text{eff}} \cong \mathbb{R}.$$

10.2 Scope of the exact theory

Parts I–III now establish three exact facts:

global shared content is the holonomy-fixed core,
universally transported operations preserve that core,
natural workspace operations are the commutant of holonomy.

These statements do not yet prove that the shared core is robust, causally used, irreducibly distributed, or temporally persistent. Those are separate questions developed in the approximate and dynamic theory that follows.

PART IV

Approximate Workspaces

The exact theory asks whether local core states agree perfectly under every transfer. Approximate theory asks three more realistic questions:

1. how strongly a candidate state violates edgewise compatibility;
2. whether a low-defect subspace is spectrally separated from incoherent directions; and
3. how that subspace changes when transfers, metrics, or operations are perturbed.

Throughout Parts IV–V, set

$$\mathcal{C}_V = \bigoplus_{i \in V} C_i, \qquad \mathcal{C}_E = \bigoplus_{e \in E} C_{t(e)},$$

with their Hilbert direct-sum inner products, and retain strictly positive weights w_e.

11. Transfer Discrepancy Operator

Definition 34 (Transfer discrepancy)**.** The transfer discrepancy operator is

$$D_H : \mathcal{C}_V \longrightarrow \mathcal{C}_E, \qquad (D_H c)_e = \sqrt{w_e}\,(c_{t(e)} - H_e c_{s(e)}).$$

Its squared norm is the coherence energy

$$\boxed{\mathcal{E}_H(c) = \|D_H c\|^2 = \sum_{e \in E} w_e \left\| c_{t(e)} - H_e c_{s(e)} \right\|^2.}$$

The transfer Laplacian is

$$\boxed{L_H = D_H^* D_H.}$$

The operator D_H is an incidence-type derivative twisted by the edge transfers. It compares a state already present at the target with the state predicted there by translation from the source. A direction can therefore have low energy even when its coordinates differ substantially across nodes, provided those coordinates differ exactly as the transfer maps prescribe.

For $z = (z_e)_{e \in E} \in \mathcal{C}_E$, the adjoint is

$$(D_H^* z)_i = \sum_{e : t(e) = i} \sqrt{w_e}\, z_e - \sum_{e : s(e) = i} \sqrt{w_e}\, H_e^* z_e.$$

Consequently,

$$\boxed{\begin{aligned}(L_H c)_i = &\sum_{e : t(e) = i} w_e (c_i - H_e c_{s(e)}) \\ &+ \sum_{e : s(e) = i} w_e H_e^* (H_e c_i - c_{t(e)}).\end{aligned}}$$

This formula makes the two contributions at a vertex explicit: incoming prediction error and adjoint-propagated outgoing error.

Proposition 35 (Positivity and variational identity)**.** *The transfer Laplacian is self-adjoint and positive semidefinite. For every $c \in \mathcal{C}_V$,*

$$\langle c, L_H c\rangle = \mathcal{E}_H(c) \geq 0.$$

Proof. Self-adjointness and positivity follow from $L_H = D_H^* D_H$. The displayed identity is

$$\langle c, D_H^* D_H c\rangle = \langle D_H c, D_H c\rangle.$$

□

Remark 36 (Weight dependence)**.** Replacing w_e by another positive family does not alter $\ker D_H$, but it generally changes every positive eigenvalue and therefore changes the approximate geometry. Edge weights are part of the metric model, not part of the exact compatibility relation.

12. Exact Kernel Theorem

Theorem 37 (Transfer-Laplacian kernel theorem)**.** *For an arbitrary finite directed core diagram, without assuming invertibility,*

$$\boxed{\ker L_H = \ker D_H = \Gamma(H).}$$

Thus the exact compatible sections are precisely the zero-energy modes of the transfer Laplacian.

Proof. If $D_H c = 0$, then $L_H c = D_H^* D_H c = 0$. Conversely, if $L_H c = 0$, then

$$0 = \langle c, L_H c\rangle = \|D_H c\|^2,$$

so $D_H c = 0$. Finally, $D_H c = 0$ if and only if

$$c_{t(e)} = H_e c_{s(e)}$$

for every edge, which is exactly the definition of $\Gamma(H)$. □

Corollary 38 (Finite rank computation)**.** *Let*

$$n_C = \sum_{i \in V} \dim C_i.$$

Then

$$\boxed{\dim \Gamma(H) = n_C - \operatorname{rank}(D_H).}$$

Hence the exact workspace dimension can be obtained by a single rank or singular-value computation, even when holonomy is unavailable because some transfers are noninvertible.

12.1 Relation to the holonomy computation

In the local-system regime, Theorems 22 and 37 give two equivalent constructions:

$$\Gamma(H) \cong C_o^{\mathrm{Hol}_o} \qquad \text{and} \qquad \Gamma(H) = \ker L_H.$$

The first reduces the problem to cycle invariants in a root fiber. The second works directly on all vertex coordinates and extends to noninvertible diagrams. The two computations therefore serve different practical regimes rather than competing definitions.

13. Approximate Structural Workspace

Let

$$0 \le \lambda_1 \le \lambda_2 \le \cdots \le \lambda_N, \qquad N = \dim \mathcal{C}_V,$$

be the eigenvalues of L_H, repeated with multiplicity, and let $P_I(L_H)$ denote the spectral projector associated with an interval I.

Definition 39 (Threshold workspace)**.** For $\varepsilon \ge 0$, define

$$\boxed{\Gamma_\varepsilon(H) = \operatorname{ran} P_{[0,\varepsilon]}(L_H).}$$

Its dimension is

$$d_\varepsilon = \#\{k : \lambda_k \le \varepsilon\}.$$

When $0 < d_\varepsilon < N$, define the boundary gap

$$\boxed{g_\varepsilon = \lambda_{d_\varepsilon+1} - \lambda_{d_\varepsilon}.}$$

The threshold ε declares which discrepancy scale is treated as structurally negligible. The dimension d_ε alone is not a robust certificate: a cluster is stable only when it is separated by a positive boundary gap.

An equivalent index-based description fixes d and writes

$$W_d = \operatorname{ran} P_d, \qquad P_d = P_{\{\lambda_1,\dots,\lambda_d\}}(L_H),$$

with cluster gap

$$g_d = \lambda_{d+1} - \lambda_d > 0.$$

The index-based form is convenient for perturbation theorems because the workspace rank is declared explicitly.

Proposition 40 (Min–max characterization)**.** *For $1 \le k \le N$,*

$$\lambda_k = \min_{\substack{S \subseteq \mathcal{C}_V \\ \dim S = k}} \max_{\substack{c \in S \\ \|c\|=1}} \mathcal{E}_H(c).$$

Consequently, the span of the first k eigenvectors minimizes the worst-case coherence energy among all k-dimensional candidate subspaces.

Proof. Expand every vector in an orthonormal eigenbasis of L_H. The span of the first k eigenvectors has maximal Rayleigh quotient λ_k. Every k-dimensional subspace intersects the orthogonal complement of the first $k-1$ eigenvectors nontrivially, and any unit vector in that intersection has Rayleigh quotient at least λ_k. □

13.1 Exact, soft, and spurious low modes

The spectrum supports three distinct interpretations:

$$\begin{aligned} \lambda_k = 0 &\iff \text{exact compatible direction,} \\ 0 < \lambda_k \ll g_d &\implies \text{approximately transported direction,} \\ \lambda_k \approx \lambda_{k+1} &\implies \text{rank selection is unstable at that boundary.} \end{aligned}$$

A low eigenvalue can still be scientifically spurious if it arises from a poorly chosen weight, a nearly disconnected graph, or a noncausal encoding. The causal and null gates in later parts remain necessary.

14. Normalized Coherence Defect

For $c \neq 0$, define the Rayleigh defect

$$\delta_{\mathrm{coh}}(c) = \frac{\mathcal{E}_H(c)}{\|c\|^2}.$$

For a nonzero subspace $W \subseteq \mathcal{C}_V$, define

$$\delta_{\mathrm{coh}}(W) = \sup_{\substack{c \in W \\ \|c\|=1}} \mathcal{E}_H(c).$$

If W is L_H-invariant, then

$$\delta_{\mathrm{coh}}(W) = \|L_H|_W\|.$$

In particular,

$$\delta_{\mathrm{coh}}(\Gamma_\varepsilon(H)) \leq \varepsilon.$$

Theorem 41 (Coherence-to-core distance estimate)**.** *Let*

$$W_0 = \Gamma(H) = \ker L_H, \qquad r = \dim W_0,$$

and suppose $r < N$. Set

$$\gamma_H = \lambda_{r+1} > 0.$$

Then every $c \in \mathcal{C}_V$ satisfies

$$\boxed{\operatorname{dist}(c, W_0)^2 \leq \frac{\mathcal{E}_H(c)}{\gamma_H}.}$$

Moreover,

$$\mathcal{E}_H(c) \leq \|L_H\| \operatorname{dist}(c, W_0)^2.$$

Proof. Write $c = P_0 c + Q_0 c$, where P_0 projects onto W_0 and $Q_0 = I - P_0$. Since $L_H P_0 = 0$,

$$\mathcal{E}_H(c) = \langle Q_0 c, L_H Q_0 c \rangle.$$

The spectrum of L_H on $W_0^\perp$ lies in $[\gamma_H, \|L_H\|]$. Hence

$$\gamma_H \|Q_0 c\|^2 \leq \mathcal{E}_H(c) \leq \|L_H\| \|Q_0 c\|^2.$$

Finally, $\|Q_0 c\| = \operatorname{dist}(c, W_0)$. □

This estimate is the exact sense in which a spectral gap turns small edgewise disagreement into closeness to a genuinely compatible global section.

14.1 Two-node consistency cycle

Let $C_1 = C_2 = \mathbb{R}$, use unit weights, and take two directed transfers

$$H_{12} = a, \qquad H_{21} = b.$$

Then

$$D_H = \begin{pmatrix} -a & 1 \\ 1 & -b \end{pmatrix}, \qquad L_H = \begin{pmatrix} a^2+1 & -(a+b) \\ -(a+b) & b^2+1 \end{pmatrix}.$$

Writing

$$\tau = a^2 + b^2 + 2,$$

the eigenvalues are

$$\lambda_\pm = \frac{1}{2}\left(\tau \pm \sqrt{\tau^2 - 4(ab-1)^2}\right),$$

and

$$\det L_H = (ab-1)^2.$$

Thus a nonzero exact shared core exists precisely when

$$ab = 1,$$

which is exactly the trivial-holonomy condition. Away from exact consistency,

$$\lambda_- \lambda_+ = (ab-1)^2.$$

When $|ab-1|$ is small while λ_+ stays bounded away from zero, the soft workspace eigenvalue is quadratic in the cycle mismatch:

$$\lambda_- = \frac{(ab-1)^2}{\lambda_+}.$$

This example connects approximate spectral coherence directly to approximate holonomy closure.

15. Perturbation Stability Theorem

The central robustness question is whether a selected low-energy subspace persists when the estimated transfer maps change. The next lemma isolates the finite-dimensional spectral fact needed for the answer.

Lemma 42 (Separated spectral subspaces)**.** *Let A and $B = A + E$ be self-adjoint operators. Let P be an A-spectral projector whose eigenvalues lie in $(-\infty, a]$, and let $\widetilde{Q}$ be a B-spectral projector whose eigenvalues lie in $[b, \infty)$, with $b > a$. Then*

$$\|P\widetilde{Q}\| \le \frac{\|E\|}{b-a}.$$

Proof. On $\operatorname{ran} P$ and $\operatorname{ran} \widetilde{Q}$, set

$$A_P = PAP|_{\operatorname{ran} P}, \qquad B_Q = \widetilde{Q}B\widetilde{Q}|_{\operatorname{ran} \widetilde{Q}}.$$

The operator $X = P\widetilde{Q}$ satisfies the Sylvester equation

$$A_P X - X B_Q = -PE\widetilde{Q}.$$

Let $\mu = (a+b)/2$. Since

$$\|e^{t(A_P - \mu I)}\| \le e^{-(b-a)t/2}, \qquad \|e^{-t(B_Q - \mu I)}\| \le e^{-(b-a)t/2},$$

the inverse Sylvester operator has the convergent integral representation

$$X = \int_0^\infty e^{t(A_P - \mu I)}(PE\widetilde{Q})e^{-t(B_Q - \mu I)}\, \mathrm{d}t.$$

Taking norms and integrating gives the result. □

Theorem 43 (Spectral workspace stability)**.** *Let L and $\widetilde{L} = L+E$ be self-adjoint positive semidefinite operators on the same N-dimensional space. Suppose*

$$\lambda_d(L) < \lambda_{d+1}(L), \qquad g_d = \lambda_{d+1}(L) - \lambda_d(L) > 0,$$

and

$$\eta = \|E\| < \frac{g_d}{2}.$$

Let P_d and $\widetilde{P}_d$ project onto the first d eigenvectors of L and $\widetilde{L}$, respectively. Then the rank-d cluster remains separated and

$$\boxed{\|P_d - \widetilde{P}_d\| \le \frac{\eta}{g_d - \eta} \le \frac{2\eta}{g_d}.}$$

Proof. The min–max characterization gives the eigenvalue perturbation bound

$$|\lambda_k(\widetilde{L}) - \lambda_k(L)| \le \eta.$$

Hence

$$\lambda_{d+1}(\widetilde{L}) - \lambda_d(L) \ge g_d - \eta$$

and

$$\lambda_{d+1}(L) - \lambda_d(\widetilde{L}) \ge g_d - \eta.$$

Apply Lemma 42 first to $P_d(I - \widetilde{P}_d)$ and then, with the roles of L and $\widetilde{L}$ exchanged, to $(I - P_d)\widetilde{P}_d$. Orthogonal projections of equal rank satisfy

$$\|P_d - \widetilde{P}_d\| = \max\left\{\|P_d(I - \widetilde{P}_d)\|, \|(I - P_d)\widetilde{P}_d\|\right\}.$$

This yields the first bound. The second follows from $\eta < g_d/2$. □

15.1 From transfer-map error to Laplacian error

Suppose the edge maps are perturbed to

$$\widetilde{H}_e = H_e + \Delta_e$$

while the weights and vertex metrics remain fixed. Define

$$\boxed{\delta_H^2 = \max_{i \in V} \sum_{e:s(e)=i} w_e \|\Delta_e\|^2.}$$

Let $F = D_{\widetilde{H}} - D_H$. Then

$$(Fc)_e = -\sqrt{w_e}\,\Delta_e c_{s(e)}, \qquad \|F\| \le \delta_H.$$

Since

$$L_{\widetilde{H}} - L_H = D_H^* F + F^* D_H + F^* F,$$

we obtain

$$\boxed{\|L_{\widetilde{H}} - L_H\| \le 2\|D_H\|\delta_H + \delta_H^2.}$$

Corollary 44 (Edgewise stability certificate)**.** *If*

$$\eta_H = 2\|D_H\|\delta_H + \delta_H^2 < \frac{g_d}{2},$$

then

$$\boxed{\|P_d(H) - P_d(\widetilde{H})\| \le \frac{\eta_H}{g_d - \eta_H}.}$$

This converts declared edge-map uncertainty into an explicit bound on the motion of the extracted workspace subspace.

15.2 Approximate operation closure

Define the vertex and edge direct-sum representations

$$\rho_V(a) = \bigoplus_{i\in V} \rho_i(a), \qquad \rho_E(a) = \bigoplus_{e\in E} \rho_{t(e)}(a).$$

Lift the edgewise intertwining defects to

$$(\mathfrak{X}_a c)_e = \sqrt{w_e}\, \Xi_e(a) c_{s(e)}.$$

Then the exact commutator identity is

$$\boxed{D_H \rho_V(a) - \rho_E(a) D_H = -\mathfrak{X}_a.}$$

Furthermore,

$$\|\mathfrak{X}_a\| \le \beta(a), \qquad \beta(a)^2 = \max_{i\in V} \sum_{e:s(e)=i} w_e \|\Xi_e(a)\|^2.$$

Theorem 45 (Approximate operation-closure theorem)**.** *Let P_d project onto the first d eigenvectors of L_H, set*

$$\lambda_d \le \varepsilon, \qquad \mu = \lambda_{d+1} > 0,$$

and let $Q_d = I - P_d$. Then every $a \in \mathcal{A}$ satisfies

$$\boxed{\|Q_d \rho_V(a) P_d\| \le \frac{\|\rho_E(a)\|\sqrt{\varepsilon} + \beta(a)}{\sqrt{\mu}}.}$$

Proof. For $x \in \operatorname{ran} P_d$,

$$\|D_H x\|^2 = \langle x, L_H x\rangle \le \varepsilon \|x\|^2.$$

The commutator identity gives

$$\|D_H \rho_V(a) x\| \le \|\rho_E(a)\|\sqrt{\varepsilon}\|x\| + \beta(a)\|x\|.$$

Because P_d and Q_d reduce L_H, the two energy components are orthogonal and

$$\|D_H Q_d y\| \le \|D_H y\|.$$

On $\operatorname{ran} Q_d$, the spectrum of L_H is bounded below by μ, so

$$\sqrt{\mu}\|Q_d y\| \le \|D_H Q_d y\|.$$

Apply these inequalities to $y = \rho_V(a)x$, then take the supremum over unit $x \in \operatorname{ran} P_d$. □

Corollary 46 (Recovery of exact operation closure)**.** *If $d = \dim\ker L_H$, $\varepsilon = 0$, and every edge intertwines a exactly, then*

$$Q_d \rho_V(a) P_d = 0.$$

Thus Theorem 45 reduces to the exact operation-closure result of Part III.

Corollary 47 (Closure under projector drift)**.** *Let $\widetilde{P}_d$ be another rank-d orthogonal projector and set*

$$\delta_P = \|P_d - \widetilde{P}_d\|.$$

Then

$$\boxed{\|(I - \widetilde{P}_d)\rho_V(a)\widetilde{P}_d\| \le \|Q_d \rho_V(a) P_d\| + 2\|\rho_V(a)\|\delta_P.}$$

Combining this estimate with Theorem 43 controls operation leakage after noisy workspace extraction.

Proof. Subtract $Q_d \rho_V(a) P_d$ and use

$$I - \widetilde{P}_d - Q_d = P_d - \widetilde{P}_d.$$

The two resulting perturbation terms are each bounded by $\|\rho_V(a)\|\delta_P$. □

15.3 Core–residual path accumulation

Let a path have vertices

$$i_0 \xrightarrow{e_1} i_1 \xrightarrow{e_2} \cdots \xrightarrow{e_m} i_m.$$

Write

$$T_r = T_{e_r}, \quad H_r = H_{e_r}, \quad V_r = V_{e_r},$$

and let

$$\iota_i^C : C_i \hookrightarrow X_i, \qquad \iota_i^R : R_i \hookrightarrow X_i$$

be the canonical inclusions. Empty products below are identities.

Theorem 48 (Pathwise core–residual Duhamel formula)**.** *The deviation of full path transport from pure core transport is exactly*

$$\boxed{\begin{aligned} &T_m \cdots T_1 \iota_{i_0}^C - \iota_{i_m}^C H_m \cdots H_1 \\ &\qquad = \sum_{r=1}^{m} T_m \cdots T_{r+1} \iota_{i_r}^R V_r H_{r-1} \cdots H_1. \end{aligned}}$$

Consequently,

$$\boxed{\begin{aligned} \|Q_{i_m} T_m \cdots T_1 P_{i_0}\| \le \sum_{r=1}^{m} &\left(\prod_{\ell=r+1}^{m} \|T_\ell\| \right) \|V_r\| \\ &\times \left(\prod_{\ell=1}^{r-1} \|H_\ell\| \right), \end{aligned}}$$

and the same right-hand side bounds

$$\|P_{i_m} T_m \cdots T_1 P_{i_0} - H_m \cdots H_1\|.$$

Proof. For every step,

$$T_r \iota^C_{i_{r-1}} = \iota^C_{i_r} H_r + \iota^R_{i_r} V_r.$$

Telescope the sequence

$$T_m \cdots T_{r+1} \iota^C_{i_r} H_r \cdots H_1, \qquad r = 0, 1, \ldots, m.$$

The difference between consecutive terms is exactly the r-th summand in the displayed formula. Projection onto R_{i_m} or C_{i_m}, followed by submultiplicativity of the operator norm, gives the two bounds. □

Corollary 49 (Uniform path bound)**.** *If along the path*

$$\|T_r\| \le \tau, \qquad \|H_r\| \le h, \qquad \|V_r\| \le \nu,$$

then the leakage and core-transport defects are bounded by

$$\boxed{\nu \sum_{r=1}^{m} \tau^{m-r} h^{r-1} = \begin{cases} \nu \dfrac{\tau^m - h^m}{\tau - h}, & \tau \neq h, \\ m \nu \tau^{m-1}, & \tau = h. \end{cases}}$$

In particular, if $\tau \le 1$ and $h \le 1$, the bound is at most $m\nu$.

The linear factor m in the merely nonexpansive regime is important. Small single-edge leakage does not by itself justify arbitrarily long transfer chains. Long-horizon stability requires contraction, cancellation, error correction, or leakage that decreases with path length.

15.4 Residual-mediated reentry

For two consecutive transfers,

$$P_{i_2} T_2 T_1 P_{i_0} = H_2 H_1 + U_2 V_1,$$
$$Q_{i_2} T_2 T_1 P_{i_0} = V_2 H_1 + K_2 V_1.$$

Thus leakage out of the core is first order in the blocks V_r, while an excursion that leaves and then returns to the core contains at least one V-block and one later U-block. For a fixed path length, if

$$\|U_r\|, \|V_r\| = O(\delta)$$

and the diagonal blocks remain bounded, the residual-mediated correction to pure core-to-core transport is $O(\delta^2)$, whereas terminal residual leakage can be $O(\delta)$. This distinction separates loss of core invariance from distortion caused by temporary excursions through local residual degrees of freedom.

PART V

Gauge Invariance and Abstract Identity

Exact compatibility is a coordinate-free notion. Approximate spectra also depend on the Hilbert metrics used to measure discrepancy. Gauge theory must therefore distinguish a change of coordinates from a physical change of metric.

16. Local Coordinate Transformations

For each $i \in V$, let

$$g_i \in GL(C_i), \qquad g = (g_i)_{i \in V}.$$

Define the vertex and edge direct sums

$$G_V = \bigoplus_{i \in V} g_i, \qquad G_E = \bigoplus_{e \in E} g_{t(e)}.$$

The transformed transfers and local operation representations are

$$H_e^g = g_{t(e)} H_e g_{s(e)}^{-1},$$

$$\rho_i^g(a) = g_i \rho_i(a) g_i^{-1}.$$

The fundamental covariance identity is

$$\boxed{D_{H^g} G_V = G_E D_H.}$$

It follows directly by evaluating the e-component:

$$\sqrt{w_e}\left(g_{t(e)} c_{t(e)} - g_{t(e)} H_e c_{s(e)}\right).$$

17. Gauge Invariance Theorem

Theorem 50 (Exact gauge covariance)**.** *The map*

$$G_V : \Gamma(H) \longrightarrow \Gamma(H^g)$$

is an isomorphism. Therefore,

$$\boxed{\Gamma(H^g) = G_V \Gamma(H), \qquad \dim \Gamma(H^g) = \dim \Gamma(H).}$$

If the core transports are invertible and o is a root, then

$$H_\gamma^g = g_o H_\gamma g_o^{-1}$$

for every loop $\gamma : o \to o$. Hence the holonomy-fixed core transforms by

$$C_*^g = g_o C_*.$$

Proof. The covariance identity implies

$$D_{H^g}(G_V c) = G_E D_H c.$$

Since G_E is invertible,

$$D_H c = 0 \quad \Longleftrightarrow \quad D_{H^g}(G_V c) = 0.$$

The holonomy formula follows by cancellation of adjacent factors $g_i^{-1} g_i$ along a path. □

Theorem 51 (Unitary gauge covariance of the Laplacian)**.** *If every g_i is unitary, then G_V and G_E are unitary and*

$$\boxed{L_{H^g} = G_V L_H G_V^{-1}.}$$

Consequently,

$$\operatorname{spec}(L_{H^g}) = \operatorname{spec}(L_H), \qquad P_I(L_{H^g}) = G_V P_I(L_H) G_V^{-1}$$

for every spectral interval I.

Proof. From

$$D_{H^g} = G_E D_H G_V^{-1}$$

and unitarity,

$$D_{H^g}^* D_{H^g} = G_V D_H^* G_E^{-1} G_E D_H G_V^{-1} = G_V L_H G_V^{-1}.$$

The spectral statements follow from unitary similarity. □

17.1 Nonunitary gauges and metric distortion

For a general invertible gauge, define

$$m_V = \sigma_{\min}(G_V), \qquad M_V = \|G_V\|,$$

$$m_E = \sigma_{\min}(G_E), \qquad M_E = \|G_E\|.$$

The covariance identity yields

$$m_E^2 \mathcal{E}_H(c) \le \mathcal{E}_{H^g}(G_V c) \le M_E^2 \mathcal{E}_H(c),$$

and

$$m_V^2 \|c\|^2 \le \|G_V c\|^2 \le M_V^2 \|c\|^2.$$

Therefore the Rayleigh quotients obey

$$\boxed{\frac{m_E^2}{M_V^2} \frac{\mathcal{E}_H(c)}{\|c\|^2} \le \frac{\mathcal{E}_{H^g}(G_V c)}{\|G_V c\|^2} \le \frac{M_E^2}{m_V^2} \frac{\mathcal{E}_H(c)}{\|c\|^2}.}$$

By the min–max principle,

$$\boxed{\frac{m_E^2}{M_V^2} \lambda_k(L_H) \le \lambda_k(L_{H^g}) \le \frac{M_E^2}{m_V^2} \lambda_k(L_H).}$$

Exact zero modes are unchanged, but fixed-metric approximate eigenvalues can be strongly distorted by an ill-conditioned coordinate transformation.

17.2 Metric-covariant formulation

A pure change of coordinates should transform the metric as well as the matrices. On each transformed space, define

$$\langle u, v\rangle_i^g = \left\langle g_i^{-1}u, g_i^{-1}v\right\rangle_i .$$

Use the analogous pullback metric on every edge copy. Then G_V and G_E are unitary between the original and transformed metric spaces. Consequently, the complete transfer-Laplacian spectrum is exactly preserved.

This gives two distinct interpretations:

- Transforming the matrices and the metrics gives the same approximate workspace in new coordinates.
- Transforming the matrices while holding the metrics fixed gives a potentially different weighting of discrepancy.

The distinction prevents arbitrary coordinate scaling from manufacturing or erasing apparent low-energy workspace directions.

18. Static Workspace Equivalence

Definition 52 (Exact static equivalence)**.** Two static realizations are exactly gauge equivalent,

$$\mathfrak{R} \sim_{\text{stat}} \mathfrak{R}',$$

when there exists an invertible family g_i such that

$$H'_e = g_{t(e)} H_e g_{s(e)}^{-1}, \qquad \rho'_i(a) = g_i \rho_i(a) g_i^{-1}.$$

The exact static abstract workspace is

$$\boxed{\mathbf{W}_{\text{stat}} = [\mathfrak{R}]_{\sim_{\text{stat}}}.}$$

For approximate theory, the realization also contains the vertex and edge metrics. Two metric realizations are equivalent only when the gauge transports those metrics, or when the gauge is unitary relative to the fixed metrics. Accordingly:

exact workspace identity is invariant under GL-gauge,
fixed-metric spectral geometry is invariant under unitary gauge,
metric-covariant spectral geometry is invariant under transported GL-gauge.

18.1 What the approximate theory establishes

Parts IV–V now prove:

$\ker L_H = \Gamma(H)$,
low energy plus a gap implies proximity to an exact core,
a separated workspace cluster is stable under transfer noise,
approximately intertwined operations nearly preserve that cluster,
core leakage and residual reentry admit explicit pathwise bounds,
exact identity is gauge invariant while approximate geometry is metric aware.

These are structural and perturbative results. They still do not prove that a workspace variable is causally used, predictively sufficient, irreducibly distributed, or temporally self-maintaining. Those questions begin in Part VI.

18.2 Finite computation protocol

For a declared graph, set of core dimensions, weights, and transfer maps, the static analysis can be executed in the following order:

1. Assemble the sparse block operator D_H from the edge equations
$$\sqrt{w_e}(c_{t(e)} - H_e c_{s(e)}).$$
2. Compute a singular-value decomposition
$$D_H = U\Sigma V^*, \qquad \lambda_k(L_H) = \sigma_k(D_H)^2.$$
Use a declared numerical tolerance to distinguish computational zero modes from positive modes.
3. Select a candidate rank d only at a reproducible spectral boundary with $g_d > 0$. A large observed gap is evidence of stability, not evidence of causal relevance.
4. Quantify transfer uncertainty by δ_H and verify whether
$$2\|D_H\|\delta_H + \delta_H^2 < \frac{g_d}{2}.$$
If this gate fails, the extracted projector has no certified perturbative identity at the stated error scale.
5. For every proposed common operation, compute $\beta(a)$ and the operation-leakage bound in Theorem 45.
6. For every long transfer chain used in an interpretation, report the pathwise leakage certificate from Theorem 48 rather than extrapolating from single-edge defects.

A minimal structural report is therefore

$$\boxed{\left(d, \lambda_d, \lambda_{d+1}, g_d, \delta_H, \|P_d - \widetilde{P}_d\|_{\mathrm{bound}}, \beta(a), \Lambda_\gamma\right),}$$

where Λ_γ denotes the right-hand side of the pathwise leakage bound. This structural report is intentionally incomplete as a workspace certificate. Parts VI–IX add the predictive, causal, null, and irreducibility gates required to interpret the low-energy subspace operationally.

PART VI

Dynamic Workspace Theory

19. Probabilistic Local and Relational Dynamics

The static theory identifies compatible directions across a declared transfer network. A dynamic workspace claim requires additional probabilistic objects: a history, a target future, a decision problem, and an intervention family. These must be fixed before sufficiency or causal efficacy is evaluated.

Definition 53 (Dynamic relational realization)**.** A dynamic relational realization is defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and consists of:

- node states $X_i(t) \in \mathsf{X}_i$ for $i \in V$;
- edge messages $M_e(t) \in \mathsf{M}_e$ for $e \in E$;
- external inputs E_t, local actions $A_i(t)$, and disturbances $\xi_i(t), \zeta_e(t)$;
- time-dependent transfer data $H_e(t)$, local core vectors $c_i(t)$, and transfer Laplacians $L_t = L_{H(t)}$; and
- a future target $Y_t^+ \in \mathsf{Y}$.

All state, history, and target spaces are assumed standard Borel unless a stronger topology is explicitly invoked.

A representative local update is

$$X_i(t+1) = F_i(X_i(t), \{M_e(t) : t(e) = i\}, A_i(t), E_t, \xi_i(t)),$$

with messages

$$M_e(t) = K_e\Big(X_{s(e)}(t), W_t, E_t, \zeta_e(t)\Big).$$

This notation permits, but does not require, the current workspace state to modulate future transfers.

For a coalition $A \subseteq V$, let $\mathbf{H}_t^A$ denote the random history available to that coalition through time t. Its generated information is

$$\mathcal{F}_t^A = \sigma(\mathbf{H}_t^A).$$

The full relational history is $\mathbf{H}_t = \mathbf{H}_t^V$, with

$$\mathcal{F}_t^A \subseteq \mathcal{F}_t^B \qquad \text{whenever } A \subseteq B.$$

The empty-coalition history $\mathbf{H}_t^\varnothing$ is the trivial constant statistic and represents prediction from the background target law alone. A concrete history may contain node trajectories, incident messages, actions, artifacts, transfer-map estimates, and the declared graph history. The exact contents must be reported because coalition sufficiency is relative to what a coalition is permitted to observe.

The target Y_t^+ can be a next output, a finite future trajectory, a research decision, a proof outcome, or any measurable downstream quantity. Every predictive workspace claim is therefore target-relative:

$$\text{WorkspaceFor}(Y_t^+) \not\Longrightarrow \text{WorkspaceFor}(\widetilde{Y}_t^+).$$

Let

$$c(t) = (c_i(t))_{i \in V} \in \mathcal{C}_V(t), \qquad P_t = \mathbf{1}_{[0,\varepsilon_t]}(L_t).$$

The instantaneous structural workspace coordinate is

$$\boxed{W_t^{\mathrm{str}} = P_t c(t) \in \Gamma_{\varepsilon_t}(H(t)).}$$

When P_t varies with time, these coordinates inhabit a moving family of subspaces. Parts X–XII will supply the transport required to compare them across time. In what follows, all sufficiency and intervention statements are made at a fixed decision time t.

Remark 54 (Three objects that must not be conflated)**.** The realization state S_t, the complete history $\mathbf{H}_t$, and the structural coordinate W_t^{str} are different objects. The first is an instantaneous physical or computational state, the second contains all admissible past information, and the third is a low-transfer-energy statistic extracted from the declared core geometry.

20. Canonical Predictive and Interventional Causal States

The quotient notation $[\mathbf{H}_t]_{\sim_Y}$ expresses the right intuition but hides measurability and null-set issues. A canonical construction is obtained from regular conditional distributions.

Definition 55 (Predictive kernel and canonical state)**.** Let H_t and Y be standard Borel spaces. A regular conditional kernel is a measurable map

$$K_t : \mathsf{H}_t \times \mathcal{B}(\mathsf{Y}) \to [0,1]$$

satisfying

$$K_t(h, A) = \mathbb{P}(Y_t^+ \in A \mid \mathbf{H}_t = h)$$

for $\mathbb{P}_{\mathbf{H}_t}$-almost every h. Let $\mathcal{P}(\mathsf{Y})$ carry its evaluation sigma-algebra. The canonical predictive state is the random probability measure

$$\boxed{Z_t = K_t(\mathbf{H}_t, \cdot) \in \mathcal{P}(\mathsf{Y}).}$$

Two histories are predictively equivalent precisely when they induce the same conditional law of Y_t^+.

Theorem 56 (Predictive-state sufficiency and minimality, T_{12})**.** *Under Definition 55:*

1. *Z_t is predictively sufficient:*
$$Y_t^+ \perp\!\!\!\perp \mathbf{H}_t \mid Z_t.$$
2. *If $S_t = s(\mathbf{H}_t)$ is any other history statistic satisfying*
$$Y_t^+ \perp\!\!\!\perp \mathbf{H}_t \mid S_t,$$
then there is a measurable map ϕ such that
$$Z_t = \phi(S_t) \qquad \text{almost surely.}$$
3. *Consequently, any two minimal sufficient predictive states are measurably isomorphic after removal of null sets.*

Proof. For every bounded measurable $f : \mathsf{Y} \to \mathbb{R}$, define

$$Z_t(f) = \int_{\mathsf{Y}} f(y)\, Z_t(\mathrm{d}y).$$

By construction,

$$\mathbb{E}[f(Y_t^+) \mid \mathbf{H}_t] = Z_t(f).$$

Since Z_t is a measurable function of $\mathbf{H}_t$, the tower property gives

$$\mathbb{E}[f(Y_t^+) \mid Z_t] = \mathbb{E}[Z_t(f) \mid Z_t] = Z_t(f).$$

Hence conditioning additionally on $\mathbf{H}_t$ does not change the conditional expectation, which is the stated conditional independence.

Now let $S_t = s(\mathbf{H}_t)$ be sufficient. Then, for every measurable $A \subseteq \mathsf{Y}$,

$$K_t(\mathbf{H}_t, A) = \mathbb{P}(Y_t^+ \in A \mid \mathbf{H}_t) = \mathbb{P}(Y_t^+ \in A \mid S_t) \qquad \text{a.s.}$$

Choose a countable determining class for probability measures on the standard Borel space Y. The values of $K_t(\mathbf{H}_t, \cdot)$ on this class are measurable functions of S_t, and therefore the complete random measure Z_t is a measurable function of S_t. Applying this conclusion in both directions to two minimal sufficient states gives the final claim. □

Remark 57 (Meaning of minimality)**.** The theorem does not say that Z_t has the fewest numerical coordinates in a chosen parametrization. It says that every sufficient history statistic must determine Z_t almost surely. This is an information order, invariant under one-to-one reparametrization.

For an arbitrary history statistic $S_t = s(\mathbf{H}_t)$, let

$$K_t^S(s, \cdot) = \mathbb{P}(Y_t^+ \in \cdot \mid S_t = s).$$

Whenever the conditional relative entropy is finite, define the sufficiency defect

$$\boxed{\delta_{\mathrm{suff}}(S_t) = \mathbb{E}\Big[\mathrm{KL}\Big(K_t(\mathbf{H}_t, \cdot) \,\Big\|\, K_t^S(S_t, \cdot)\Big)\Big].}$$

Proposition 58 (Conditional-information identity)**.** *If S_t is a measurable function of $\mathbf{H}_t$, then*

$$\boxed{\delta_{\mathrm{suff}}(S_t) = I(Y_t^+; \mathbf{H}_t \mid S_t).}$$

In particular, the defect vanishes if and only if S_t is predictively sufficient, up to the usual almost-sure qualifications.

Proof. The conditional mutual information is the expected relative entropy between the conditional law given $(\mathbf{H}_t, S_t)$ and the conditional law given S_t. Because $S_t = s(\mathbf{H}_t)$, the first of these is exactly $K_t(\mathbf{H}_t, \cdot)$. □

Sufficiency has a direct decision-theoretic interpretation. Use the convention

$$\mathrm{TV}(P, Q) = \sup_A |P(A) - Q(A)|.$$

Let A be a finite action set and $\ell : \mathsf{Y} \times \mathsf{A} \to [0, 1]$ a bounded loss. Define

$$\mathcal{R}_\ell(S_t) = \inf_d \mathbb{E}\Big[\ell\big(Y_t^+, d(S_t)\big)\Big].$$

Proposition 59 (Approximate sufficiency bounds excess Bayes risk)**.** *Under the preceding assumptions,*

$$\boxed{0 \le \mathcal{R}_\ell(S_t) - \mathcal{R}_\ell(\mathbf{H}_t) \le \mathbb{E}\Big[\mathrm{TV}\Big(K_t(\mathbf{H}_t), K_t^S(S_t)\Big)\Big] \le \sqrt{\frac{\delta_{\mathrm{suff}}(S_t)}{2}}.}$$

Proof. More information cannot increase optimal risk, giving the first inequality. For any fixed action, its conditional expected loss changes by at most total variation when the conditional law changes. Taking the infimum over actions preserves this bound. Conditional Pinsker inequality followed by Jensen's inequality gives the final estimate. □

For a finite target alphabet and logarithmic scoring loss

$$\ell_{\log}(y, q) = -\log q(y), \qquad q \in \mathcal{P}(\mathsf{Y}),$$

the identity is exact:

$$\boxed{\mathcal{R}_{\log}(S_t) = H(Y_t^+ \mid S_t), \qquad \mathcal{R}_{\log}(S_t) - \mathcal{R}_{\log}(\mathbf{H}_t) = I(Y_t^+; \mathbf{H}_t \mid S_t).}$$

Natural logarithms measure the gap in nats; base-two logarithms measure it in bits.

Intervention-indexed causal state

Observational prediction may merge histories that behave identically under the current policy but respond differently to future interventions. Let $\mathcal{I}$ be a finite or countable family of interventions applied only after time t. For each $\iota \in \mathcal{I}$, let

$$K_t^\iota(h, \cdot) = \mathbb{P}^\iota(Y_t^+ \in \cdot \mid \mathbf{H}_t = h)$$

be a regular post-intervention kernel on the common pre-intervention history space. Define

$$\boxed{Z_t^{\mathcal{I}} = \big(K_t^\iota(\mathbf{H}_t, \cdot)\big)_{\iota \in \mathcal{I}} \in \mathcal{P}(\mathsf{Y})^{\mathcal{I}}.}$$

Corollary 60 (Minimal simultaneous interventional state)**.** *The random element $Z_t^{\mathcal{I}}$ is sufficient simultaneously for all post-intervention target laws. If a statistic $S_t = s(\mathbf{H}_t)$ determines the conditional target law for every $\iota \in \mathcal{I}$, then there exists a measurable $\phi_{\mathcal{I}}$ such that*

$$Z_t^{\mathcal{I}} = \phi_{\mathcal{I}}(S_t) \qquad a.s.$$

Proof. Apply Theorem 56 componentwise. A countable product of standard Borel spaces is standard Borel, so the component maps combine into a single measurable factor map. □

The observational state is the special case $\mathcal{I} = \{\mathrm{id}\}$. A causal workspace claim should align with $Z_t^{\mathcal{I}}$ for a declared intervention family rather than only with the observational state.

21. Structural–Operational Alignment

A structural workspace coordinate becomes operational only when it retains the canonical predictive or interventional state relevant to the declared task. Let

$$W_t = \pi_t(\mathbf{H}_t)$$

be a measurable candidate workspace statistic satisfying the structural support condition

$$W_t \in \Gamma_{\varepsilon_t}(H(t)) \qquad \text{almost surely.}$$

The map π_t may include encoding into a moving low-energy bundle; no canonical operational interpretation follows from the spectral projector alone.

Definition 61 (Exact structural–operational alignment)**.** The candidate W_t is observationally aligned when there exists a measurable map ϕ_t such that

$$Z_t = \phi_t(W_t) \qquad \text{a.s.}$$

It is $\mathcal{I}$-causally aligned when

$$Z_t^{\mathcal{I}} = \phi_t^{\mathcal{I}}(W_t) \qquad \text{a.s.}$$

Theorem 62 (Exact alignment criterion)**.** *For any workspace statistic $W_t = \pi_t(\mathbf{H}_t)$, the following are equivalent:*

1. $Y_t^+ \perp\!\!\!\perp \mathbf{H}_t \mid W_t$*;*
2. $I(Y_t^+; \mathbf{H}_t \mid W_t) = 0$*;*
3. *the canonical predictive state factors through W_t:*

$$Z_t = \phi_t(W_t) \quad \textit{a.s.}$$

for some measurable ϕ_t.

The analogous equivalence holds simultaneously over $\mathcal{I}$ with $Z_t^{\mathcal{I}}$.

Proof. The equivalence of the first two statements is the zero-divergence characterization of conditional mutual information. The first statement and Theorem 56 imply the factorization in the third. Conversely, if $Z_t = \phi_t(W_t)$, then

$$\mathbb{P}(Y_t^+ \in A \mid \mathbf{H}_t) = Z_t(A) = \phi_t(W_t)(A),$$

which depends on the history only through W_t, proving sufficiency. The interventional statement follows from Corollary 60. □

Define the observational alignment defect by

$$\boxed{\delta_{\text{align}}^{\text{obs}}(W_t) = I_{\mathbb{P}}(Y_t^+; \mathbf{H}_t \mid W_t).}$$

For a countable intervention family with declared weights $q_\iota > 0$, $\sum_\iota q_\iota = 1$, define

$$\boxed{\delta_{\text{align}}^{\mathcal{I}}(W_t) = \sum_{\iota \in \mathcal{I}} q_\iota I_{\mathbb{P}^\iota}(Y_t^+; \mathbf{H}_t \mid W_t).}$$

For bounded decision losses, Proposition 59 applies under each intervention and gives the weighted bound

$$\sum_\iota q_\iota \left[\mathcal{R}_\ell^\iota(W_t) - \mathcal{R}_\ell^\iota(\mathbf{H}_t)\right] \le \sqrt{\frac{\delta_{\text{align}}^{\mathcal{I}}(W_t)}{2}}.$$

Structural coherence and operational alignment are independent

The static defect

$$\delta_{\mathrm{coh}}(W_t)$$

and the operational defect

$$\delta^{\mathcal{I}}_{\mathrm{align}}(W_t)$$

answer different questions. The first measures compatibility under transfer. The second measures loss of task-relevant conditional law.

Regime	Structural status	Operational interpretation
Low coherence defect, low alignment defect	Stable transported coordinate	Candidate operational workspace for the declared target and interventions.
Low coherence defect, high alignment defect	Stable but task-irrelevant shared structure	Archive, convention, or synchronized representation.
High coherence defect, low alignment defect	Predictively adequate but not transport-stable	Localized or brittle sufficient statistic.
High coherence defect, high alignment defect	Neither stable nor sufficient	Rejected candidate at the stated scales.

Thus neither low transfer energy nor high predictive performance alone is enough. The combined gate is

$$\boxed{\delta_{\mathrm{coh}}(W_t) \le \varepsilon_{\mathrm{coh}}, \qquad \delta^{\mathcal{I}}_{\mathrm{align}}(W_t) \le \varepsilon_{\mathrm{align}}.}$$

Example 63 (Coherent but irrelevant archive)**.** Suppose every node perfectly copies a common random variable R, so the corresponding transfer energy is zero. If $R \perp\!\!\!\perp Y_t^+$, then

$$\delta_{\mathrm{coh}}(R) = 0, \qquad \delta^{\mathrm{obs}}_{\mathrm{align}}(R) = I(Y_t^+; \mathbf{H}_t \mid R),$$

which can be as large as the entire predictive information in the history. Perfect agreement therefore does not imply workspace functionality.

Example 64 (Sufficient but localized statistic)**.** If one node contains the canonical state Z_t and all transfers from that node are unstable, then $W_t = Z_t$ is sufficient but need not lie in a stable low-energy transfer subspace. This is a local controller or archive, not yet a stratified shared workspace.

PART VII

Relational Irreducibility

22. Coalition Information and Bayes Risk

For $A \subseteq V$, the coalition statistic is $\mathbf{H}_t^A$. Given an action space A and loss ℓ, define its unrestricted Bayes risk by

$$\boxed{\mathcal{R}_\ell(A) = \inf_d \mathbb{E}\Big[\ell(Y_t^+, d(\mathbf{H}_t^A))\Big].}$$

The infimum is over measurable decision rules for which the expectation is defined. The full-system risk is $\mathcal{R}_\ell(V)$.

Proposition 65 (Coalition monotonicity)**.** *If $A \subseteq B \subseteq V$, then*

$$\mathcal{R}_\ell(B) \le \mathcal{R}_\ell(A).$$

Proof. Every decision rule available to coalition A is also available to coalition B, which can ignore its additional information. □

For finite Y, logarithmic loss converts risk gaps into conditional information exactly.

Theorem 66 (Coalition log-risk identity)**.** *Assume $H(Y_t^+) < \infty$. If $A \subseteq B \subseteq V$, then*

$$\boxed{\mathcal{R}_{\log}(A) - \mathcal{R}_{\log}(B) = I\Big(Y_t^+; \mathbf{H}_t^B \mid \mathbf{H}_t^A\Big).}$$

In particular,

$$\mathcal{R}_{\log}(A) - \mathcal{R}_{\log}(V) = I\Big(Y_t^+; \mathbf{H}_t \mid \mathbf{H}_t^A\Big).$$

Proof. Under logarithmic loss the optimal report is the conditional target law, so

$$\mathcal{R}_{\log}(A) = H(Y_t^+ \mid \mathbf{H}_t^A).$$

Because $\mathbf{H}_t^A$ is a function of $\mathbf{H}_t^B$, subtracting the two conditional entropies gives the stated conditional mutual information. □

The theorem makes the decomposition explicit: irreducibility is not an unqualified property of a system. It is a property relative to the chosen node partition, accessible histories, target, and decision loss.

23. Joint Advantage and Its Scope

Define the coalition gap

$$\Delta_A^\ell = \mathcal{R}_\ell(A) - \mathcal{R}_\ell(V) \ge 0.$$

The node-level joint advantage is

$$\boxed{\Delta_{\text{joint}}^\ell = \min_{i \in V} \Delta_{\{i\}}^\ell.}$$

Under logarithmic loss,

$$\boxed{\Delta_{\text{joint}}^{\log} = \min_{i \in V} I\left(Y_t^+; \mathbf{H}_t \mid \mathbf{H}_t^{\{i\}}\right).}$$

Therefore

$$\Delta_{\text{joint}}^{\log} > 0$$

means that every individual component omits some predictive information available to the complete system.

This quantity should not automatically be identified with a particular notion of partial-information synergy. A positive gap can arise from complementary information distributed across components, from higher-order interaction, or from both. The theory uses the operationally weaker and more directly testable statement:

no individual node is sufficient for the declared target.

For $0 \le r \le |V|$, define the order-r coalition gap profile by

$$\boxed{\Delta_\ell^{(r)} = \min_{\substack{A \subseteq V \\ |A| \le r}} \left[\mathcal{R}_\ell(A) - \mathcal{R}_\ell(V)\right].}$$

The profile is nonincreasing in r and reaches zero by $r = |V|$. It records how rapidly operative capacity becomes reconstructible as larger coalitions are permitted.

24. Minimal Sufficient Coalition

For tolerance $\varepsilon \ge 0$, define

$$\boxed{k_\varepsilon = \min\left\{|A| : \mathcal{R}_\ell(A) \le \mathcal{R}_\ell(V) + \varepsilon\right\}.}$$

The set of minimal coalitions is

$$\mathcal{K}_\varepsilon = \left\{A \subseteq V : |A| = k_\varepsilon, \mathcal{R}_\ell(A) \le \mathcal{R}_\ell(V) + \varepsilon\right\}.$$

Corollary 67 (Information form of coalition size)**.** *Under the assumptions of Theorem 66,*

$$\boxed{k_\varepsilon = \min\left\{|A| : I(Y_t^+; \mathbf{H}_t \mid \mathbf{H}_t^A) \le \varepsilon\right\}.}$$

Equivalently, k_ε is the first r for which $\Delta_{\log}^{(r)} \le \varepsilon$.

This yields a graded classification:

$$\begin{aligned}
k_0 = 0 &\implies \text{the background law alone is sufficient,} \\
k_0 = 1 &\implies \text{some individual node is sufficient,} \\
k_0 \ge 2 &\implies \text{node-level relational irreducibility,} \\
k_0 = |V| &\implies \text{every proper coalition is insufficient.}
\end{aligned}$$

The last condition is much stronger and should not be inferred from the first.

25. Reconstructibility of the Canonical State

Let $d_{\mathcal{P}}$ be a bounded metric generating the measurable structure on $\mathcal{P}(\mathsf{Y})$. Define the reconstruction defect of the canonical state from coalition A by

$$\delta_A^Z = \inf_{\psi_A} \mathbb{E}\Big[d_{\mathcal{P}}\Big(Z_t, \psi_A(\mathbf{H}_t^A)\Big)\Big].$$

Proposition 68 (Sufficiency–reconstruction equivalence)**.** *For a coalition $A \subseteq V$, the following are equivalent:*

1. $\mathbf{H}_t^A$ *is sufficient for* Y_t^+ *relative to the full history;*
2. Z_t *is measurable with respect to* $\mathcal{F}_t^A$*;*
3. *there exists a measurable* ψ_A *such that*

$$Z_t = \psi_A(\mathbf{H}_t^A) \qquad \textit{almost surely.}$$

Each condition implies $\delta_A^Z = 0$. Under finite-target logarithmic loss, they are also equivalent to

$$\mathcal{R}_{\log}(A) = \mathcal{R}_{\log}(V).$$

If the reconstruction infimum is known to be attained, then $\delta_A^Z = 0$ is itself equivalent to the three conditions.

Proof. The equivalence of the first two statements is Theorem 56; the second and third are two formulations of measurability up to null sets. An exact reconstructor has zero expected metric error. The log-risk statement follows from Theorem 66 and the zero condition for conditional mutual information. Without an attainment or compactness condition, an infimum of zero alone certifies arbitrarily accurate reconstruction but need not supply a single exact minimizing map. □

For an arbitrary enriched workspace variable W_t, local non-reconstructibility need not imply predictive irreducibility: W_t may contain task-irrelevant material. The equivalence is exact for the canonical state Z_t, or for a candidate W_t known to be minimally aligned with it.

26. Matched Relational Nulls

A null control must preserve the local facts that the claim is not intended to challenge while destroying the relational organization being tested. Merely reducing model capacity or replacing the workspace by noise is not a matched null.

Let P be the intact law of $(Y_t^+, \mathbf{H}_t^1, \dots, \mathbf{H}_t^n)$. Define the local-marginal null class

$$\mathcal{N}_{\mathrm{marg}}(P) = \Big\{Q : Q_{\mathbf{H}_t^i} = P_{\mathbf{H}_t^i} \text{ for every } i\Big\},$$

and the stronger local-predictive-matched class

$$\boxed{\mathcal{N}_{\mathrm{pred}}(P) = \Big\{Q : Q_{(Y_t^+, \mathbf{H}_t^i)} = P_{(Y_t^+, \mathbf{H}_t^i)} \text{ for every } i\Big\}.}$$

Because $\mathbf{H}_t^i$ can include the complete local trajectory, the latter class can preserve all single-component temporal and predictive behavior while altering cross-component organization.

Proposition 69 (Local-risk invariance under predictive matching)**.** *If $Q \in \mathcal{N}_{\mathrm{pred}}(P)$, then for every node i and every loss determined by $(Y_t^+, d(\mathbf{H}_t^i))$,*

$$\boxed{\mathcal{R}_{\ell,Q}(\{i\}) = \mathcal{R}_{\ell,P}(\{i\}).}$$

Proof. Every such risk depends only on the joint law of $(Y_t^+, \mathbf{H}_t^i)$, which is identical under P and Q. □

A valid null must also satisfy a declared relation-breaking condition. Examples include:

- permuting source–target message pairings while preserving each message marginal;
- shuffling temporal correspondence across nodes while preserving every local trajectory;
- permuting role bindings while preserving the local token or state distributions;
- replacing an artifact by a matched-size, matched-format artifact whose relation to the surrounding history has been destroyed; and
- randomizing transfer channels through a physical or software intervention while preserving message rate and marginal content.

For a predeclared null Q, define

$$\boxed{\Delta_{\mathrm{tr}}(Q) = \mathcal{R}_{\ell,Q}(V) - \mathcal{R}_{\ell,P}(V).}$$

For a finite validated panel $\mathcal{Q} = \{Q_1, \dots, Q_m\} \subseteq \mathcal{N}_{\mathrm{pred}}(P)$, report both the complete vector

$$\boldsymbol{\Delta}_{\mathrm{tr}}(\mathcal{Q}) = (\Delta_{\mathrm{tr}}(Q_1), \dots, \Delta_{\mathrm{tr}}(Q_m))$$

and the robust panel minimum

$$\Delta_{\mathrm{tr}}^{\min}(\mathcal{Q}) = \min_{Q \in \mathcal{Q}} \Delta_{\mathrm{tr}}(Q).$$

The panel must be fixed independently of the observed favorable gap. Otherwise null selection can manufacture apparent relational dependence.

27. Relational Witness Pairs

Definition 70 (Relational witness pair)**.** A pair of laws (P, Q) is a predictive relational witness pair for the decomposition V and target Y_t^+ when

$$Q \in \mathcal{N}_{\mathrm{pred}}(P)$$

but

$$\mathcal{R}_{\ell,Q}(V) \neq \mathcal{R}_{\ell,P}(V).$$

The two systems agree on every single-node predictive marginal but differ in full-system operative capacity.

Example 71 (XOR witness pair)**.** Under the intact law P, let

$$X_1, X_2 \overset{\mathrm{iid}}{\sim} \mathrm{Bernoulli}(1/2), \qquad Y = X_1 \oplus X_2.$$

Then each (X_i, Y) pair is independent and uniform, so

$$H_P(Y \mid X_i) = 1 \text{ bit}, \qquad H_P(Y \mid X_1, X_2) = 0.$$

Let Q make X_1, X_2, Y mutually independent fair bits. Then

$$Q_{(X_i,Y)} = P_{(X_i,Y)} \qquad (i = 1, 2),$$

but

$$H_Q(Y \mid X_1, X_2) = 1 \text{ bit}.$$

Therefore

$$\Delta^{\log}_{\text{joint}} = 1 \text{ bit}, \qquad k_0 = 2, \qquad \Delta_{\text{tr}}(Q) = 1 \text{ bit}.$$

This proves predictive relational structure relative to the declared variables. It does not, by itself, prove that any workspace variable is causally active.

Example 72 (Redundant-copy control)**.** Let Y be a fair bit and set $X_1 = X_2 = Y$. The two nodes are perfectly coherent and the full system predicts perfectly, but

$$H(Y \mid X_i) = 0, \qquad \Delta^{\log}_{\text{joint}} = 0, \qquad k_0 = 1.$$

Duplication is therefore not relational irreducibility.

The witness-pair criterion is stronger than a performance comparison against a smaller model because it holds local predictive behavior fixed. It remains observational. Part VIII adds interventions and converts predictive relationality into an operational causal claim.

PART VIII

Causal Efficacy and Reentry

28. Feasible Workspace Interventions

A notation such as $\mathrm{do}(W_t = w)$ is meaningful only if the candidate workspace is an endogenous variable in a structural causal model or if an intervention on underlying variables is shown to realize that value. Abstractly setting a projection coordinate need not correspond to any possible physical or computational manipulation.

Definition 73 (Time-unrolled causal realization)**.** A time-unrolled causal realization is a structural causal model

$$\mathcal{M} = (\mathcal{U}, \mathcal{V}, F, \mathbb{P}_U)$$

whose endogenous variables include the local states, messages, actions, artifacts, target variables, and a measurable workspace extractor

$$\pi_t : \mathsf{S}_t \to \mathsf{W}_t, \qquad W_t = \pi_t(S_t).$$

The unrolled graph is acyclic even when the rolled process is recurrent.

Definition 74 (Feasible implementation intervention)**.** Let B_t denote declared background variables that should remain matched. A family $\{\iota_w : w \in \mathsf{W}_t^\circ\}$ is a feasible implementation of workspace values on the experimental domain W_t° when:

1. each ι_w acts only on declared manipulable parents, channels, or artifacts at or after time t;
2. the intervened realization satisfies
$$\pi_t(S_t^{\iota_w}) = w \qquad \text{almost surely;}$$
3. the background law is matched across compared interventions; and
4. the intervention does not directly modify Y_t^+ except through paths included in the declared causal design.

Let D be a divergence or metric on probability laws. For a matched background value b, define

$$\mathrm{CE}_{\mathbf{W}}(w, w'; b) = D\Big(\mathbb{P}^{\iota_w}(Y_t^+ \in \cdot \mid B_t = b), \mathbb{P}^{\iota_{w'}}(Y_t^+ \in \cdot \mid B_t = b)\Big).$$

For a declared comparison design ν over (w, w', b), define

$$\boxed{\eta^{\nu}_{\mathrm{causal}} = \mathbb{E}_{(w,w',b)\sim\nu}\left[\mathrm{CE}_{\mathbf{W}}(w, w'; b)\right].}$$

The maximum

$$\eta^{\max}_{\mathrm{causal}} = \sup_{w,w',b} \mathrm{CE}_{\mathbf{W}}(w, w'; b)$$

may be useful for discovery, but a design-weighted effect is harder to inflate by selecting a single exceptional pair after inspection.

Proposition 75 (Coordinate invariance of intervention effects)**.** *Let $g : \mathsf{W}_t \to \mathsf{W}'_t$ be a measurable bijection and suppose the reparametrized intervention family satisfies*

$$\iota'_{g(w)} = \iota_w.$$

Then every divergence-based causal effect is preserved:

$$\mathrm{CE}_{\mathbf{W}'}(g(w), g(w'); b) = \mathrm{CE}_{\mathbf{W}}(w, w'; b).$$

Proof. The two labels designate the same underlying interventions and therefore induce the same pair of target laws. Only the coordinate name has changed. □

This is the causal analogue of gauge invariance. It also shows why effects should be attached to intervention implementations rather than to arbitrary numerical coordinates.

29. Causal Specificity and Transfer Mediation

A causal effect on the target is not workspace-specific if an equal-cost perturbation outside the candidate workspace produces the same effect. Let $c(\iota)$ be a declared intervention cost and let

$$\mathcal{I}_{\mathbf{W}}(r), \qquad \mathcal{I}_{\perp}(r)$$

be matched families with cost r, the first changing workspace coordinates and the second changing complementary coordinates while preserving the workspace value as closely as the implementation permits.

Define

$$\eta_{\mathbf{W}}(r) = \mathbb{E}_{\iota,\iota' \sim \mathcal{I}_{\mathbf{W}}(r)} D\Big(\mathbb{P}^{\iota}(Y_t^+), \mathbb{P}^{\iota'}(Y_t^+)\Big),$$

$$\eta_{\perp}(r) = \mathbb{E}_{\jmath,\jmath' \sim \mathcal{I}_{\perp}(r)} D\Big(\mathbb{P}^{\jmath}(Y_t^+), \mathbb{P}^{\jmath'}(Y_t^+)\Big),$$

and

$$\boxed{\Delta_{\mathrm{specific}}(r) = \eta_{\mathbf{W}}(r) - \eta_{\perp}(r).}$$

A positive value supports causal concentration in the candidate workspace at the declared perturbation scale.

If the complement is defined by $\Gamma_{\varepsilon}(H)^{\perp}$, the result is metric dependent. Under nonunitary gauge changes, either the metric must be transported with the coordinates or interventions must be matched by a coordinate-independent physical cost. Otherwise specificity can change merely because the representation was rescaled.

Causal transfer nulls

Let $\jmath_E$ randomize selected communication edges while preserving their marginal message laws, rates, and timing budget. The causal transfer gap is

$$\boxed{\Delta_{\mathrm{tr}}^{\mathrm{do}}(\jmath_E) = \mathcal{R}_{\ell,\mathbb{P}^{\jmath_E}}(V) - \mathcal{R}_{\ell,\mathbb{P}}(V).}$$

Unlike an observational null law, this quantity is anchored to an implemented channel intervention. A positive effect is evidence that the intact alignment of messages, rather than merely their marginal content, contributes to task performance.

A stronger mediation test equalizes the workspace state after two distinct channel interventions. Let μ be a common imposed workspace distribution, implemented by a feasible stochastic intervention in the sense of Definition 73. The notation below is shorthand for that implementation. Define the residual mediation defect

$$\delta_{\mathrm{med}} = \sup_{\jmath,\jmath'} D\Big(\mathbb{P}^{\jmath,\mathrm{do}(W_t\sim\mu)}(Y_t^+), \mathbb{P}^{\jmath',\mathrm{do}(W_t\sim\mu)}(Y_t^+)\Big).$$

A small value means that once the workspace state is equalized, little of the channel intervention's effect remains. This supports, but does not by itself identify without the causal assumptions, the claim that W_t mediates the relevant relational influence.

30. Causal Recurrence, Loop-Mediated Reentry, and Relationality

Temporal correlation is weaker than causal recurrence. Define the total workspace recurrence effect at lag τ by

$$\boxed{\eta_{\mathrm{rec}}(\tau) = \mathbb{E}_{(w,w',b)\sim\nu} D(\mathbb{P}^{\iota_w}(W_{t+\tau}\in\cdot \mid B_t = b), \mathbb{P}^{\iota_{w'}}(W_{t+\tau}\in\cdot \mid B_t = b))\,.}$$

A nonzero value shows that manipulating the present workspace changes a later workspace state. It may, however, be explained by direct copying or a retained memory register.

To isolate reentry, consider the unrolled causal graph and remove every direct retention edge whose source and target are both workspace variables at successive times. Let

$$\mathcal{M}^{\backslash\mathrm{dir}}$$

denote the resulting path-specific model, preserving paths that leave the workspace through a nonworkspace participant, action, message, or artifact and later return. Define

$$\boxed{\eta_{\mathrm{loop}}(\tau) = \mathbb{E}_{(w,w',b)\sim\nu} D\Big(\mathbb{P}^{\iota_w}_{\mathcal{M}^{\backslash\mathrm{dir}}}(W_{t+\tau}\in\cdot \mid B_t = b), \mathbb{P}^{\iota_{w'}}_{\mathcal{M}^{\backslash\mathrm{dir}}}(W_{t+\tau}\in\cdot \mid B_t = b)\Big).}$$

Then the distinctions are:

$$\begin{aligned} I(W_t; W_{t+\tau}) > 0 &\quad \text{temporal dependence only,}\\ \eta_{\mathrm{rec}}(\tau) > 0 &\quad \text{causal persistence,}\\ \eta_{\mathrm{loop}}(\tau) > 0 &\quad \text{causal exit-and-return reentry.}\end{aligned}$$

Path-specific effects require the usual consistency and no-unmeasured-confounding assumptions appropriate to the implementation. In software systems, edge interventions can often be literal. In human collaboration, these assumptions are substantially harder to justify.

Theorem 76 (Operational relationality, T_{13})**.** *Let Y be finite, let P be the intact law at a fixed decision time, and let $W_t = \pi_t(\mathbf{H}_t)$ be a structurally supported candidate. Let $\mathcal{Q} \subseteq \mathcal{N}_{\mathrm{pred}}(P)$ be a nonempty, predeclared panel of validated relation-breaking nulls. Assume:*

1. *exact structural–operational alignment,*

$$I_P(Y_t^+; \mathbf{H}_t \mid W_t) = 0;$$

2. *a positive node-level information gap,*

$$\delta_{\mathrm{node}} = \min_{i\in V} I_P(Y_t^+; \mathbf{H}_t \mid \mathbf{H}_t^i) > 0;$$

3. *a positive robust matched-null gap,*

$$\delta_{\text{null}} = \min_{Q \in \mathcal{Q}} \left[H_Q(Y_t^+ \mid \mathbf{H}_t) - H_P(Y_t^+ \mid \mathbf{H}_t) \right] > 0;$$

4. *a feasible intervention design with*

$$\eta^{\nu}_{\text{causal}} > 0;$$

5. *positive matched causal specificity at some declared scale r,*

$$\Delta_{\text{specific}}(r) > 0.$$

Then, relative to the declared decomposition, target, log loss, null panel, and intervention family:

1. *no statistic of any single-node history is sufficient for Y_t^+, and $k_0 \geq 2$;*
2. *W_t cannot be reconstructed almost surely from any single-node history;*
3. *every null $Q \in \mathcal{Q}$ preserves all single-node predictive risks but has strictly worse full-system log risk than the intact law;*
4. *the candidate workspace is causally efficacious and its effect exceeds that of the declared matched off-workspace perturbations.*

If, in addition, $\eta_{\text{loop}}(\tau) > 0$, the workspace is loop-reentrant at lag τ.

Proof. By Theorem 66, for each node i,

$$\mathcal{R}_{\log,P}(\{i\}) - \mathcal{R}_{\log,P}(V) = I_P(Y_t^+; \mathbf{H}_t \mid \mathbf{H}_t^i) \geq \delta_{\text{node}} > 0.$$

Hence no single-node history, and therefore no statistic of one, is sufficient; this gives $k_0 \geq 2$.

Exact alignment and Theorem 62 imply that the canonical predictive state factors through W_t. If $W_t = \psi_i(\mathbf{H}_t^i)$ almost surely for some node, then the canonical state would also be a function of $\mathbf{H}_t^i$, contradicting the positive node gap through Proposition 68.

For every $Q \in \mathcal{Q}$, Proposition 69 gives

$$\mathcal{R}_{\log,Q}(\{i\}) = \mathcal{R}_{\log,P}(\{i\}) \qquad \forall i.$$

The assumed null gap gives

$$\mathcal{R}_{\log,Q}(V) > \mathcal{R}_{\log,P}(V).$$

Thus a description determined only by the collection of single-node predictive marginals cannot reproduce the intact joint capacity: it would assign the same local data to P and Q while their full-system risks differ.

Finally, the feasible intervention effect establishes causal efficacy relative to the declared implementation, and the positive specificity gap shows that the effect exceeds the matched complementary perturbation effect. The final reentry statement is exactly the definition of η_{loop}. □

Remark 77 (Scope of the theorem)**.** The conclusion is operational and indexed by

$$(V, Y_t^+, \ell, \mathcal{Q}, \mathcal{I}, \nu, r).$$

Changing the decomposition, target, loss, nulls, or intervention design may change the verdict. The theorem does not establish metaphysical irreducibility, autonomous agency, selfhood, or consciousness.

Finite certification sequence

A reproducible dynamic-causal analysis should proceed in the following order:

1. Freeze the node decomposition, accessible coalition histories, target, horizon, and proper scoring loss.
2. Estimate the structural projector on training data only and encode W_t without using held-out targets.
3. Cross-fit conditional target laws from the full history, the workspace, and every declared coalition. Report alignment and coalition risk gaps.
4. Construct the null panel before inspecting null performance. Verify local-predictive matching and the intended destruction of transfer organization separately.
5. Implement feasible workspace and complementary interventions with a common background and cost design. Report both causal effect and specificity.
6. Test recurrence and then cut direct retention paths to test loop-mediated reentry.
7. Report confidence intervals or finite-sample uncertainty for every gate; do not substitute training accuracy for Bayes-risk or causal estimates.

The resulting certification report is

$$\boxed{\begin{aligned}\mathcal{C}_{\mathrm{dyn}}(W_t) = (&\delta^{\mathcal{I}}_{\mathrm{align}}, \{\Delta^{\ell}_{A}\}_{A\subseteq V}, k_{\varepsilon}, \boldsymbol{\Delta}_{\mathrm{tr}}(\mathcal{Q}),\\ &\eta^{\nu}_{\mathrm{causal}}, \Delta_{\mathrm{specific}}(r), \eta_{\mathrm{rec}}(\tau), \eta_{\mathrm{loop}}(\tau)).\end{aligned}}$$

No single coordinate certifies a relational workspace.

PART IX

Formal Workspace Definition

31. Certified Stratified Relational Workspace

Let

$$\Theta = \left(\varepsilon_{\mathrm{coh}}, \varepsilon_{\mathrm{align}}, \varepsilon_k, \underline{\eta}_c, \underline{\eta}_{\mathrm{loop}}, \delta_{\mathrm{node}}, \delta_{\mathrm{null}}, \delta_{\mathrm{spec}}, k, \varepsilon_{\mathrm{pers}}\right).$$

Let

$$\varsigma = (\mathcal{D}, Y^+, \ell, \tau, \mathcal{I}, \mathcal{Q}, \Theta)$$

be the complete declared scientific specification, where $\mathcal{Q} \subseteq \mathcal{N}_{\mathrm{pred}}(P)$ is a validated null panel. A dynamic realization contains a certified stratified relational workspace when there exists a candidate W_t satisfying all gates required for the claimed level under ς. The detailed robustness and identifiability requirements for varying ς are developed in Parts XVII and XVIII.

Structural coherence

$$\delta_{\mathrm{coh}}(W_t) \le \varepsilon_{\mathrm{coh}}.$$

Interventional structural–operational alignment

$$\delta^{\mathcal{I}}_{\mathrm{align}}(W_t) \le \varepsilon_{\mathrm{align}}.$$

Causal efficacy

$$\eta^{\nu}_{\mathrm{causal}} \ge \underline{\eta}_c.$$

Loop-mediated reentry, when claimed

$$\eta_{\mathrm{loop}}(\tau) \ge \underline{\eta}_{\mathrm{loop}}.$$

Node-level joint advantage

$$\Delta^{\ell}_{\mathrm{joint}} \ge \delta_{\mathrm{node}}.$$

Matched-null transfer dependence

$$\Delta^{\min}_{\mathrm{tr}}(\mathcal{Q}) \ge \delta_{\mathrm{null}}.$$

Causal specificity

$$\Delta_{\text{specific}}(r) \geq \delta_{\text{spec}}.$$

Distributedness

$$k_{\varepsilon_k} \geq k.$$

Persistence, when claimed

$$d_{\text{id}}(W_t, W_{t+\tau}) \leq \varepsilon_{\text{pers}}.$$

Symbolically,

$$\boxed{\text{SRW}(W_t;\varsigma) = \bigwedge_{q\in\mathcal{J}(\varsigma)} \text{Gate}_q(W_t;\varsigma).}$$

The gate set is claim-relative. A one-step operational workspace need not pass persistence or reentry gates, while a persistent reentrant workspace must. No scalar total score is sufficient: failure of a required gate cannot be hidden by high performance on another gate.

Remark 78 (Structural, predictive, causal, and temporal claims)**.** The formal hierarchy now distinguishes four logically independent questions:

Does a stable shared coordinate exist?

Does it retain the task-relevant conditional law?

Does manipulating it change downstream behavior specifically?

Does it persist or reenter through time?

The preceding structural and causal analysis answers the first three at the level of definitions and conditional theorems. The following parts develop the fourth.

PART X

Persistence

The preceding parts identify exact and approximate workspace subspaces at a single time. Persistence adds three further requirements. The structural subspace must remain trackable while the realization changes, actual states must remain near that moving subspace, and the transported content must not collapse to a trivial zero state. These are distinct conditions.

32. Coherence-Error Dynamics

32.1 Time-indexed transfer systems

At time t, let

$$\mathcal{C}_V(t) = \bigoplus_{i \in V} C_i(t), \qquad \mathcal{C}_E(t) = \bigoplus_{e \in E} C_{t(e)}(t),$$

and let

$$D_t : \mathcal{C}_V(t) \longrightarrow \mathcal{C}_E(t), \qquad L_t = D_t^* D_t$$

be the transfer discrepancy and transfer Laplacian. At time t, the combinatorial graph is held fixed while the local representations, transfers, and metrics may vary. Changes of graph are treated later by a graph correspondence.

To compare successive times, choose unitary identification maps

$$U_t^V : \mathcal{C}_V(t) \longrightarrow \mathcal{C}_V(t+1), \qquad U_t^E : \mathcal{C}_E(t) \longrightarrow \mathcal{C}_E(t+1).$$

They are comparison maps, not assumptions that the physical carriers are identical. Define the one-step diagram defect

$$\boxed{\Delta_t^D = D_{t+1} U_t^V - U_t^E D_t, \qquad \delta_t^D = \|\Delta_t^D\|.}$$

Exact temporal compatibility is the condition $\Delta_t^D = 0$.

Proposition 79 (Temporal Laplacian drift)**.** *The carried Laplacian satisfies*

$$\boxed{\left\| (U_t^V)^* L_{t+1} U_t^V - L_t \right\| \le 2\|D_t\|\delta_t^D + (\delta_t^D)^2.}$$

Proof. Set

$$B_t = (U_t^E)^* \Delta_t^D.$$

Because the comparison maps are unitary,

$$(U_t^E)^* D_{t+1} U_t^V = D_t + B_t, \qquad \|B_t\| = \delta_t^D.$$

Hence

$$(U_t^V)^* L_{t+1} U_t^V = (D_t + B_t)^* (D_t + B_t).$$

Subtract $D_t^* D_t$ and apply the triangle inequality. □

Let P_t project onto a selected rank-d low-energy workspace cluster of L_t, and set

$$\mathcal{W}_t = \operatorname{ran} P_t.$$

Write

$$g_t = \lambda_{d+1}(L_t) - \lambda_d(L_t) > 0$$

for the separating spectral gap and

$$\eta_t^L = 2\|D_t\|\delta_t^D + (\delta_t^D)^2.$$

Corollary 80 (Temporal workspace tracking)**.** *If*

$$\eta_t^L < \frac{g_t}{2},$$

then the transported and newly extracted workspace projectors satisfy

$$\boxed{q_t^{\mathrm{str}} := \left\| P_{t+1} - U_t^V P_t (U_t^V)^* \right\| \le \frac{\eta_t^L}{g_t - \eta_t^L} \le \frac{2\eta_t^L}{g_t}.}$$

Proof. Carry L_{t+1} back to time t using U_t^V, apply Theorem 43, and conjugate the resulting projector bound forward. □

The quantity q_t^{str} measures structural motion after the declared coordinate change has been removed. A workspace may therefore move substantially in raw coordinates while remaining structurally unchanged.

32.2 State tracking inside a moving workspace

Let the realized vertex state satisfy

$$x_{t+1} = F_t x_t + \xi_t,$$

where $F_t : \mathcal{C}_V(t) \to \mathcal{C}_V(t+1)$ is the actual update and ξ_t is an unmodeled disturbance. Define the off-workspace error

$$\boxed{e_t = \|(I - P_t)x_t\|.}$$

Introduce

$$\alpha_t = \|(I - P_{t+1})F_t(I - P_t)\|,$$

which measures persistence of prior off-workspace error, and

$$\ell_t = \|(I - P_{t+1})F_t P_t\|,$$

which measures leakage of a valid workspace state under the update.

Proposition 81 (Grounded coherence recurrence)**.** *Every trajectory satisfies*

$$\boxed{e_{t+1} \le \alpha_t e_t + \sigma_t,}$$

where

$$\boxed{\sigma_t = \ell_t \|P_t x_t\| + \|(I - P_{t+1})\xi_t\|.}$$

Proof. Insert $I = P_t + (I - P_t)$ before x_t, apply $I - P_{t+1}$, and use submultiplicativity:

$$\begin{aligned} e_{t+1} \le \|(I - P_{t+1})F_t(I - P_t)x_t\| \\ + \|(I - P_{t+1})F_tP_tx_t\| + \|(I - P_{t+1})\xi_t\|. \end{aligned}$$

The displayed recurrence follows from the definitions of α_t and ℓ_t. □

This derivation gives the contraction and disturbance terms an operational meaning. The scalar recurrence is not an independent postulate. It follows from an update rule, a moving workspace projector, and a declared disturbance model.

32.3 Persistence theorem T_6

Theorem 82 (Nonautonomous persistence theorem)**.** *Let $e_t \ge 0$ satisfy*

$$e_{t+1} \le \alpha_t e_t + \sigma_t, \qquad \alpha_t, \sigma_t \ge 0.$$

Then, for every $t \ge 1$,

$$\boxed{e_t \le \left(\prod_{r=0}^{t-1} \alpha_r\right) e_0 + \sum_{j=0}^{t-1} \left(\prod_{r=j+1}^{t-1} \alpha_r\right) \sigma_j,}$$

where an empty product equals one. If

$$\alpha_t \le \bar{\alpha} < 1, \qquad \sigma_t \le \bar{\sigma},$$

then

$$\boxed{e_t \le \bar{\alpha}^t e_0 + \frac{\bar{\sigma}(1 - \bar{\alpha}^t)}{1 - \bar{\alpha}},}$$

and consequently

$$\boxed{\limsup_{t\to\infty} e_t \le \frac{\bar{\sigma}}{1 - \bar{\alpha}}.}$$

Proof. Repeated substitution yields

$$e_t \le \alpha_{t-1} \cdots \alpha_0 e_0 + \alpha_{t-1} \cdots \alpha_1 \sigma_0 + \cdots + \sigma_{t-1},$$

which is the first formula. Under the uniform bounds, dominate each product by the corresponding power of $\bar{\alpha}$ and sum the geometric series. □

Corollary 83 (Vanishing-disturbance convergence)**.** *If $\alpha_t \le \bar{\alpha} < 1$ and $\sigma_t \to 0$, then*

$$e_t \to 0.$$

More generally, if $\sum_{t\ge0} \sigma_t < \infty$, then the disturbance contribution in Theorem 82 vanishes asymptotically.

Proof. Split the convolution sum into a finite early part, whose coefficients decay to zero, and a late tail on which σ_t is uniformly small. The summable case follows by the same argument. □

32.4 Mean-square persistence

For stochastic systems, pathwise bounded disturbances may be too strong. Let $\{\mathcal{F}_t\}$ be a filtration containing the history through time t.

Proposition 84 (Mean-square persistence)**.** *Suppose*

$$\mathbb{E}[e_{t+1}^2 \mid \mathcal{F}_t] \le q e_t^2 + v_t, \qquad 0 \le q < 1,$$

and $\mathbb{E}[v_t] \le \bar{v}$*. Then*

$$\boxed{\mathbb{E}[e_t^2] \le q^t \mathbb{E}[e_0^2] + \frac{\bar{v}(1-q^t)}{1-q},}$$

so

$$\limsup_{t\to\infty} \mathbb{E}[e_t^2] \le \frac{\bar{v}}{1-q}.$$

Proof. Take expectations, iterate the resulting deterministic recurrence, and sum a geometric series. □

32.5 The non-collapse condition

Low off-workspace error is not sufficient for meaningful persistence. The zero update $F_t = 0$ drives every state exactly into every linear workspace while destroying all content. Define the internal transport

$$B_t = P_{t+1} F_t P_t : \mathcal{W}_t \longrightarrow \mathcal{W}_{t+1}$$

and its non-collapse margin

$$\boxed{\mu_t^{\mathrm{keep}} = \sigma_{\min}(B_t|_{\mathcal{W}_t}).}$$

A content-preserving persistence claim requires

$$\mu_t^{\mathrm{keep}} > 0$$

for the relevant interval, or an information-theoretic replacement when the workspace state is stochastic. Structural tracking, state tracking, and content retention are therefore separate gates.

A finite-horizon persistence certificate may be recorded as

$$\boxed{\mathcal{P}_{0:T} = \left(\max_{t<T} q_t^{\mathrm{str}}, \max_{t\le T} e_t, \min_{t<T} \mu_t^{\mathrm{keep}}, \sum_{t<T} \sigma_t \right).}$$

No single coordinate of this vector subsumes the others.

33. Interpretation of the Contraction Parameter

The inequality

$$\alpha_t < 1$$

means that deviations already outside the current workspace are damped by the next update. Mechanisms that can reduce α_t or σ_t include

verification and error feedback,
shared notation and stable interfaces,
artifact persistence and version control,
proof checking and experimental replication,
handoff reconstruction and redundancy.

Sources of disturbance include

context loss and representational drift,
miscommunication and transfer-map error,
participant or model replacement,
artifact corruption,
unsupported inference and objective drift.

33.1 Four persistence levels

The theory distinguishes:

structural persistence	q_t^{str} remains small,
state persistence	e_t remains small,
content persistence	μ_t^{keep} remains bounded away from zero,
operational persistence	predictive and causal capacities remain matched.

The first three are developed in this part. Operational matching is incorporated into the identity theory of Part XII.

33.2 Moving-line example

Let $\mathcal{C}_V(t) = \mathbb{R}^2$ and let P_t project onto

$$u_t = (\cos\theta_t, \sin\theta_t).$$

If U_t^V is rotation through $\theta_{t+1} - \theta_t$ and $F_t = U_t^V$, then a state $x_t = u_t$ obeys

$$e_t = e_{t+1} = 0, \qquad \mu_t^{\mathrm{keep}} = 1.$$

The workspace changes raw coordinates but not abstract identity. If instead $F_t = I$, then

$$e_{t+1} = |\sin(\theta_{t+1} - \theta_t)|.$$

The same structural motion now produces a state-tracking defect because the actual update fails to transport the content with the workspace.

PART XI

Reconstruction and Resilience

Persistence can be maintained by direct transport, but it can also be restored from distributed traces after a disruption. Reconstruction theory quantifies when a coalition contains enough information to recover the abstract core and how noise is amplified by that recovery.

34. Core Encoding Maps

Let C_* be a finite-dimensional abstract core. The exact static theory provides a complete local chart

$$\overline{E}_i = \mathrm{ev}_i \circ \Psi : C_* \longrightarrow C_i,$$

where $\Psi : C_* \to \Gamma(H)$ is the root-to-section isomorphism. A participant or artifact need not expose that complete chart. Let

$$O_i : C_i \longrightarrow \mathsf{Y}_i$$

be the declared observation or access map and define the *accessible encoding*

$$\boxed{E_i = O_i \overline{E}_i : C_* \longrightarrow \mathsf{Y}_i.}$$

For notational economy, the direct sums below use the accessible spaces Y_i. Thus injectivity of a full structural chart does not make a singleton operationally reconstructive unless its access map preserves the whole core.

For a coalition $S \subseteq V$, define the analysis operator

$$A_S : C_* \longrightarrow \bigoplus_{i \in S} \mathsf{Y}_i, \qquad A_S c = (E_i c)_{i \in S},$$

and the coalition frame operator

$$\boxed{F_S = A_S^* A_S = \sum_{i \in S} E_i^* E_i.}$$

Set

$$\alpha_S = \lambda_{\min}(F_S), \qquad \beta_S = \lambda_{\max}(F_S).$$

Thus

$$\alpha_S \|c\|^2 \le \|A_S c\|^2 \le \beta_S \|c\|^2.$$

Proposition 85 (Coalition reconstruction criterion)**.** *For a coalition S, the following are equivalent:*

1. *A_S is injective;*
2. *F_S is positive definite;*
3. *$\alpha_S > 0$;*
4. *there exists a linear decoder R_S with $R_S A_S = I_{C_*}$.*

When these conditions hold, the canonical decoder is

$$\boxed{R_S = A_S^\dagger = F_S^{-1} A_S^*.}$$

Proof. In finite dimensions,

$$\ker F_S = \ker A_S$$

because

$$\langle c, F_S c \rangle = \|A_S c\|^2.$$

This proves the equivalence of the first three statements. An injective linear map admits a left inverse, while a left inverse implies injectivity. The formula for the Moore–Penrose left inverse follows from full column rank. □

The lower frame bound α_S measures recoverability. The upper bound β_S measures amplification in the forward encoding. Their ratio

$$\kappa_S = \sqrt{\frac{\beta_S}{\alpha_S}}$$

is the coalition condition number.

35. Stable Reconstruction Condition

35.1 Stable reconstruction theorem T_7

Theorem 86 (Stable core reconstruction)**.** *Assume $\alpha_S > 0$ and observe*

$$y = A_S c + \eta.$$

With the canonical estimate

$$\widehat{c} = A_S^{\dagger} y,$$

one has

$$\boxed{\|\widehat{c} - c\| \le \frac{\|\eta\|}{\sqrt{\alpha_S}}.}$$

Proof. Since $A_S^{\dagger} A_S = I_{C_*}$,

$$\widehat{c} - c = A_S^{\dagger} \eta.$$

The singular values of A_S are the square roots of the eigenvalues of F_S, so

$$\|A_S^{\dagger}\| = \frac{1}{\sigma_{\min}(A_S)} = \frac{1}{\sqrt{\alpha_S}}.$$

□

Corollary 87 (Encoding mismatch)**.** *If the true observation is*

$$y = (A_S + \Delta_S) c + \eta$$

but the nominal decoder $A_S^{\dagger}$ is used, then

$$\boxed{\|\widehat{c} - c\| \le \frac{\|\Delta_S\|\,\|c\| + \|\eta\|}{\sqrt{\alpha_S}}.}$$

Proof. Now

$$\widehat{c} - c = A_S^\dagger \Delta_S c + A_S^\dagger \eta,$$

and the result follows from Theorem 86. □

35.2 Reconstruction as operational recovery

Geometric recovery becomes behaviorally relevant when the workspace state controls a predictive kernel. Let

$$K_w = \mathbb{P}(Y \in \cdot \mid W = w)$$

and suppose

$$\mathrm{TV}(K_w, K_{w'}) \le L_K d_W(w, w').$$

Proposition 88 (Exact reconstruction implies coalition sufficiency)**.** *Let W be sufficient for Y relative to the full history $\mathbf{H}$:*

$$Y \perp\!\!\!\perp \mathbf{H} \mid W.$$

If a coalition reconstructs the workspace exactly,

$$W = R_S(\mathbf{H}^S) \qquad \textit{almost surely},$$

then

$$\boxed{Y \perp\!\!\!\perp \mathbf{H} \mid \mathbf{H}^S.}$$

Under logarithmic loss,

$$\boxed{\mathcal{R}_{\log}(S) = \mathcal{R}_{\log}(V).}$$

Proof. Sufficiency means that the conditional target law given $\mathbf{H}$ is a measurable function of W. Exact reconstruction makes that function measurable with respect to $\mathbf{H}^S$, which is the asserted conditional independence. The log-risk equality follows from Theorem 66. □

Proposition 89 (Reconstruction-to-risk transfer)**.** *Let W be sufficient for a target Y, let losses take values in $[0,1]$, and let a coalition construct $\widehat{W}$ from its local data. Then its optimal Bayes risk satisfies*

$$\boxed{0 \le \mathcal{R}(S) - \mathcal{R}(W) \le 2L_K\, \mathbb{E}[d_W(W, \widehat{W})].}$$

Proof. For every pair $w, \widehat{w}$, using the Bayes action for $K_{\widehat{w}}$ under the true law K_w incurs conditional regret at most

$$2\,\mathrm{TV}(K_w, K_{\widehat{w}}).$$

This follows by inserting and subtracting the risk under $K_{\widehat{w}}$ for both the chosen action and the Bayes action for K_w. Average the bound and use the Lipschitz assumption. The coalition's optimal rule performs at least as well as this decoder-based rule. □

This proposition links Part XI to the causal-state theory of Part VI. A small geometric reconstruction error is operationally meaningful only when the outcome kernel changes continuously with workspace state.

36. k-Failure Resilience

For $k \geq 0$, define the worst surviving lower frame bound

$$\boxed{\alpha^{(k)} = \min_{\substack{A \subseteq V \\ |A| \leq k}} \alpha_{V \setminus A}.}$$

The sequence $\alpha^{(k)}$ is nonincreasing in k.

Theorem 90 (Uniform erasure resilience)**.** *The encoding family is exactly reconstructible after every loss of at most k nodes if and only if*

$$\boxed{\alpha^{(k)} > 0.}$$

If surviving observations have the form

$$y_{V \setminus A} = A_{V \setminus A} c + \eta_A, \qquad |A| \leq k,$$

then the canonical survivor decoder obeys the uniform estimate

$$\boxed{\|\widehat{c}_A - c\| \leq \frac{\|\eta_A\|}{\sqrt{\alpha^{(k)}}}.}$$

Proof. Exact reconstruction for a particular survivor set is equivalent to a positive lower frame bound by Proposition 85. Taking the minimum over all permitted erasures gives the first claim. The uniform noise estimate follows from Theorem 86. □

Proposition 91 (Dimensional obstruction)**.** *Let $d = \dim C_*$. If*

$$\sum_{i \in S} \dim \mathsf{Y}_i < d,$$

then $\alpha_S = 0$, so S cannot reconstruct the core. Consequently, k-failure resilience requires

$$\sum_{i \in V \setminus A} \dim \mathsf{Y}_i \geq d$$

for every $A \subseteq V$ with $|A| \leq k$.

Proof. The rank of A_S is at most the dimension of its codomain. If that dimension is below d, then A_S is not injective. □

Define the accessible block support of $c \neq 0$ by

$$\operatorname{supp}_E(c) = \{i \in V : E_i c \neq 0\}$$

and define the node-erasure distance

$$\boxed{d_{\mathrm{er}} = \min_{c \neq 0} |\operatorname{supp}_E(c)|.}$$

Theorem 92 (Exact erasure-distance criterion)**.** *Exact reconstruction after every loss of at most k nodes is possible if and only if*

$$\boxed{d_{\mathrm{er}} > k.}$$

Proof. If $d_{\mathrm{er}} \le k$, choose $c \neq 0$ whose nonzero accessible blocks are contained in a set A with $|A| \le k$. The surviving analysis operator satisfies $A_{V\setminus A}c = 0$, so it cannot distinguish c from zero. Conversely, if some erasure set A of size at most k destroys injectivity, there exists $c \neq 0$ with $E_i c = 0$ for every $i \notin A$. Thus $\mathrm{supp}_E(c) \subseteq A$, and $d_{\mathrm{er}} \le k$. □

The distance criterion is combinatorial. The frame quantity $\alpha^{(k)}$ is quantitative. A system may tolerate k erasures exactly while amplifying noise severely because $\alpha^{(k)}$ is close to zero.

36.1 Reconstructive handoff after replacement

Suppose node r is removed and the surviving coalition is

$$S = V \setminus \{r\}.$$

A replacement node r' is to receive the encoding

$$E'_{r'} : C_* \to \mathsf{Y}'_{r'}.$$

Corollary 93 (Handoff initialization)**.** *If $\alpha_S > 0$ and the survivors observe $y = A_S c + \eta$, initialize the replacement by*

$$x^{\mathrm{init}}_{r'} = E'_{r'} A_S^{\dagger} y.$$

Then

$$\boxed{\|x^{\mathrm{init}}_{r'} - E'_{r'}c\| \le \frac{\|E'_{r'}\|\,\|\eta\|}{\sqrt{\alpha_S}}.}$$

This is the basic mathematical form of artifact-mediated or coalition-mediated handoff. It does not require the removed component to transfer its full internal state to the replacement.

37. Redundancy–Irreducibility Tension

Define the minimum reconstructing coalition size

$$\boxed{m_{\mathrm{rec}} = \min\{|S| : \alpha_S > 0\}}$$

and the maximum exact erasure tolerance

$$\boxed{k_{\mathrm{erase}} = \max\{k : \alpha^{(k)} > 0\}.}$$

The condition

$$m_{\mathrm{rec}} \ge 2$$

means that no individual encoding is sufficient for exact core recovery. The condition

$$k_{\mathrm{erase}} \ge 1$$

means that the system survives at least one node loss. These can hold simultaneously because nonlocality of encoding is not the same as fragility.

37.1 Three-node distributed-frame example

Let

$$C_* = \mathbb{R}^2$$

and define scalar node encodings

$$E_1(c_1, c_2) = c_1, \qquad E_2(c_1, c_2) = c_2, \qquad E_3(c_1, c_2) = c_1 + c_2.$$

Every singleton has rank one, so no node reconstructs the two-dimensional core. Every pair does reconstruct. For $S = \{1, 2\}$,

$$F_S = I, \qquad \alpha_S = 1.$$

For $S = \{1, 3\}$ or $S = \{2, 3\}$, the frame eigenvalues are

$$\frac{3 \pm \sqrt{5}}{2}.$$

Therefore

$$\boxed{m_{\mathrm{rec}} = 2, \qquad k_{\mathrm{erase}} = 1, \qquad d_{\mathrm{er}} = 2, \qquad \alpha^{(1)} = \frac{3 - \sqrt{5}}{2}.}$$

This realizes a workspace core that is irreducible to any singleton while remaining reconstructible after any one node fails.

37.2 Relation to predictive irreducibility

The geometric quantity m_{rec} and the predictive quantity k_0 from Part VII answer different questions. The first asks whether a coalition can reconstruct a declared core coordinate. The second asks whether a coalition is sufficient for a declared future target. Equality between them requires structural–operational alignment and regularity such as Proposition 89. Neither quantity should be substituted for the other without those assumptions.

PART XII

Identity Through Time and Replacement

Identity is treated as invariance under a structured correspondence, not as literal persistence of tokens or carriers. Exact correspondences form a groupoid. Approximate identity is measured by structural, operational, causal, and reconstructive defects that must be reported separately before they are combined.

38. Temporal Realization Maps

Let

$$\mathfrak{R} = (G, \mathbf{C}, D, \rho, \mathcal{I}, \mathcal{K})$$

and

$$\mathfrak{R}' = (G', \mathbf{C}', D', \rho', \mathcal{I}', \mathcal{K}')$$

be two realizations. A correspondence begins with a graph isomorphism $\varphi : G \to G'$. After relabeling by φ, let

$$J_i : C_i \to C_i'$$

be unitary local identifications and define

$$J_V = \bigoplus_{i \in V} J_i, \qquad J_E = \bigoplus_{e \in E} J_{t(e)}.$$

Definition 94 (Exact structural correspondence)**.** The correspondence is structurally exact when

$$\boxed{D' J_V = J_E D.}$$

It is operation exact for $a \in \mathcal{A}$ when

$$\boxed{\rho_V'(a) J_V = J_V \rho_V(a).}$$

Theorem 95 (Exact replacement isomorphism)**.** *Under an exact structural correspondence,*

$$\boxed{\Gamma(D') = J_V \Gamma(D), \qquad L' J_V = J_V L.}$$

Consequently, every spectral workspace subspace is transported exactly:

$$P_I(L') = J_V P_I(L) J_V^*$$

for every spectral interval I. If the correspondence is also operation exact, then the transported workspace actions are intertwined exactly.

Proof. For every c,

$$D' J_V c = J_E D c.$$

Because J_E is invertible, $Dc = 0$ if and only if $D' J_V c = 0$, proving the kernel statement. Taking adjoints of the exact covariance identity gives

$$D'^* J_E = J_V D^*.$$

Therefore

$$L'J_V = D'^*D'J_V = D'^*J_ED = J_VD^*D = J_VL.$$

The spectral and operation claims follow by functional calculus and the intertwining assumptions. □

Exact correspondences compose and invert. They therefore form a groupoid of realizations. The abstract exact workspace is an isomorphism class in this groupoid.

39. Structural Identity Defect

For an approximate correspondence define

$$\boxed{\Delta_D = D'J_V - J_ED, \qquad \delta_D = \|\Delta_D\|.}$$

If the graph weights are unchanged, define edgewise comparison defects

$$\delta_e = \|H'_eJ_{s(e)} - J_{t(e)}H_e\|.$$

The discrepancy operator has one changed source block per edge, so

$$\boxed{\delta_D^2 \le \max_{i\in V} \sum_{e:s(e)=i} w_e\delta_e^2 \le \sum_{e\in E} w_e\delta_e^2.}$$

When a single node is replaced, every nonincident edge has $\delta_e = 0$. The replacement defect is therefore locally measurable even though its effect on the global workspace is controlled by the global spectral gap.

The corresponding carried-Laplacian defect obeys

$$\boxed{\eta_L := \|J_V^*L'J_V - L\| \le 2\|D\|\delta_D + \delta_D^2.}$$

This is Proposition 79 with the time indices removed.

Let P and P' be rank-d workspace projectors selected from L and L', and let $g > 0$ be the separating source gap.

Theorem 96 (Approximate structural replacement)**.** *If*

$$\eta_L < \frac{g}{2},$$

then

$$\boxed{q_{\mathrm{str}} := \|P' - J_VPJ_V^*\| \le \frac{\eta_L}{g-\eta_L} \le \frac{2\eta_L}{g}.}$$

If $q_{\mathrm{str}} < 1$*, then*

$$B = P'J_VP : \operatorname{ran} P \longrightarrow \operatorname{ran} P'$$

is bijective and satisfies

$$\sigma_{\min}(B) \ge \sqrt{1-q_{\mathrm{str}}^2}.$$

Its polar factor

$$\boxed{\beta = B(B^*B)^{-1/2}}$$

is a canonical unitary workspace correspondence. Moreover,

$$\boxed{\|\beta - J_VP\| \le q_{\mathrm{str}} + 1 - \sqrt{1-q_{\mathrm{str}}^2}.}$$

Proof. The projector estimate is Theorem 43 applied after carrying L' back by J_V. For unit $w \in \operatorname{ran} P$,

$$\|(I - P')J_V w\| \le q_{\mathrm{str}}.$$

Orthogonality gives

$$\|Bw\|^2 = 1 - \|(I - P')J_V w\|^2 \ge 1 - q_{\mathrm{str}}^2.$$

Thus B is injective, hence bijective because the source and target have the same finite dimension. Its singular values lie in $[\sqrt{1 - q_{\mathrm{str}}^2}, 1]$, so polar normalization moves B by at most $1 - \sqrt{1 - q_{\mathrm{str}}^2}$. Finally,

$$\|B - J_V P\| = \|(I - P')J_V P\| \le q_{\mathrm{str}}.$$

□

39.1 Transport of workspace operations

Write

$$A = \rho_V(a), \qquad A' = \rho'_V(a),$$

and define

$$\delta_\rho(a) = \|A'J_V - J_V A\|, \qquad \ell_a = \|(I - P)AP\|.$$

Let

$$A_{\mathcal{W}} = PAP, \qquad A'_{\mathcal{W}'} = P'A'P'.$$

Proposition 97 (Operation transport bound)**.** *For $B = P'J_V P$,*

$$\boxed{\|A'_{\mathcal{W}'}B - BA_{\mathcal{W}}\| \le \|A'\|q_{\mathrm{str}} + \delta_\rho(a) + \ell_a.}$$

For the unitary polar correspondence β,

$$\boxed{\begin{aligned}\|A'_{\mathcal{W}'}\beta - \beta A_{\mathcal{W}}\| \le{}& \|A'\|q_{\mathrm{str}} + \delta_\rho(a) + \ell_a \\ &+ (\|A'_{\mathcal{W}'}\| + \|A_{\mathcal{W}}\|)\left(1 - \sqrt{1 - q_{\mathrm{str}}^2}\right).\end{aligned}}$$

Proof. Insert and subtract $P'A'J_V P$ and $P'J_V AP$:

$$\begin{aligned}A'_{\mathcal{W}'}B - BA_{\mathcal{W}} = {}& P'A'(P' - I)J_V P \\ &+ P'(A'J_V - J_V A)P \\ &+ P'J_V(I - P)AP.\end{aligned}$$

Bound the three terms. Replacing B by β adds

$$A'_{\mathcal{W}'}(\beta - B) - (\beta - B)A_{\mathcal{W}},$$

which gives the second estimate. □

40. Behavioral Identity Defect

Structural correspondence does not determine operational identity. Let

$$\{\iota_{w,b} : (w, b) \in \mathcal{D}\} \quad \text{and} \quad \{\iota'_{\beta w, \chi b} : (w, b) \in \mathcal{D}\}$$

be matched feasible implementation-intervention families in the sense of Part VIII. Here b denotes declared background variables and χ is the background correspondence. Define

$$K(w,b) = \mathbb{P}^{\iota_{w,b}}(Y^+ \in \cdot),$$

$$K'(\beta w, \chi b) = \mathbb{P}'^{\iota'_{\beta w,\chi b}}(Y'^+ \in \cdot).$$

For a chosen divergence D_Y, define

$$\boxed{\delta_{\text{causal}}(\beta) = \sup_{(w,b)\in\mathcal{D}} D_Y\left(K(w,b), K'(\beta w, \chi b)\right).}$$

The observational response is included only when the null intervention belongs to the declared family. This definition compares actual implementation interventions rather than formally assigning an arbitrary projection coordinate.

A reconstructive handoff may also be imperfect. Let

$$\delta_{\text{rec}}(\beta) = \mathbb{E}[d_{\mathcal{W}'}(W'_{\text{init}}, \beta W)]$$

measure replacement initialization error. Corollary 93 provides one explicit upper bound on this term.

40.1 Structural identity does not imply causal identity

Take two realizations with identical

$$G, \quad D, \quad L, \quad P, \quad \rho.$$

In the first, let $Y = W$. In the second, let $Y' = 1 - W$ for a binary workspace coordinate. Then

$$q_{\text{str}} = 0, \qquad \delta_\rho(a) = 0$$

for every operation, while the observational and interventional outcome kernels differ maximally under the identity correspondence. Therefore no purely structural theorem can prove causal identity without an additional kernel-matching assumption.

41. Identity Distance and Replacement Invariance

For an admissible correspondence $\mathfrak{J}$, define the operation defect on a declared unit ball $\mathcal{A}_1$ by

$$\delta_{\text{op}}(\mathfrak{J}) = \sup_{a\in\mathcal{A}_1} \|A'_{\mathcal{W}'}\beta - \beta A_{\mathcal{W}}\|.$$

The primary certification object is the identity profile

$$\boxed{\mathbf{d}_{\text{id}}(\mathfrak{J}) = \left(q_{\text{str}}, \delta_{\text{op}}(\mathfrak{J}), \delta_{\text{causal}}(\beta), \delta_{\text{rec}}(\beta); \mu^{\text{keep}}\right).}$$

The first four coordinates must remain below declared tolerances, while the non-collapse margin μ^{keep} must remain above a declared lower bound. This vector, rather than a weighted average, determines whether a specific identity claim passes.

A composite identity cost is

$$\boxed{\text{Cost}(\mathfrak{J}) = a\, q_{\text{str}} + b\, \delta_{\text{op}}(\mathfrak{J}) + c\, \delta_{\text{causal}}(\beta) + d\, \delta_{\text{rec}}(\beta),}$$

where $a, b, c, d > 0$ are declared scales. Define

$$\boxed{d_{\mathrm{id}}(\mathfrak{R}, \mathfrak{R}') = \inf_{\mathfrak{J}} \mathrm{Cost}(\mathfrak{J}).}$$

This quantity is a task-relative optimization surrogate until the admissible correspondence class, intervention matching, divergence, and scales are fixed. It does not replace the coordinatewise identity profile or the non-collapse gate.

41.1 Replacement invariance theorem T_{14}

Theorem 98 (Quantitative replacement invariance)**.** *Let $\mathfrak{J}$ be a unitary graph correspondence with diagram defect δ_D. Define*

$$\eta_L = 2\|D\|\delta_D + \delta_D^2$$

and assume

$$\eta_L < \frac{g}{2}.$$

Set

$$q_* = \frac{\eta_L}{g - \eta_L}.$$

Then there exists a unitary workspace correspondence β for which

$$q_{\mathrm{str}} \le q_*.$$

Moreover,

$$\boxed{\begin{aligned}\delta_{\mathrm{op}}(\mathfrak{J}) \le \sup_{a \in \mathcal{A}_1} \Big[&\|\rho'_V(a)\| q_* + \delta_\rho(a) + \ell_a \\ &+ \left(\|A'_{\mathcal{W}'}\| + \|A_{\mathcal{W}}\|\right)\left(1 - \sqrt{1 - q_*^2}\right)\Big].\end{aligned}}$$

If, in addition,

$$\delta_{\mathrm{causal}}(\beta) \le \delta_c, \qquad \delta_{\mathrm{rec}}(\beta) \le \delta_r,$$

then

$$\boxed{\begin{aligned}d_{\mathrm{id}}(\mathfrak{R}, \mathfrak{R}') \le &a\, q_* + b\, \delta_{\mathrm{op}}(\mathfrak{J}) \\ &+ c\, \delta_c + d\, \delta_r.\end{aligned}}$$

In particular, if all local structural, operation, closure, causal, and reconstruction defects are $O(\delta_{\mathrm{rep}})$, the relevant operator norms are uniformly bounded, and $g \ge g_0 > 0$, then

$$\boxed{d_{\mathrm{id}}(\mathfrak{R}, \mathfrak{R}') \le C_1 \frac{\delta_{\mathrm{rep}}}{g_0} + C_2 \delta_{\mathrm{rep}} + O\left(\frac{\delta_{\mathrm{rep}}^2}{g_0}\right).}$$

After a fixed normalization of the non-spectral terms, this recovers the schematic target

$$d_{\mathrm{id}} \lesssim C \frac{\delta_{\mathrm{rep}}}{g_0}.$$

Proof. The structural bound and existence of β follow from Theorem 96. The operation estimate is Proposition 97 with $q_{\mathrm{str}} \le q_*$. Insert these estimates and the assumed causal and reconstructive bounds into the definition of $\mathrm{Cost}(\mathfrak{J})$. The asymptotic statement follows from

$$q_* = O\left(\frac{\delta_D}{g_0} + \frac{\delta_D^2}{g_0}\right)$$

and

$$1 - \sqrt{1 - q_*^2} = O(q_*^2).$$

□

Corollary 99 (Preservation of coalition frame bounds)**.** *Identify the old and new abstract cores with* $\operatorname{ran} P$ *and* $\operatorname{ran} P'$ *through their chosen core parameterizations. Let* S *be a declared coalition, let*

$$A_S : \operatorname{ran} P \to \mathsf{Y}_S, \qquad A'_S : \operatorname{ran} P' \to \mathsf{Y}'_S$$

be the resulting accessible coalition encoders, and let K_S *be a unitary identification of their observation spaces. Suppose*

$$\|A'_S\beta - K_S A_S\| \le \delta_S^{\mathrm{frame}} < \sqrt{\alpha_S}.$$

Then the replacement coalition remains reconstructive and its new lower frame bound satisfies

$$\boxed{\alpha'_S \ge \left(\sqrt{\alpha_S} - \delta_S^{\mathrm{frame}}\right)^2.}$$

Consequently, k-failure resilience survives whenever the bound holds uniformly for every coalition remaining after at most k erasures.

Proof. For $w' \in \operatorname{ran} P'$, write $w' = \beta w$. Since β and K_S are unitary,

$$\begin{aligned}\|A'_S w'\| &\ge \|K_S A_S w\| - \|(A'_S\beta - K_S A_S)w\| \\ &\ge \left(\sqrt{\alpha_S} - \delta_S^{\mathrm{frame}}\right)\|w'\|.\end{aligned}$$

Square the inequality. □

Corollary 100 (Exact carrier replacement)**.** *If*

$$\delta_D = 0, \qquad \delta_\rho(a) = 0, \qquad \ell_a = 0, \qquad \delta_c = 0, \qquad \delta_r = 0,$$

then

$$d_{\mathrm{id}}(\mathfrak{R}, \mathfrak{R}') = 0.$$

Thus a participant, model, artifact, or notation system may be replaced without changing the abstract workspace when the replacement realizes the same structural, operational, causal, and reconstructive organization.

41.2 Temporal chains and identity paths

For a sequence of realizations

$$\mathfrak{R}_0, \mathfrak{R}_1, \dots, \mathfrak{R}_T,$$

let $\mathfrak{J}_t : \mathfrak{R}_t \to \mathfrak{R}_{t+1}$ be admissible correspondences. Define the path length

$$\mathrm{Len}(\mathfrak{J}_{0:T}) = \sum_{t=0}^{T-1} \mathrm{Cost}(\mathfrak{J}_t)$$

and the induced directed path quasi-pseudometric

$$\boxed{d^{\rightarrow}_{\mathrm{path}}(\mathfrak{R}_0, \mathfrak{R}_T) = \inf_{\mathfrak{J}_{0:T}} \mathrm{Len}(\mathfrak{J}_{0:T}).}$$

Concatenation gives the directed triangle inequality. Because replacement costs and intervention divergences need not be symmetric, symmetry is not assumed. When reverse correspondences exist, define the symmetrized pseudometric

$$\boxed{d^{\mathrm{sym}}_{\mathrm{path}}(\mathfrak{R}, \mathfrak{R}') = \max\left\{d^{\rightarrow}_{\mathrm{path}}(\mathfrak{R}, \mathfrak{R}'), d^{\rightarrow}_{\mathrm{path}}(\mathfrak{R}', \mathfrak{R})\right\}.}$$

Distinct physical realizations can have zero symmetrized pseudodistance when an exact correspondence exists in both directions. This is intentional: identity is assigned to the invariant relational organization rather than to the carrier tokens.

41.3 Identity failure modes

A complete replacement audit must report at least

$$\boxed{\mathcal{I}_{\mathrm{rep}} = \left(q_{\mathrm{str}}, \delta_{\mathrm{op}}, \delta_{\mathrm{causal}}, \delta_{\mathrm{rec}}, \mu^{\mathrm{keep}}\right).}$$

The failure modes are different:

q_{str} large	the shared subspace changed,
δ_{op} large	the same operations are no longer realized,
δ_{causal} large	interventions no longer have matching consequences,
δ_{rec} large	the new realization was not initialized from the old core,
μ^{keep} small	content collapsed despite apparent coherence.

The theorem does not permit one successful coordinate to conceal failure of another.

PART XIII

Gluing and Workspaces of Workspaces

The preceding parts treat one interaction diagram at a time. A large collaborative system is usually assembled from overlapping local laboratories, subteams, artifacts, models, or task-specific workspaces. This part asks when those local workspaces determine one coherent global workspace, how failed gluing is detected, and whether recursive assembly changes the resulting object. The exact theory is a descent theorem. The approximate theory is a second transfer-Laplacian construction on the cover itself.

42. Local Workspace Covers

42.1 Edge-complete covers

Let

$$G = (V, E)$$

be a finite directed interaction graph with local core spaces C_i, core transfers H_e, vertex space $\mathcal{C}_V(G)$, discrepancy operator D_G, and exact workspace

$$\mathscr{W}(G) := \Gamma(H|_G) = \ker D_G.$$

Definition 101 (Finite workspace cover)**.** A finite family of subgraphs

$$\mathcal{U} = \{G_\alpha\}_{\alpha \in A}$$

is an *edge-complete workspace cover* of G when

$$V = \bigcup_{\alpha \in A} V_\alpha, \qquad E = \bigcup_{\alpha \in A} E_\alpha.$$

Every edge in E_α carries the transfer inherited from G. For a nonempty finite index set

$$I = \{\alpha_0, \dots, \alpha_p\} \subseteq A,$$

define the intersection subgraph

$$G_I = G_{\alpha_0 \cdots \alpha_p} := \bigcap_{\alpha \in I} G_\alpha$$

by intersecting both vertex and edge sets. Isolated vertices are retained.

Covering only the vertices is insufficient. If an edge belongs to no chart, its transfer equation is absent from every local workspace and cannot be recovered by overlap agreement. Edge completeness is therefore the minimal cover assumption used below.

If $K \subseteq L \subseteq G$ are subgraphs, coordinate restriction gives

$$\rho_K^L : \mathcal{C}_V(L) \longrightarrow \mathcal{C}_V(K).$$

Because every edge equation on K is also an edge equation on L, restriction maps exact sections to exact sections:

$$r_K^L := \rho_K^L|_{\mathscr{W}(L)} : \mathscr{W}(L) \longrightarrow \mathscr{W}(K).$$

They obey

$$r_L^L = I, \qquad r_J^K r_K^L = r_J^L$$

whenever $J \subseteq K \subseteq L$. Thus $K \mapsto \mathscr{W}(K)$ is a contravariant system of local workspaces.

42.2 The cover nerve and overlap interfaces

The nerve $N(\mathcal{U})$ has one vertex for each chart G_α and a p-simplex

$$[\alpha_0 \cdots \alpha_p]$$

whenever $G_{\alpha_0 \cdots \alpha_p}$ is nonempty. The nerve records which local workspaces can compare states directly. It does not replace the underlying interaction graph: G carries the internal transfer equations, while $N(\mathcal{U})$ carries overlap-consistency equations between already constructed local objects.

For pairwise overlaps write

$$\mathscr{W}_\alpha = \mathscr{W}(G_\alpha), \qquad \mathscr{W}_{\alpha\beta} = \mathscr{W}(G_{\alpha\beta}),$$

and

$$r_{\alpha\beta}^\alpha : \mathscr{W}_\alpha \to \mathscr{W}_{\alpha\beta}, \qquad r_{\alpha\beta}^\beta : \mathscr{W}_\beta \to \mathscr{W}_{\alpha\beta}.$$

The interface is therefore a cospan

$$\mathscr{W}_\alpha \xrightarrow{r_{\alpha\beta}^\alpha} \mathscr{W}_{\alpha\beta} \xleftarrow{r_{\alpha\beta}^\beta} \mathscr{W}_\beta,$$

not necessarily a direct transfer map from one complete local workspace to the other. Both charts are compared through the content they expose on their common domain.

42.3 Čech cochains of local workspaces

For $p \geq 0$, define the finite cochain space

$$\check{C}^p(\mathcal{U}; \mathscr{W}) = \bigoplus_{\alpha_0 < \cdots < \alpha_p} \mathscr{W}(G_{\alpha_0 \cdots \alpha_p}),$$

where empty intersections contribute the zero space. The coboundary

$$\delta^p : \check{C}^p \longrightarrow \check{C}^{p+1}$$

is

$$\boxed{(\delta^p s)_{\alpha_0 \cdots \alpha_{p+1}} = \sum_{j=0}^{p+1} (-1)^j r_{\alpha_0 \cdots \widehat{\alpha_j} \cdots \alpha_{p+1}}^{\alpha_0 \cdots \alpha_{p+1}} s_{\alpha_0 \cdots \widehat{\alpha_j} \cdots \alpha_{p+1}}.}$$

The restriction identities imply

$$\delta^{p+1} \delta^p = 0.$$

For a local family $w = (w_\alpha)_\alpha$,

$$\boxed{(\delta^0 w)_{\alpha\beta} = r_{\alpha\beta}^\beta w_\beta - r_{\alpha\beta}^\alpha w_\alpha.}$$

Thus δ^0 is the overlap-discrepancy operator at the level of exact local workspaces.

43. Compatibility and Gluing Obstructions

43.1 Direct overlap mismatch

A family $w = (w_\alpha)_\alpha$ is exactly compatible when

$$\delta^0 w = 0.$$

Its direct mismatch is

$$\boxed{\mathfrak{m}_{\mathcal{U}}(w) = \delta^0 w.}$$

After choosing inner products and positive overlap weights, the mismatch energy is

$$\boxed{\mathcal{E}_{\mathrm{ov}}(w) = \|\delta^0 w\|^2 = \sum_{\alpha<\beta} \nu_{\alpha\beta} \left\| r^{\beta}_{\alpha\beta} w_\beta - r^{\alpha}_{\alpha\beta} w_\alpha \right\|^2.}$$

This is the appropriate obstruction when the full local states are already given. Such a family glues exactly if and only if its mismatch vector itself vanishes. No cohomological quotient is needed for that direct question.

43.2 Relative-only and affine descent data

A different problem arises when only relative transition data are supplied. Let

$$b = (b_{\alpha\beta}) \in \check{C}^1(\mathcal{U}; \mathscr{W})$$

be a prescribed family of overlap offsets. The triple-overlap consistency condition is

$$\delta^1 b = 0.$$

One seeks local potentials $q = (q_\alpha) \in \check{C}^0$ satisfying

$$\delta^0 q = b.$$

Define

$$\boxed{\check{H}^1(\mathcal{U}; \mathscr{W}) = \frac{\ker \delta^1}{\operatorname{im} \delta^0}.}$$

Then the affine gluing problem is solvable precisely when

$$\boxed{[b] = 0 \quad \text{in} \quad \check{H}^1(\mathcal{U}; \mathscr{W}).}$$

A nonzero class means that all local transition statements can be mutually consistent on triple overlaps while still failing to arise from any family of local workspace states. This is a cover-level analogue of nontrivial holonomy.

43.3 Harmonic representative of the obstruction

Equip the finite cochain spaces with Hilbert inner products and define the first Čech Laplacian

$$\boxed{\Delta_1 = \delta^0(\delta^0)^* + (\delta^1)^* \delta^1.}$$

Let

$$\mathcal{H}^1_{\mathcal{U}} = \ker \Delta_1.$$

Proposition 102 (Harmonic gluing obstruction)**.** *For every cocycle $b \in \ker \delta^1$, there is an orthogonal decomposition*

$$\boxed{b = \delta^0 q^\dagger + h, \qquad h \in \mathcal{H}^1_{\mathcal{U}},}$$

where $q^\dagger = (\delta^0)^\dagger b$ is a least-squares potential. Moreover,

$$\boxed{\inf_q \|b - \delta^0 q\| = \|h\| = \|P_{\mathcal{H}^1_{\mathcal{U}}} b\|.}$$

Hence b is exactly integrable if and only if its harmonic component vanishes.

Proof. The finite-dimensional Hodge decomposition gives

$$\check{C}^1 = \operatorname{im} \delta^0 \oplus \ker \Delta_1 \oplus \operatorname{im}(\delta^1)^*.$$

If $b \in \ker \delta^1$, then b is orthogonal to $\operatorname{im}(\delta^1)^*$. Its remaining two orthogonal components are therefore an exact part $\delta^0 q^\dagger$ and a harmonic part h. The orthogonal projection theorem gives the least-squares identity. □

The scalar

$$\omega_{\mathcal{U}}(b) := \|P_{\mathcal{H}^1_{\mathcal{U}}} b\|$$

is an irreducible overlap-frustration measure. Its vanishing is invariant under admissible invertible coordinate changes, although its numerical norm depends on the chosen metric. It cannot be removed by adding local potentials alone.

43.4 Cycle example

Consider three charts arranged in a cycle, with one-dimensional overlap spaces and oriented offsets

$$b_{12}, \qquad b_{23}, \qquad b_{31}.$$

Local potentials would have to satisfy

$$q_2 - q_1 = b_{12}, \qquad q_3 - q_2 = b_{23}, \qquad q_1 - q_3 = b_{31}.$$

Adding the equations gives the necessary and sufficient condition

$$\boxed{b_{12} + b_{23} + b_{31} = 0.}$$

With the standard edge norm, the harmonic obstruction has magnitude

$$\boxed{\omega_{\mathcal{U}}(b) = \frac{|b_{12} + b_{23} + b_{31}|}{\sqrt{3}}.}$$

Every pairwise relation may be individually meaningful while the complete cycle remains globally inconsistent.

44. Exact Gluing Theorem T8

Let

$$R_{\mathcal{U}} : \mathscr{W}(G) \longrightarrow \check{C}^0(\mathcal{U}; \mathscr{W})$$

be the restriction map

$$R_{\mathcal{U}}(w) = (w|_{G_\alpha})_{\alpha \in A}.$$

Theorem 103 (Exact workspace gluing)**.** *For every finite edge-complete workspace cover $\mathcal{U}$, the sequence*

$$\boxed{0 \longrightarrow \mathscr{W}(G) \xrightarrow{R_{\mathcal{U}}} \check{C}^0(\mathcal{U};\mathscr{W}) \xrightarrow{\delta^0} \check{C}^1(\mathcal{U};\mathscr{W})}$$

is exact. Equivalently,

$$\boxed{\mathscr{W}(G) \cong \ker \delta^0 \cong \mathrm{Eq}\left(\prod_{\alpha} \mathscr{W}_\alpha \rightrightarrows \prod_{\alpha<\beta} \mathscr{W}_{\alpha\beta}\right).}$$

Every pairwise compatible family of exact local workspace states therefore glues to a unique exact global workspace state.

Proof. The cover contains every vertex, so $R_{\mathcal{U}}$ is injective. A global section has identical restrictions on every overlap, hence $\delta^0 R_{\mathcal{U}} = 0$.

Conversely, let $w = (w_\alpha)_\alpha \in \ker \delta^0$. For each vertex $i \in V$, choose any chart G_α containing i and set

$$c_i = (w_\alpha)_i.$$

If another chart G_β contains i, then $i \in G_{\alpha\beta}$ and overlap compatibility gives $(w_\alpha)_i = (w_\beta)_i$. Thus c_i is well defined.

For every edge $e : i \to j$, edge completeness provides a chart G_α containing e. Since $w_\alpha \in \mathscr{W}(G_\alpha)$,

$$c_j = (w_\alpha)_j = H_e(w_\alpha)_i = H_e c_i.$$

Hence $c = (c_i) \in \mathscr{W}(G)$, and by construction $R_{\mathcal{U}} c = w$. Uniqueness follows because the charts cover all vertices. □

Corollary 104 (Zeroth descent cohomology)**.** *For every finite edge-complete cover,*

$$\boxed{\mathscr{W}(G) \cong \check{H}^0(\mathcal{U};\mathscr{W}) = \ker \delta^0.}$$

In finite dimensions,

$$\boxed{\dim \mathscr{W}(G) = \sum_{\alpha \in A} \dim \mathscr{W}_\alpha - \operatorname{rank}(\delta^0).}$$

The theorem does not require the nerve to be acyclic. Cycles in the cover are allowed; their effect is encoded in the rank of δ^0 and, for relative-only data, in $\check{H}^1$. It also shows that the local-workspace assignment is a genuine sheaf on the finite subgraph cover used here rather than merely a presheaf.

44.1 Local nontriviality does not imply global nontriviality

Let G be the directed triangle

$$1 \xrightarrow{a} 2 \xrightarrow{b} 3 \xrightarrow{c} 1$$

with one-dimensional vertex cores. Cover it by its three edge charts. Their local workspaces are all one-dimensional:

$$\mathscr{W}_{12} = \{(x, ax) : x \in \mathbb{R}\},$$
$$\mathscr{W}_{23} = \{(y, by) : y \in \mathbb{R}\},$$
$$\mathscr{W}_{31} = \{(z, cz) : z \in \mathbb{R}\}.$$

Overlap agreement requires

$$y = ax, \qquad z = by, \qquad x = cz.$$

Therefore

$$(1 - abc)x = 0,$$

and

$$\boxed{\dim \mathscr{W}(G) = \begin{cases} 1, & abc = 1, \\ 0, & abc \neq 1. \end{cases}}$$

The meta-level gluing obstruction is exactly the original cycle holonomy. Strong local workspaces can therefore coexist with a trivial global core.

45. Approximate Gluing and Meta-Workspaces

45.1 The cover-level transfer Laplacian

First assume each local state already lies in its exact local workspace. Define

$$B_{\mathcal{U}} := \delta^0 : \check{C}^0(\mathcal{U}; \mathscr{W}) \longrightarrow \check{C}^1(\mathcal{U}; \mathscr{W})$$

and the meta-Laplacian

$$\boxed{L_{\mathcal{U}}^{\mathrm{meta}} = B_{\mathcal{U}}^* B_{\mathcal{U}}.}$$

By Theorem 103,

$$\boxed{\ker L_{\mathcal{U}}^{\mathrm{meta}} = R_{\mathcal{U}} \mathscr{W}(G).}$$

Thus the gluing problem is itself a workspace problem whose "nodes" are local workspaces and whose interfaces are overlaps.

Let

$$\gamma_{\mathcal{U}}^{\mathrm{meta}} = \lambda_{\min}^{+}\left(L_{\mathcal{U}}^{\mathrm{meta}}\right)$$

be the first positive eigenvalue.

Proposition 105 (Approximate overlap gluing)**.** *For every $w \in \check{C}^0(\mathcal{U}; \mathscr{W})$,*

$$\boxed{\operatorname{dist}\left(w, R_{\mathcal{U}} \mathscr{W}(G)\right)^2 \leq \frac{\|B_{\mathcal{U}} w\|^2}{\gamma_{\mathcal{U}}^{\mathrm{meta}}}.}$$

Proof. Decompose w into its component in $\ker L_{\mathcal{U}}^{\mathrm{meta}}$ and its orthogonal complement. On the complement, $L_{\mathcal{U}}^{\mathrm{meta}} \geq \gamma_{\mathcal{U}}^{\mathrm{meta}} I$. Therefore

$$\|B_{\mathcal{U}} w\|^2 = \langle w, L_{\mathcal{U}}^{\mathrm{meta}} w \rangle \geq \gamma_{\mathcal{U}}^{\mathrm{meta}} \operatorname{dist}\left(w, \ker L_{\mathcal{U}}^{\mathrm{meta}}\right)^2.$$

□

The gap $\gamma_{\mathcal{U}}^{\mathrm{meta}}$ measures how strongly the cover penalizes disagreement. A small overlap mismatch need not imply a nearby global workspace when this gap is itself very small.

45.2 Simultaneous local and overlap error

Local candidates may also violate their internal transfer equations. Define the local-copy space

$$\widehat{\mathcal{C}}_{\mathcal{U}} = \bigoplus_{\alpha \in A} \mathcal{C}_V(G_\alpha)$$

and the block internal discrepancy

$$\widehat{D}_{\mathcal{U}} = \bigoplus_{\alpha \in A} D_{G_\alpha}.$$

For $u = (u_\alpha)_\alpha \in \widehat{\mathcal{C}}_{\mathcal{U}}$, define the raw overlap synchronization operator

$$(S_{\mathcal{U}} u)_{\alpha\beta} = \rho^{G_\beta}_{G_{\alpha\beta}} u_\beta - \rho^{G_\alpha}_{G_{\alpha\beta}} u_\alpha.$$

For an overlap weight $\lambda_{\mathrm{ov}} > 0$, set

$$\boxed{K_{\mathcal{U}} = \begin{pmatrix} \widehat{D}_{\mathcal{U}} \\ \sqrt{\lambda_{\mathrm{ov}}}\, S_{\mathcal{U}} \end{pmatrix}, \qquad L^{\mathrm{tot}}_{\mathcal{U}} = K^*_{\mathcal{U}} K_{\mathcal{U}}.}$$

Its energy is

$$\boxed{\begin{aligned} \mathcal{E}^{\mathrm{tot}}_{\mathcal{U}}(u) = &\sum_\alpha \|D_{G_\alpha} u_\alpha\|^2 \\ &+ \lambda_{\mathrm{ov}} \sum_{\alpha<\beta} \left\| \rho^{G_\beta}_{G_{\alpha\beta}} u_\beta - \rho^{G_\alpha}_{G_{\alpha\beta}} u_\alpha \right\|^2. \end{aligned}}$$

Let

$$J_{\mathcal{U}} : \mathcal{C}_V(G) \to \widehat{\mathcal{C}}_{\mathcal{U}}, \qquad J_{\mathcal{U}} c = (c|_{G_\alpha})_\alpha,$$

be the duplication map.

Theorem 106 (Total descent kernel)**.** *For every finite edge-complete cover,*

$$\boxed{\ker K_{\mathcal{U}} = J_{\mathcal{U}} \mathscr{W}(G).}$$

Consequently, if

$$\gamma^{\mathrm{tot}}_{\mathcal{U}} = \lambda^+_{\min}\left(L^{\mathrm{tot}}_{\mathcal{U}}\right),$$

then

$$\boxed{\mathrm{dist}\left(u, J_{\mathcal{U}} \mathscr{W}(G)\right)^2 \le \frac{\mathcal{E}^{\mathrm{tot}}_{\mathcal{U}}(u)}{\gamma^{\mathrm{tot}}_{\mathcal{U}}}.}$$

Proof. If $u = J_{\mathcal{U}} c$ for a global exact section, every local discrepancy and every overlap discrepancy vanishes. Conversely, if $K_{\mathcal{U}} u = 0$, then every u_α is an exact local section and the family agrees on all overlaps. Theorem 103 gives a unique $c \in \mathscr{W}(G)$ with $u = J_{\mathcal{U}} c$. The distance estimate follows from the spectral lower bound on the orthogonal complement of the kernel. □

This theorem combines two distinct failure modes:

$$\text{local incoherence} \qquad \text{and} \qquad \text{cross-chart disagreement}.$$

Neither can be hidden by the other because they occupy separate blocks of $K_{\mathcal{U}}$.

45.3 Least-squares global assembly

Let

$$m_i = |\{\alpha : i \in V_\alpha\}|$$

be the multiplicity of vertex i. Then

$$J_{\mathcal{U}}^* J_{\mathcal{U}} = M_{\mathcal{U}}, \qquad (M_{\mathcal{U}} c)_i = m_i c_i.$$

The pseudoinverse performs coordinatewise averaging:

$$\boxed{(J_{\mathcal{U}}^\dagger u)_i = \frac{1}{m_i} \sum_{\alpha : i \in V_\alpha} (u_\alpha)_i.}$$

Let

$$\widehat{u} = P_{\ker K_{\mathcal{U}}} u = J_{\mathcal{U}} \widehat{c}.$$

Then

$$\widehat{c} = J_{\mathcal{U}}^\dagger \widehat{u}$$

and

$$\boxed{\|u - J_{\mathcal{U}} \widehat{c}\| \le \frac{\|K_{\mathcal{U}} u\|}{\sqrt{\gamma_{\mathcal{U}}^{\mathrm{tot}}}}.}$$

If $m_{\min} = \min_i m_i$, then

$$\boxed{\left\|\widehat{c} - J_{\mathcal{U}}^\dagger u\right\| \le \frac{\|K_{\mathcal{U}} u\|}{\sqrt{m_{\min} \gamma_{\mathcal{U}}^{\mathrm{tot}}}}.}$$

The simple overlap average and the globally coherent least-squares assembly are close only when the total descent energy is small relative to the cover gap.

45.4 A workspace of workspaces

Define the exact meta-workspace

$$\boxed{\mathscr{W}_{\mathrm{meta}}(\mathcal{U}) := \ker B_{\mathcal{U}} \subseteq \bigoplus_\alpha \mathscr{W}_\alpha.}$$

Theorem 103 gives the canonical identification

$$\boxed{\mathscr{W}_{\mathrm{meta}}(\mathcal{U}) \cong \mathscr{W}(G).}$$

Thus exact formation of a workspace from local workspaces does not create an additional mysterious carrier. It gives a second presentation of the same global compatible-section object.

Suppose local operations

$$A_\alpha : \mathscr{W}_\alpha \to \mathscr{W}_\alpha$$

and overlap operations

$$A_{\alpha\beta} : \mathscr{W}_{\alpha\beta} \to \mathscr{W}_{\alpha\beta}$$

satisfy

$$r_{\alpha\beta}^\alpha A_\alpha = A_{\alpha\beta} r_{\alpha\beta}^\alpha, \qquad r_{\alpha\beta}^\beta A_\beta = A_{\alpha\beta} r_{\alpha\beta}^\beta.$$

Then, with $A^0 = \bigoplus_\alpha A_\alpha$ and $A^1 = \bigoplus_{\alpha<\beta} A_{\alpha\beta}$,

$$B_{\mathcal{U}} A^0 = A^1 B_{\mathcal{U}}.$$

Therefore

$$A^0 \ker B_{\mathcal{U}} \subseteq \ker B_{\mathcal{U}}.$$

The operation-closure theorem from Part III reappears at the meta level: operations on subworkspaces descend globally precisely when they agree on interfaces.

45.5 Descent-faithful compression

A practical hierarchy rarely transmits each complete local workspace. Let

$$Q^p : \check{C}^p(\mathcal{U}; \mathscr{W}) \to \overline{C}^p$$

be compression maps into smaller local and overlap representations. Suppose there is a compressed coboundary $\overline{\delta}^p$ such that

$$\boxed{Q^{p+1}\delta^p = \overline{\delta}^p Q^p.}$$

The maps $Q^\bullet$ then form a cochain map.

Definition 107 (Descent-faithful compression)**.** The compression is *descent-faithful in degree zero* when the induced map

$$\boxed{H^0(Q) : \ker \delta^0 \longrightarrow \ker \overline{\delta}^0}$$

is an isomorphism.

Injectivity prevents two distinct global workspace states from collapsing to the same compressed meta-state. Surjectivity prevents the compressed hierarchy from introducing spurious compatible states that have no full-resolution lift. Under this condition,

$$\boxed{\ker \overline{\delta}^0 \cong \mathscr{W}(G).}$$

This is the exact criterion for a compressed workspace of workspaces to retain the global core.

45.6 Structural and operational gluing are distinct

Exact or approximate overlap coherence does not by itself establish global predictive sufficiency. Let W_t denote a glued structural state and let

$$Z_t^{\mathrm{loc}} = (Z_{\alpha,t})_{\alpha\in A}$$

be local predictive summaries recovered from its chart restrictions.

Proposition 108 (Local-to-global predictive alignment)**.** *Assume W_t is measurable with respect to the complete history $\mathbf{H}_t$, every $Z_{\alpha,t}$ is a measurable function of $W_t|_{G_\alpha}$, and*

$$I(Y_t^+; \mathbf{H}_t \mid Z_t^{\mathrm{loc}}) \le \varepsilon.$$

Then

$$\boxed{I(Y_t^+; \mathbf{H}_t \mid W_t) \le \varepsilon.}$$

In particular, exact sufficiency of the complete local-summary tuple implies exact sufficiency of the glued workspace.

Proof. The tuple Z_t^{loc} is a measurable function of W_t, while W_t is itself a history statistic. Conditioning on the finer statistic cannot increase the residual information that the full history carries about the target. Equivalently, use the conditional-law characterization from Theorem 56. □

Causal gluing requires an additional interface gate: interventions described in two overlapping charts must induce matched outcome laws on the overlap. A structurally compatible family can fail this gate, just as structurally identical replacement systems can fail causal identity in Part XII.

45.7 Anti-centralization gate

The isomorphism

$$\mathscr{W}_{\mathrm{meta}}(\mathcal{U}) \cong \mathscr{W}(G)$$

is a compositional statement, not a relational-irreducibility statement. A single chart may still expose enough accessible information to reconstruct the whole meta-state. A claim that the hierarchy is genuinely distributed must therefore reuse the coalition gates from Parts VII and XI. At the chart level, one may require, for example,

$$\alpha_{\{\alpha\}}^{\mathrm{meta}} = 0 \quad \text{for every } \alpha, \qquad \alpha_A^{\mathrm{meta}} > 0 \quad \text{for some } |A| \geq 2,$$

or the predictive analogue

$$k_\varepsilon^{\mathrm{meta}} \geq 2, \qquad \Delta_{\mathrm{joint}}^{\mathrm{meta}} > 0.$$

Capacity-matched local-memory and chart-ablation controls are still required. Hierarchical notation cannot convert a centralized copy into a relational workspace.

46. Hierarchical Collapse Theorem T9

46.1 Finite cover hierarchies

A finite workspace hierarchy is a rooted tree $\mathcal{T}$. Each node $a \in \mathcal{T}$ carries a subgraph G_a. The root r carries

$$G_r = G.$$

For every internal node a, its children

$$\mathrm{Ch}(a)$$

form a finite edge-complete cover of G_a. Leaves carry primitive local workspace spaces

$$W_\ell = \mathscr{W}(G_\ell).$$

For an internal node, define recursively

$$\boxed{W_a = \ker\left(\delta_a^0 : \bigoplus_{b \in \mathrm{Ch}(a)} W_b \longrightarrow \bigoplus_{b<c} W_{bc}\right),}$$

where the overlap spaces and restrictions are inherited from the child cover.

Theorem 109 (Exact hierarchical collapse)**.** *For every finite workspace hierarchy as above, there is a canonical isomorphism*

$$\boxed{W_a \cong \mathscr{W}(G_a)}$$

for every node a. In particular,

$$\boxed{W_r \cong \mathscr{W}(G).}$$

Therefore recursively gluing local workspaces, using any other finite parenthesization built from edge-complete intermediate covers that retain all required overlap objects, and solving the flattened global transfer system all produce the same exact workspace up to canonical isomorphism.

Proof. Proceed upward from the leaves. The claim holds at every leaf by definition. Assume it holds for all children of an internal node a. Replacing each child space W_b by its canonical copy $\mathscr{W}(G_b)$, the recursive definition of W_a is exactly the compatible-family kernel for the edge-complete child cover of G_a. Theorem 103 therefore gives

$$W_a \cong \mathscr{W}(G_a).$$

Finite induction reaches the root. □

46.2 Flattening and limit interchange

Let $L(\mathcal{T})$ be the leaf set. Repeated edge completeness implies that the flattened family

$$\mathcal{U}_{\mathrm{flat}} = \{G_\ell : \ell \in L(\mathcal{T})\}$$

is itself an edge-complete cover of G. Let

$$\delta^0_{\mathrm{flat}}$$

be its overlap coboundary and define

$$W_{\mathrm{flat}} = \ker \delta^0_{\mathrm{flat}}.$$

Theorem 103 and Theorem 109 give

$$\boxed{W_r \cong W_{\mathrm{flat}} \cong \mathscr{W}(G).}$$

Equivalently,

$$\boxed{\mathrm{Lim}^{\mathrm{rec}}_{\mathcal{T}} \left(\{\mathscr{W}(G_\ell)\}_{\ell \in L(\mathcal{T})}\right) \cong \mathrm{Eq}_{\mathcal{U}_{\mathrm{flat}}} \left(\{\mathscr{W}(G_\ell)\}_{\ell \in L(\mathcal{T})}\right).}$$

The notation on the left means recursive finite descent along the tree; the right side means one-shot equalization over the flattened leaf cover. In finite-dimensional vector spaces, both are kernels of the same complete family of local and overlap constraints, grouped in different orders. Hierarchical collapse is therefore a finite-limit interchange property, not a claim that intermediate levels are physically absent or operationally irrelevant.

46.3 Approximate hierarchical collapse

At level r, let K_r be the corresponding total descent operator,

$$\gamma_r = \lambda^+_{\min}(K_r^* K_r), \qquad \|K_r u_r\| \le \varepsilon_r.$$

The nearest exact compatible state then differs from u_r by at most

$$\frac{\varepsilon_r}{\sqrt{\gamma_r}}.$$

Allow an additional implementation, compression, or causal-summary defect ζ_r, and define

$$\delta_r = \frac{\varepsilon_r}{\sqrt{\gamma_r}} + \zeta_r.$$

Theorem 110 (Approximate hierarchical collapse bound)**.** *Let*

$$F_r : \mathcal{H}_r \to \mathcal{H}_{r+1}, \qquad \widetilde{F}_r : \mathcal{H}_r \to \mathcal{H}_{r+1}$$

be exact and implemented level-collapse maps for $r = 0, \dots, R-1$. Let $\mathcal{R}_r \subseteq \mathcal{H}_r$ contain the exact and implemented states reachable at level r. Assume, on $\mathcal{R}_r$,

$$\|F_r(x) - F_r(y)\| \le L_r \|x - y\|$$

and

$$\sup_{x \in \mathcal{R}_r} \|\widetilde{F}_r(x) - F_r(x)\| \le \delta_r.$$

Then

$$\boxed{\begin{aligned}&\left\|\widetilde{F}_{R-1} \circ \cdots \circ \widetilde{F}_0(x) - F_{R-1} \circ \cdots \circ F_0(x)\right\| \\ &\qquad \le \sum_{r=0}^{R-1} \left(\prod_{s=r+1}^{R-1} L_s \right) \delta_r.\end{aligned}}$$

In particular, if $L_r \le 1$, then the total error is at most

$$\sum_{r=0}^{R-1} \delta_r.$$

If $L_r \le L < 1$ and $\delta_r \le \delta$, then

$$\boxed{\text{total error} \le \frac{\delta(1 - L^R)}{1 - L} \le \frac{\delta}{1 - L}.}$$

Proof. Let x_r and $\widetilde{x}_r$ be the exact and implemented states after r levels. Then

$$\begin{aligned}\|\widetilde{x}_{r+1} - x_{r+1}\| &\le \|\widetilde{F}_r(\widetilde{x}_r) - F_r(\widetilde{x}_r)\| \\ &\quad + \|F_r(\widetilde{x}_r) - F_r(x_r)\| \\ &\le \delta_r + L_r \|\widetilde{x}_r - x_r\|.\end{aligned}$$

Iterating this recurrence gives the displayed product-sum estimate. The final two bounds follow by summing a finite or infinite geometric series. □

This theorem identifies three regimes:

$L_r < 1$	hierarchical correction is contractive,
$L_r = 1$	level errors accumulate at most additively,
$L_r > 1$	deep hierarchies can amplify early mismatch.

A deep hierarchy is therefore stable only when its level gaps remain separated and its assembly maps do not magnify small overlap errors faster than those errors are corrected.

46.4 Causal-kernel accumulation

Let K_r and $\widetilde{K}_r$ now denote exact and implemented Markov kernels at successive hierarchy levels, and suppose

$$\sup_x \mathrm{TV}\left(\widetilde{K}_r(x, \cdot), K_r(x, \cdot)\right) \le \zeta_r.$$

Because Markov kernels contract total variation, a telescoping argument gives

$$\boxed{\sup_x \mathrm{TV}\left(\widetilde{K}_0 \cdots \widetilde{K}_{R-1}(x,\cdot), K_0 \cdots K_{R-1}(x,\cdot)\right) \le \sum_{r=0}^{R-1} \zeta_r.}$$

Thus causal semantics can also be transported through a hierarchy, but only under an explicit kernel-level compatibility test. Structural gluing alone provides no such guarantee.

46.5 Direct versus hierarchical computation

The exact theorem states

$$\boxed{\text{recursive exact gluing} = \text{one-shot exact gluing.}}$$

The approximate theorem instead supplies a measurable flattening defect:

$$\boxed{\delta_{\mathrm{flat}}^{(R)} = \left\| \widetilde{F}_{R-1} \cdots \widetilde{F}_0(x) - F_{R-1} \cdots F_0(x) \right\|.}$$

A hierarchy is empirically justified when it reduces computation, communication, or local model complexity while keeping $\delta_{\mathrm{flat}}^{(R)}$, predictive risk, causal-kernel mismatch, and anti-centralization gates within declared tolerances. Hierarchy by itself is not evidence of emergence.

46.6 Gluing and hierarchical synthesis

The exact and approximate constructions can be compressed as

$$\boxed{\begin{aligned} &\text{local transfer systems} \xrightarrow{\ker D_\alpha} \text{local workspaces} \\ &\qquad\qquad\qquad\qquad \xrightarrow{\ker \delta^0} \text{meta-workspace} \\ &\qquad\qquad\qquad\qquad \xrightarrow{\text{finite recursive descent}} \text{global workspace.} \end{aligned}}$$

At exact resolution, all finite parenthesizations are canonically equivalent. At approximate resolution, every level contributes a defect scaled by its spectral gap and by downstream amplification. Relative-only overlap data can carry a harmonic H^1 obstruction even when every local chart is internally coherent. Operational and relational claims still require predictive, causal, coalition, and null-model gates in addition to structural descent.

PART XIV

Scaffolded and Endogenous Formation

47. Formation Systems and Scaffold Specification

The preceding parts begin with a transfer system that has already been chosen. A formation theory must instead explain how the nodes, edges, transfer maps, workspace channels, and operational boundaries are obtained. This requires a strict separation between the structure supplied by the experimenter and the structure selected from interaction history.

Definition 111 (Formation system)**.** A *workspace formation system* is a tuple

$$\mathfrak{F} = (\Sigma, \mathcal{D}_{\text{form}}, \mathcal{L}, \Omega, \widehat{\theta}, \mathcal{D}_{\text{cert}}, \mathcal{I}_{\text{form}}),$$

where:

- Σ is the externally supplied scaffold;
- $\mathcal{D}_{\text{form}}$ is the formation or training history;
- $\mathcal{L}$ is the learning or selection rule;
- Ω is its internal random seed or stochastic state;
- $\widehat{\theta} = \mathcal{L}(\Sigma, \mathcal{D}_{\text{form}}, \Omega)$ is the learned organization;
- $\mathcal{D}_{\text{cert}}$ is held-out certification data; and
- $\mathcal{I}_{\text{form}}$ is a declared family of interventions on scaffold and formation history.

The learned organization is decomposed as

$$\widehat{\theta} = (\widehat{V}, \widehat{E}, \widehat{H}, \widehat{P}, \widehat{\mathcal{O}}, \widehat{B}),$$

where $\widehat{B}$ is a learned operational boundary.

The scaffold is itself stratified:

$$\boxed{\Sigma = (\Sigma_V, \Sigma_E, \Sigma_H, \Sigma_P, \Sigma_{\mathcal{O}}, \Sigma_B, \Sigma_Y).}$$

The final coordinate Σ_Y records labels, targets, privileged masks, or other supervision that could directly reveal the intended workspace. This coordinate is necessary because an architecture may appear to discover a boundary even when the desired membership mask was encoded in its training labels.

47.1 Scaffolded architecture versus scaffolded solution

A scaffold can constrain the search space without specifying its solution. For example, a graph neural architecture can fix the class of admissible local message-passing rules while leaving the active edges and transfer maps to be learned. Conversely, a system can use a flexible architecture while receiving labels that directly encode the desired workspace. The former is architecturally scaffolded but solution-open. The latter is solution-scaffolded.

For any component q of $\widehat{\theta}$, define its scaffold-only Bayes risk under loss ℓ_q by

$$\mathcal{R}_q^{\Sigma} = \inf_f \mathbb{E}\left[\ell_q(q, f(\Sigma, \Omega))\right].$$

Define the full formation risk by

$$\mathcal{R}_q^{\Sigma,\mathcal{H}} = \inf_g \mathbb{E}\left[\ell_q(q, g(\Sigma, \mathcal{D}_{\text{form}}, \Omega))\right].$$

The history advantage is

$$\boxed{\Delta_q^{\text{hist}} = \mathcal{R}_q^{\Sigma} - \mathcal{R}_q^{\Sigma,\mathcal{H}}.}$$

A positive history advantage means that formation history contains information about q that is not already recoverable from the declared scaffold and random seed.

Proposition 112 (Scaffold-only impossibility)**.** *If*

$$q = f(\Sigma, \Omega) \qquad \textit{almost surely,}$$

then for every formation-history intervention that leaves (Σ, Ω) *fixed,*

$$\mathcal{L}(q \mid \text{do}(\mathcal{D}_{\text{form}} = h)) = \mathcal{L}(q \mid \text{do}(\mathcal{D}_{\text{form}} = h')),$$

and $\Delta_q^{\text{hist}} = 0$.

Proof. The value of q is measurable with respect to (Σ, Ω), so an intervention on formation history cannot change its conditional law when those variables are held fixed. The scaffold-only predictor f attains the same risk as any predictor that also receives formation history, hence the risk gap vanishes. □

This proposition is elementary but methodologically decisive. A structure that is deterministically encoded by the scaffold cannot count as discovered from interaction history merely because a learned network reproduces it.

47.2 Formation and certification split

The same sample must not be used both to select a workspace and to certify its coherence, causal efficacy, or relational irreducibility. The required split is

$$\mathcal{D} = \mathcal{D}_{\text{form}} \sqcup \mathcal{D}_{\text{cal}} \sqcup \mathcal{D}_{\text{cert}}.$$

The formation sample selects $\widehat{\theta}$. The calibration sample fixes thresholds, regularization strengths, and intervention magnitudes. The certification sample evaluates the final gates without further structural selection. In a multi-environment study, the stronger requirement is

$$\mathcal{E}_{\text{form}} \cap \mathcal{E}_{\text{cert}} = \varnothing,$$

so that the learned organization is tested under environments not used to construct it.

Definition 113 (Formation leakage)**.** A certification variable Y_{cert} leaks into formation when the conditional law of $\widehat{\theta}$ changes under an intervention on Y_{cert} that should be inaccessible during formation:

$$\mathcal{L}\left(\widehat{\theta} \mid \text{do}(Y_{\text{cert}} = y)\right) \neq \mathcal{L}\left(\widehat{\theta} \mid \text{do}(Y_{\text{cert}} = y')\right).$$

A valid discovery claim requires a documented information barrier or an audit showing that this effect is absent up to tolerance.

47.3 Origin and reality are orthogonal

The formation provenance of a workspace and its operational certification are separate coordinates. A fully scaffolded structure may satisfy every structural and causal workspace gate. A highly endogenous structure may fail to become coherent or causally useful. Symbolically,

$$\text{Endogenous}(W) \not\Longrightarrow \text{SRW}_\Theta(W),$$

$$\text{SRW}_\Theta(W) \not\Longrightarrow \text{Endogenous}(W).$$

The theory therefore treats endogeneity as a provenance axis rather than a higher level on the workspace hierarchy.

48. Operational Endogeneity Signature

The ideal description-complexity expression from the symbolic outline is retained as a conceptual limit, but raw Kolmogorov complexity is not computable. A usable theory must declare a model class, a code, and an intervention family.

Definition 114 (Code-relative history contribution)**.** Let $\mathfrak{M}_q$ be a declared coding or predictive model class for a learned component q. Let

$$L_{\mathfrak{M}_q}(q \mid Z)$$

be its held-out codelength given side information Z. Define

$$\boxed{e_q^{\text{MDL}} = \left[\frac{L_{\mathfrak{M}_q}(q \mid \Sigma) - L_{\mathfrak{M}_q}(q \mid \Sigma, \mathcal{D}_{\text{form}})}{\max\{L_{\mathfrak{M}_q}(q \mid \Sigma), 1\}}\right]_0^1,}$$

where $[x]_0^1 = \min\{1, \max\{0, x\}\}$.

This quantity is explicitly code-relative. It must not be presented as an observer-independent metaphysical measure. Its role is to test whether the formation history provides compressive information about the learned organization beyond the declared scaffold.

Definition 115 (Causal formation sensitivity)**.** Let ι_h and $\iota_{h'}$ be matched feasible interventions that alter formation history while preserving the scaffold, randomization law, sample size, and intervention cost. For a divergence D, define

$$\boxed{e_q^{\text{causal}} = \mathbb{E}_{(h,h')\sim\nu_q} D(\mathcal{L}^{\iota_h}(q), \mathcal{L}^{\iota_{h'}}(q))\,.}$$

A common null uses temporally or relationally shuffled histories that preserve local marginals but destroy the candidate formation signal.

Proposition 116 (Reparameterization invariance of causal provenance)**.** *Let ϕ be a measurable bijection on the range of q, and let the divergence D be invariant under measurable bijections. Then*

$$e_{\phi(q)}^{\text{causal}} = e_q^{\text{causal}}.$$

Proof. Every intervention-induced law of $\phi(q)$ is the pushforward of the corresponding law of q. Bijection invariance of D preserves each pairwise divergence, and therefore their design-weighted expectation. □

The causal coordinate is therefore insensitive to a mere renaming of the learned structure. The MDL coordinate need not be invariant unless the coding class is transformed with the reparameterization.

48.1 Scaffold dependence and cross-scaffold invariance

Endogeneity does not require independence from all scaffolds. Every learning system has an architecture, objective, and resource budget. The relevant question is whether the learned organization depends on the particular solution-bearing details of the scaffold.

For a family $\mathfrak{S}$ of admissible scaffold variants, define the cross-scaffold identity defect

$$\boxed{\delta_q^{\mathrm{scaf}} = \inf_{\{J_{\sigma\sigma'}\}} \mathbb{E}_{\sigma,\sigma'\sim\pi} d_q\left(J_{\sigma\sigma'} q_\sigma, q_{\sigma'}\right),}$$

where $J_{\sigma\sigma'}$ transports the learned object between admissible coordinate systems. A low value indicates that the same abstract organization is recovered under multiple non-solution-bearing scaffolds.

The complete formation signature is

$$\boxed{\mathbf{E}_{\mathrm{form}}(W) = \left(e_V, e_E, e_H, e_P, e_{\mathcal{O}}, e_B; \delta^{\mathrm{scaf}}, \delta^{\mathrm{leak}}\right),}$$

where every e_q is reported as the pair

$$e_q = \left(e_q^{\mathrm{MDL}}, e_q^{\mathrm{causal}}\right).$$

No scalar average is permitted to hide a zero history contribution for one of the structures claimed to have been discovered.

48.2 Formation certificate

For thresholds Θ_{form}, define

$$\mathrm{Form}_{\Theta_{\mathrm{form}}}(W) = 1$$

only when all applicable gates pass:

$$\begin{aligned} &e_q^{\mathrm{MDL}} \geq \tau_q^{\mathrm{MDL}}, \qquad e_q^{\mathrm{causal}} \geq \tau_q^{\mathrm{causal}}, \\ &\delta_q^{\mathrm{scaf}} \leq \varepsilon_q^{\mathrm{scaf}}, \qquad \delta^{\mathrm{leak}} \leq \varepsilon_{\mathrm{leak}}, \\ &\mathrm{SRW}_\Theta(W) = 1, \quad \mathrm{Pass}(N_{\mathrm{form}}) = 1. \end{aligned}$$

The final workspace gate is included because endogenous discovery of an unstable or causally irrelevant pattern is not endogenous workspace formation.

49. Learned Transfer Structure and Formation Classes

49.1 Gated membership and edge learning

Let

$$z_i \in \{0,1\}, \qquad a_e \in [0,1]$$

be node-membership and edge-activation variables satisfying

$$a_e \leq z_{s(e)}, \qquad a_e \leq z_{t(e)}.$$

A weighted learned discrepancy operator is

$$(D_{a,H}c)_e = \sqrt{w_e a_e}\left(c_{t(e)} - H_e c_{s(e)}\right).$$

One generic formation objective is

$$\boxed{\widehat{\theta} \in \operatorname*{arg\,min}_{z,a,H,P,\mathcal{O},B} \Big\{ \widehat{\mathcal{R}}_{\text{task}} + \lambda_{\text{coh}}\|D_{a,H}c\|^2 + \lambda_E\|a\|_1 + \lambda_V\|z\|_1 + \lambda_{\text{stab}}\mathcal{J}_{\text{stab}} + \lambda_{\text{leak}}\mathcal{J}_{\text{leak}} \Big\}.}$$

The objective is schematic rather than universal. Its purpose is to expose which quantities were fixed and which were optimized. If z and a are fixed by Σ, only the transfer maps or local channels can be claimed as learned.

49.2 Gauge-identifiable transfer learning

A learned transfer map is not absolutely identifiable when local coordinates are latent. Under local invertible changes

$$X'_u = g_u X_u, \qquad X'_v = g_v X_v,$$

the same abstract transfer is represented by

$$H'_{uv} = g_v H_{uv} g_u^{-1}.$$

The identifiable object is therefore the gauge class

$$[H]_{\mathcal{G}} = \left\{ (g_{t(e)} H_e g_{s(e)}^{-1})_{e\in E} : (g_i)_{i\in V} \in \mathcal{G} \right\}.$$

Any formation claim that compares raw matrices across independently trained systems without gauge alignment is ill posed.

49.3 Transfer identification theorem T_{15}

Theorem 117 (Population transfer identification)**.** *Let $X \in \mathbb{R}^{d_u}$ and $Y \in \mathbb{R}^{d_v}$ satisfy*

$$Y = H_* X + \xi, \qquad \mathbb{E}[\xi X^{\mathsf{T}}] = 0,$$

and assume

$$\Sigma_{XX} = \mathbb{E}[XX^{\mathsf{T}}] \succeq \kappa I$$

for some $\kappa > 0$. Then H_* is the unique minimizer of

$$\mathcal{R}(H) = \mathbb{E}\|Y - HX\|_2^2,$$

with

$$\boxed{H_* = \Sigma_{YX}\Sigma_{XX}^{-1}.}$$

Moreover,

$$\boxed{\mathcal{R}(H) - \mathcal{R}(H_*) = \operatorname{tr}\left((H - H_*)\Sigma_{XX}(H - H_*)^{\mathsf{T}}\right) \geq \kappa\|H - H_*\|_F^2.}$$

Proof. Substitute $Y = H_*X + \xi$ into the risk. The cross term vanishes because $\mathbb{E}[\xi X^\mathsf{T}] = 0$, giving

$$\mathcal{R}(H) = \mathcal{R}(H_*) + \mathbb{E}\|(H - H_*)X\|_2^2.$$

The remaining expectation is the displayed trace. Positive definiteness of Σ_{XX} gives uniqueness and the lower bound. The normal equations give the covariance formula. □

Corollary 118 (Deterministic empirical recovery bound)**.** *For samples $(X_r, Y_r)_{r=1}^n$, write*

$$\widehat{\Sigma}_{XX} = \frac{1}{n}\sum_{r=1}^n X_r X_r^\mathsf{T}, \qquad \widehat{M} = \frac{1}{n}\sum_{r=1}^n \xi_r X_r^\mathsf{T}.$$

If $\widehat{\Sigma}_{XX} \succeq \widehat{\kappa} I$, then ordinary least squares satisfies

$$\boxed{\|\widehat{H} - H_*\|_F \le \frac{\|\widehat{M}\|_F}{\widehat{\kappa}}.}$$

For ridge parameter $\tau > 0$,

$$\boxed{\|\widehat{H}_\tau - H_*\|_F \le \frac{\|\widehat{M}\|_F + \tau\|H_*\|_F}{\widehat{\kappa} + \tau}.}$$

Proof. The empirical normal equations give

$$\widehat{H} - H_* = \widehat{M}\widehat{\Sigma}_{XX}^{-1}.$$

For ridge regression,

$$\widehat{H}_\tau - H_* = (\widehat{M} - \tau H_*)(\widehat{\Sigma}_{XX} + \tau I)^{-1}.$$

Taking Frobenius and operator norms yields both estimates. □

Proposition 119 (Gauge covariance of least squares)**.** *Under invertible coordinate transformations*

$$X' = g_u X, \qquad Y' = g_v Y,$$

the population least-squares solution transforms as

$$H'_* = g_v H_* g_u^{-1}.$$

Proof. The transformed covariances are

$$\Sigma_{Y'X'} = g_v \Sigma_{YX} g_u^\mathsf{T}, \qquad \Sigma_{X'X'} = g_u \Sigma_{XX} g_u^\mathsf{T}.$$

Substitution into the covariance formula and inversion of the second factor give the result. □

The theorem establishes identification only under excitation and exogeneity. When Σ_{XX} is singular, the transfer is identified only on the excited source subspace. When $\mathbb{E}[\xi X^\mathsf{T}] \neq 0$, regression recovers a predictive map rather than necessarily the intended causal transfer.

49.4 Held-out structure-selection theorem T_{16}

The same finite-margin argument applies to node membership, edge selection, boundary selection, or a finite family of gauge-normalized transfer models.

Theorem 120 (Uniform structure-selection recovery)**.** *Let $\mathfrak{T}$ be a finite candidate family with population score $J : \mathfrak{T} \to \mathbb{R}$. Let*

$$\theta_* = \arg\min_{\theta \in \mathfrak{T}} J(\theta)$$

be unique, with margin

$$\Delta_J = \min_{\theta \neq \theta_*} \big(J(\theta) - J(\theta_*)\big) > 0.$$

If a held-out estimate $\widehat{J}$ satisfies

$$\sup_{\theta \in \mathfrak{T}} |\widehat{J}(\theta) - J(\theta)| \leq \varepsilon,$$

then every empirical minimizer $\widehat{\theta}$ obeys

$$\boxed{J(\widehat{\theta}) - J(\theta_*) \leq 2\varepsilon.}$$

If $2\varepsilon < \Delta_J$, then

$$\boxed{\widehat{\theta} = \theta_*.}$$

Proof. By empirical optimality and uniform approximation,

$$J(\widehat{\theta}) \leq \widehat{J}(\widehat{\theta}) + \varepsilon \leq \widehat{J}(\theta_*) + \varepsilon \leq J(\theta_*) + 2\varepsilon.$$

If $\widehat{\theta} \neq \theta_*$, the margin would force an excess larger than 2ε, a contradiction. □

The theorem does not guarantee that the candidate class contains the correct organization. It states exactly what follows once a finite class, score, and uniform certification error are declared.

49.5 Formation classes

The provenance classes are now defined relative to

$$(\Sigma, \mathcal{L}, \mathcal{I}_{\text{form}}, \mathfrak{M}).$$

They are not absolute labels.

Exogenously scaffolded.
The claimed nodes, edges, transfer maps, channels, operations, or boundary are recoverable from the scaffold alone at the declared tolerance. Their history advantages and causal formation sensitivities vanish.

Adaptively parameterized.
The membership and channel topology are fixed, while transfer parameters, local operators, or readouts are learned from interaction history. Typically

$$e_H + e_P + e_{\mathcal{O}} > 0, \qquad e_V \approx e_E \approx e_B \approx 0.$$

Structurally adaptive.
At least one of membership, edges, or boundary is selected from history, but the search family and objective remain externally specified.

Endogenously constituted relative to the scaffold class.
Membership, transfer organization, and operational boundary have nonzero history contribution and causal formation sensitivity, survive cross-scaffold transport, pass leakage and matched-null tests, and produce a certified workspace on held-out environments.

The phrase "relative to the scaffold class" is mandatory. No finite experiment can establish independence from every possible external influence.

PART XV

Boundary Formation and Stability

50. Candidate Workspace Boundaries

A boundary is not merely a subset of physical components. It is a subset relative to a declared node decomposition, prediction target, intervention family, environment distribution, and time scale.

Definition 121 (Boundary problem)**.** A boundary problem is a tuple

$$\mathfrak{B} = (V, \mathfrak{S}, \tau, \mathcal{Y}, \mathcal{I}, \Pi, \mathcal{G}_B)\,,$$

where $\mathfrak{S} \subseteq 2^V$ is the admissible candidate family, τ is the temporal scale, $\mathcal{Y}$ is the operational target, $\mathcal{I}$ is the feasible intervention family, Π is the environment distribution, and $\mathcal{G}_B$ is a symmetry group under which boundaries are considered equivalent.

Unless the application requires otherwise, the candidate family excludes

$$\varnothing, \qquad V, \qquad \{i\} \text{ for } i \in V.$$

This anti-triviality rule prevents the empty set, the entire realization, or a single centralized controller from winning solely through a degenerate score.

50.1 Observational recurrence, dependence, and outward influence

For $S \in \mathfrak{S}$, define

$$\begin{aligned}
\mathrm{Ret}_\tau(S) &= I(X_S(t+\tau); X_S(t) \mid X_{\bar{S}}(t), E_t)\,,\\
\mathrm{Dep}_\tau(S) &= I(X_S(t+\tau); X_{\bar{S}}(t) \mid X_S(t), E_t)\,,\\
\mathrm{Out}_\tau(S) &= I(X_{\bar{S}}(t+\tau); X_S(t) \mid X_{\bar{S}}(t), E_t)\,.
\end{aligned}$$

These quantities are observational. They can be nonzero because of common causes, selection, or feedback paths and therefore cannot alone establish a causal boundary.

50.2 Causal boundary terms

Let $\iota_{S,x}$ be a feasible intervention that changes the internal workspace coordinate of S to x while matching background variables and intervention cost. Define outward causal efficacy

$$\boxed{\mathrm{COut}_\tau(S) = \mathbb{E}_{x,x' \sim \nu_S} D(\mathcal{L}^{\iota_{S,x}}(X_{\bar{S}}(t+\tau)), \mathcal{L}^{\iota_{S,x'}}(X_{\bar{S}}(t+\tau)))\,.}$$

Define inward causal dependence analogously by intervening on $\bar{S}$ and measuring the future law of X_S:

$$\boxed{\mathrm{CIn}_\tau(S) = \mathbb{E}_{u,u' \sim \nu_{\bar{S}}} D(\mathcal{L}^{\iota_{\bar{S},u}}(X_S(t+\tau)), \mathcal{L}^{\iota_{\bar{S},u'}}(X_S(t+\tau)))\,.}$$

Neither term should be minimized to zero in general. A useful workspace may receive substantial external input. The boundary objective must instead express the intended balance between internal recurrence, external dependence, and outward efficacy.

50.3 Structural and anti-centralization gates

For the induced subsystem on S, define

$$\mathsf{Coh}(S) = \operatorname{tr}\left(P_{S,d} L_S P_{S,d}\right),$$

where $P_{S,d}$ projects onto the selected low-energy workspace modes. Let

$$\mathsf{Cut}(S) = \sum_{e:s(e)\in S,\ t(e)\notin S} w_e + \sum_{e:s(e)\notin S,\ t(e)\in S} w_e.$$

Let $k_\varepsilon(S)$ be the minimal sufficient coalition size inside the candidate subsystem. A candidate is admissible only if

$$\mathrm{SRW}_{\Theta_S}(W_S) = 1, \qquad k_\varepsilon(S) \geq 2,$$

unless the claim explicitly concerns a centralized workspace.

51. Boundary Objective, Recovery, and Persistence

A general score is

$$\boxed{\begin{aligned}\mathcal{B}_t(S) = {} & a_R \operatorname{Ret}_\tau(S) + a_O \operatorname{Out}_\tau(S) + a_C \operatorname{COut}_\tau(S) \\ & - a_D \operatorname{Dep}_\tau(S) - a_I \operatorname{CIn}_\tau(S) - a_L \mathsf{Coh}(S) \\ & - a_K \mathsf{Cut}(S) - a_Z \mathsf{Size}(S) - a_T \operatorname{Instability}_t(S).\end{aligned}}$$

The coefficients and all normalizations are part of the boundary problem. No single choice is universal. The selected boundary is

$$\widehat{S}_t \in \arg\max_{S\in\mathfrak{S}_{\mathrm{pass}}} \widehat{\mathcal{B}}_t(S),$$

where $\mathfrak{S}_{\mathrm{pass}}$ contains only candidates that satisfy the required workspace and anti-centralization gates.

51.1 Boundary equivalence and identifiability

If a symmetry $g \in \mathcal{G}_B$ leaves the full observational and interventional law invariant, S and gS cannot be distinguished by the declared experiment. The identifiable boundary is then the orbit

$$[S]_{\mathcal{G}_B} = \{gS : g \in \mathcal{G}_B\}.$$

A unique subset should not be claimed when only an equivalence class is identifiable.

51.2 Stable boundary recovery T_{17}

Theorem 122 (Held-out boundary recovery)**.** *Let $\mathfrak{S}_{\mathrm{pass}}$ be finite and let*

$$S_t^* = \arg\max_{S\in\mathfrak{S}_{\mathrm{pass}}} \mathcal{B}_t(S)$$

be unique. Define the population margin

$$\Delta_t^B = \mathcal{B}_t(S_t^*) - \max_{S\neq S_t^*} \mathcal{B}_t(S) > 0.$$

If

$$\sup_{S\in\mathfrak{S}_{\mathrm{pass}}} \left|\widehat{\mathcal{B}}_t(S) - \mathcal{B}_t(S)\right| \leq \varepsilon_t,$$

then every empirical maximizer $\widehat{S}_t$ *satisfies*

$$\boxed{\mathcal{B}_t(S_t^*) - \mathcal{B}_t(\widehat{S}_t) \le 2\varepsilon_t.}$$

If $2\varepsilon_t < \Delta_t^B$*, then*

$$\boxed{\widehat{S}_t = S_t^*.}$$

Proof. Apply Theorem 120 to $J(S) = -\mathcal{B}_t(S)$. □

The theorem converts uncertainty in mutual-information estimation, causal interventions, structural scores, and stability penalties into an explicit boundary-identification condition. It also exposes when the data are insufficient: a small margin makes the boundary intrinsically fragile.

51.3 Temporal boundary stability

Theorem 123 (Boundary persistence under score drift)**.** *Assume* S_t^* *is the unique maximizer of* $\mathcal{B}_t$ *with margin* Δ_t^B*, and suppose*

$$\sup_{S \in \mathfrak{S}_{\text{pass}}} |\mathcal{B}_{t+1}(S) - \mathcal{B}_t(S)| \le \delta_t.$$

If

$$2\delta_t < \Delta_t^B,$$

then

$$\boxed{S_{t+1}^* = S_t^*.}$$

More generally,

$$\mathcal{B}_{t+1}(S_{t+1}^*) - \mathcal{B}_{t+1}(S_t^*) \le 2\delta_t.$$

Proof. For every competitor $S \neq S_t^*$,

$$\begin{aligned}\mathcal{B}_{t+1}(S_t^*) - \mathcal{B}_{t+1}(S) &\ge \mathcal{B}_t(S_t^*) - \mathcal{B}_t(S) - 2\delta_t \\ &\ge \Delta_t^B - 2\delta_t > 0.\end{aligned}$$

Hence the maximizer is unchanged. The approximate bound follows by comparing both time-indexed scores at their respective maximizers. □

If the score depends on a parameter ϑ_t and satisfies

$$\sup_S |\mathcal{B}_\vartheta(S) - \mathcal{B}_{\vartheta'}(S)| \le L_B d(\vartheta, \vartheta'),$$

then the sufficient stability condition becomes

$$2L_B d(\vartheta_t, \vartheta_{t+1}) < \Delta_t^B.$$

51.4 Analyst-discovered, system-instantiated, and self-maintained boundaries

Three claims must remain separate.

Analyst-discovered boundary.
An external procedure fits $\widehat{S}$ after observing the system. The system need not represent or use that boundary.

System-instantiated boundary.
The system contains a channel, mask, or routing variable causally equivalent to $\widehat{S}$, and interventions on that variable alter downstream operation.

Self-maintained boundary.
The system detects degradation of its boundary certificate and takes actions that restore membership, transfer, or reconstruction properties.

Only the second and third claims concern a boundary realized within the system. A post hoc partition that predicts behavior well is not by itself a self-discovered boundary.

51.5 Boundary formation certificate

A boundary $\widehat{S}$ is certified as endogenously formed only if:

1. the membership mask was not directly encoded in Σ_B or Σ_Y;
2. e_B^{MDL} and e_B^{causal} exceed declared thresholds;
3. held-out recovery satisfies Theorem 122;
4. the boundary remains stable under admissible environment drift or its changes are predicted by score-margin closure;
5. matched history-shuffle and label-leakage controls fail to reproduce the result, as do full-system, empty-set, singleton, and centralized boundary nulls;
6. the induced subsystem satisfies its workspace certificate; and
7. a system-instantiation intervention demonstrates that the boundary is used by the system rather than only by the analyst.

The results below establish the statistical and deterministic stability components. They does not prove that an arbitrary adaptive system will spontaneously learn such a boundary.

PART XVI

Agency, Counterfactual Control, and Self-Maintenance

52. Controlled Workspace Agents and Goal States

The preceding parts define when a distributed structure is coherent, operationally sufficient, causally efficacious, relationally irreducible, persistent, reconstructible, and endogenously formed. None of those properties by itself establishes agency. Agency adds a controlled relation among a represented objective, a set of feasible actions, counterfactual consequences, and a policy that uses those quantities to alter the future. Self-maintenance adds a further loop in which damage to the workspace becomes an input to actions that preserve or restore the workspace certificate.

The purpose of this part is not to infer intentions from fluent behavior. It is to define an intervention-based certificate whose components can fail separately.

52.1 Controlled relational realization

A controlled relational realization extends Definition 1 by the tuple

$$\boxed{\mathfrak{A}_{\mathbf{W}} = \left(\mathfrak{R}, \mathcal{G}, \mathcal{A}_t, \mathcal{Y}^+, \{U_g\}_{g\in\mathcal{G}}, \Pi, \mathcal{I}_A, \mathcal{I}_G, \mathcal{I}_D, \mathcal{I}_R\right).}$$

Here:

- $\mathfrak{R}$ is the underlying relational realization;
- $\mathcal{G}$ is a measurable goal-code space;
- $\mathcal{A}_t$ is the feasible action set at time t;
- $\mathcal{Y}^+$ is the relevant future trajectory space;
- $U_g : \mathcal{Y}^+ \to \mathbb{R}$ evaluates a trajectory under goal code g;
- Π is an admissible class of workspace-level policies;
- $\mathcal{I}_A$ and $\mathcal{I}_G$ are feasible action and goal interventions; and
- $\mathcal{I}_D$ and $\mathcal{I}_R$ are feasible damage and repair interventions.

A controlled state contains at least

$$C_t^{\mathrm{ctl}} = (W_t, G_t, M_t, B_t, E_t)\,,$$

where G_t is the represented goal, M_t is a workspace-integrity monitor, B_t is the current boundary or routing state, and E_t is the declared external context. An action is generated by

$$A_t \sim \pi_t\left(\cdot \mid W_t, G_t, M_t, B_t, E_t\right).$$

The future controlled state obeys a time-unrolled structural model

$$S_{t+1} = F_t(S_t, A_t, E_{t+1}, \xi_{t+1}), \qquad W_{t+1} = \omega_{t+1}(S_{t+1}).$$

The intervention semantics are inherited from Part VII. A symbol such as $\mathrm{do}(A_t = a)$ is meaningful only when an implementation intervention changes an actual action mechanism while the declared background variables are matched.

52.2 Goal content and goal-code equivalence

A persistent internal variable is not automatically a goal. A goal code must index a stable counterfactual preference over relevant future trajectories. For a fixed state w, define the goal-conditioned action value

$$Q_t^g(a \mid w) = \mathbb{E}\left[U_g(Y_t^+) \mid \mathrm{do}(A_t = a), W_t = w, G_t = g\right].$$

When action costs are relevant, they are incorporated into U_g or declared as a constraint on $\mathcal{A}_t$.

Two goal codes are operationally equivalent,

$$g \sim_{\mathrm{goal}} g',$$

when they induce the same action ordering in every declared state and intervention background:

$$Q_t^g(a \mid w) \geq Q_t^g(a' \mid w) \quad \Longleftrightarrow \quad Q_t^{g'}(a \mid w) \geq Q_t^{g'}(a' \mid w).$$

The abstract goal is the equivalence class

$$[g]_{\mathrm{goal}}.$$

This quotient prevents a mere invertible recoding of a goal channel from being reported as a change of objective. For quantitative regret bounds, a representative normalization of U_g must nevertheless be fixed.

Remark 124 (Goals are claim-relative)**.** The same internal code can count as a goal for one declared trajectory family and fail to do so for another. A code that predicts a short-horizon action may not encode a persistent long-horizon objective. Every agency claim must state its horizon, outcome space, feasible actions, and utility normalization.

52.3 Local, distributed, and workspace-level goals

Let $\mathcal{H}_t^S$ be the history available to coalition $S \subseteq V$. For a loss ℓ_G, define goal-reconstruction risk

$$\mathcal{R}_G(S) = \inf_f \mathbb{E}\left[\ell_G\left(G_t, f(\mathcal{H}_t^S)\right)\right].$$

For tolerance ε, define

$$k_{G,\varepsilon} = \min\left\{|S| : \mathcal{R}_G(S) \leq \mathcal{R}_G(V) + \varepsilon\right\}.$$

A goal is locally stored when $k_{G,\varepsilon} = 1$. It is distributed relative to the declared decomposition when $k_{G,\varepsilon} \geq 2$.

This distinction is independent of whether the rest of the workspace is relational. A relational workspace can obey a goal copied in full into one controller. Such a system may possess a workspace but not a *workspace-level goal* in the strong distributed sense.

For the complete control state, define

$$\mathcal{R}_{\mathrm{ctl}}(S) = \inf_f \mathbb{E}\left[\ell_{\mathrm{ctl}}\left(C_t^{\mathrm{ctl}}, f(\mathcal{H}_t^S)\right)\right],$$

and let $k_{\mathrm{ctl},\varepsilon}$ be its minimal sufficient coalition size. A strong workspace-level agency claim requires

$$\boxed{k_{\mathrm{ctl},\varepsilon} \geq 2.}$$

This is a location criterion. It does not yet establish goal persistence, counterfactual evaluation, or self-maintenance.

52.4 Action mediation and outcome control

Goal-dependent action mediation is measured by feasible interventions on the goal channel while matching the current workspace and external context. For a divergence D_A, define

$$\eta_{G\to A} = \sup_{g,g',w,e} D_A\left(\mathbb{P}(A_t \in \cdot \mid \iota_g, w, e), \mathbb{P}(A_t \in \cdot \mid \iota_{g'}, w, e)\right).$$

Workspace-dependent action mediation is

$$\eta_{W\to A} = \sup_{w,w',g,e} D_A\left(\mathbb{P}(A_t \in \cdot \mid \iota_w, g, e), \mathbb{P}(A_t \in \cdot \mid \iota_{w'}, g, e)\right).$$

An action must also alter the declared future:

$$\eta_{A\to Y} = \sup_{a,a',b} D_Y\left(\mathbb{P}(Y_t^+ \in \cdot \mid \iota_a, b), \mathbb{P}(Y_t^+ \in \cdot \mid \iota_{a'}, b)\right).$$

A positive $\eta_{G\to A}$ without $\eta_{A\to Y}$ describes goal-sensitive output selection with no demonstrated control of the relevant environment. A positive $\eta_{A\to Y}$ without a goal or counterfactual-value channel may describe a reactive controller.

52.5 Why observational goal persistence is insufficient

The observational condition

$$I(G_{t+\tau}; G_t \mid E_{t+1:t+\tau}) > 0$$

can arise from several non-agentic mechanisms:

1. a persistent external command writes the same code at every step;
2. a clock, episode label, or task identifier drives both variables;
3. a frozen memory cell repeats a symbol that is not used by the policy;
4. selection on successful trajectories induces apparent temporal dependence; or
5. a downstream action is copied back into the nominal goal channel.

A goal-persistence claim therefore requires a cue-controlled intervention that changes G_t, removes or matches subsequent command inputs, and checks both future goal state and action consequences.

52.6 Cue-controlled goal persistence

Let C_t denote an external command or task cue. Fix a neutral or matched future command trajectory $c^0_{t+1:t+\tau}$. For goal interventions ι_g and $\iota_{g'}$, define the cue-controlled retention effect

$$\eta_G(\tau; g, g') = D_G\left(\mathbb{P}\left(G_{t+\tau} \in \cdot \mid \iota_g, \mathrm{do}(C_{t+1:t+\tau} = c^0)\right), \mathbb{P}\left(G_{t+\tau} \in \cdot \mid \iota_{g'}, \mathrm{do}(C_{t+1:t+\tau} = c^0)\right)\right).$$

The corresponding action-retention effect is

$$\eta_{G\to A}(\tau; g, g') = D_A\left(\mathbb{P}(A_{t+\tau} \in \cdot \mid \iota_g, c^0), \mathbb{P}(A_{t+\tau} \in \cdot \mid \iota_{g'}, c^0)\right).$$

The first quantity tests whether a goal difference survives. The second tests whether the surviving difference remains behaviorally operative.

Cue dependence must be reported separately. For matched initial goal g, define

$$\delta_G^{\mathrm{cue}}(\tau) = \sup_{c,c'} D_G\left(\mathbb{P}(G_{t+\tau} \in \cdot \mid \iota_g, \mathrm{do}(C = c)), \mathbb{P}(G_{t+\tau} \in \cdot \mid \iota_g, \mathrm{do}(C = c'))\right).$$

High cue dependence is not a failure when the claimed agent is intentionally instruction following. It does mean that persistence cannot be attributed to an autonomous internal objective without an additional cue-removal result.

52.7 T_{19}: Cue-controlled goal-retention theorem

Theorem 125 (Cue-controlled goal retention)**.** *Let $G_s \in \mathbb{R}^{d_G}$ obey*

$$G_{s+1} = R_s G_s + B_s C_{s+1} + \zeta_{s+1} \qquad (s = t, \dots, t+\tau-1).$$

Consider two intervention runs that set

$$G_t = g, \qquad G'_t = g',$$

and then use the same command trajectory and the same disturbance realization. Let

$$R_{t+\tau:t} = R_{t+\tau-1} \cdots R_t.$$

Then

$$\boxed{G_{t+\tau} - G'_{t+\tau} = R_{t+\tau:t}(g - g').}$$

If

$$m_s \le \sigma_{\min}(R_s), \qquad \|R_s\| \le M_s,$$

then

$$\boxed{\left(\prod_{s=t}^{t+\tau-1} m_s\right)\|g - g'\| \le \|G_{t+\tau} - G'_{t+\tau}\| \le \left(\prod_{s=t}^{t+\tau-1} M_s\right)\|g - g'\|.}$$

Proof. Subtract the two update equations. The matched command and disturbance terms cancel, giving

$$G_{s+1} - G'_{s+1} = R_s(G_s - G'_s).$$

Iteration gives the exact product formula. The upper bound follows from submultiplicativity of the operator norm. The lower bound follows from

$$\sigma_{\min}(AB) \ge \sigma_{\min}(A)\sigma_{\min}(B)$$

for square linear maps. □

The theorem isolates internal retention from repeated external rewriting. It also makes clear that persistence is graded. Uniformly contractive maps can retain a goal for a finite horizon while erasing it asymptotically. A neutral mode with product singular value near one supports long retention. An expansive mode can preserve distinctions but may be dynamically unstable.

Corollary 126 (Behavioral readout of a retained goal)**.** *Assume, in addition, that under the matched intervention background the action readout has conditional mean*

$$\mathbb{E}[A_{t+\tau} \mid G_{t+\tau}, W_{t+\tau}] = K_{t+\tau} G_{t+\tau} + b(W_{t+\tau}),$$

and that the two runs are matched on $W_{t+\tau}$. Then

$$\mathbb{E}[A_{t+\tau} \mid \iota_g] - \mathbb{E}[A_{t+\tau} \mid \iota_{g'}] = K_{t+\tau} R_{t+\tau:t}(g - g').$$

If $\sigma_{\min}(K_{t+\tau}) \geq k_{t+\tau} > 0$, the norm of this action-mean difference is at least

$$k_{t+\tau} \left(\prod_{s=t}^{t+\tau-1} m_s \right) \|g - g'\|.$$

The corollary is a sufficient condition, not a universal model of action selection. For discrete actions, the corresponding test should compare the full interventional action distributions.

52.8 Goal-persistence certificate

A minimal goal-persistence report contains:

1. the goal code and its declared trajectory-level semantics;
2. the feasible goal intervention and its manipulation check;
3. the matched or neutralized future cue trajectory;
4. $\eta_G(\tau)$ across a declared horizon grid;
5. $\eta_{G \to A}(\tau)$ over the same grid;
6. cue dependence $\delta_G^{\mathrm{cue}}(\tau)$;
7. goal-channel ablation and permutation controls;
8. coalition reconstruction risks $\mathcal{R}_G(S)$; and
9. uncertainty intervals over environments, episodes, and random seeds.

A repeated label with no intervention effect on future action fails this certificate even when its observational autocorrelation is one.

53. Counterfactual Policy, Repair, and Agency Certification

53.1 Feasible action interventions and value functions

Let $\mathcal{A}_t(w, e)$ denote the actions that can be implemented from workspace state w in environment e. The interventional action value is

$$Q_t^g(a \mid w, e) = \mathbb{E}^{\iota_a} \left[U_g(Y_t^+) \mid W_t = w, E_t = e, G_t = g \right].$$

The optimal action set is

$$\mathcal{A}_t^*(w, g, e) = \arg\max_{a \in \mathcal{A}_t(w,e)} Q_t^g(a \mid w, e).$$

A counterfactual evaluator is an internal state

$$\widehat{Q}_t = \left(\widehat{Q}_t(a)\right)_{a\in\mathcal{A}_t(w,e)}$$

that estimates these intervention values. It is not enough for $\widehat{Q}_t$ to predict the action selected by the system. It must be validated against externally estimated intervention outcomes.

53.2 Counterfactual-use criterion

Three properties must be separated:

Counterfactual accuracy.

$$\delta_Q = \sup_{a\in\mathcal{A}_t} |\widehat{Q}_t(a) - Q_t^g(a \mid w, e)|.$$

Policy use. A feasible intervention that changes $\widehat{Q}_t(a)$ while preserving the action affordances changes the action distribution in the direction predicted by the decision rule.

Outcome validity. The resulting action change alters $U_g(Y_t^+)$ as predicted by the external interventional value estimate.

A system can satisfy any two while failing the third. A learned critic can be accurate but ignored. A policy can respond to a manipulated score that does not predict outcomes. A successful action can arise from habit or replay without any counterfactual evaluator.

53.3 T_{20}: Approximate counterfactual policy theorem

Theorem 127 (Approximate counterfactual policy selection)**.** *Let $\mathcal{A}$ be a finite feasible action set. Let $Q : \mathcal{A} \to \mathbb{R}$ be the true interventional value and let $\widehat{Q}$ satisfy*

$$\sup_{a\in\mathcal{A}} |\widehat{Q}(a) - Q(a)| \le \varepsilon_Q.$$

Suppose a policy $\widehat{\pi}$ is δ_{opt}-optimal for $\widehat{Q}$:

$$\max_{\pi} \mathbb{E}_{a\sim\pi}[\widehat{Q}(a)] - \mathbb{E}_{a\sim\widehat{\pi}}[\widehat{Q}(a)] \le \delta_{\text{opt}}.$$

Then, for every true-value-optimal policy π^*,

$$\boxed{\mathbb{E}_{a\sim\pi^*}[Q(a)] - \mathbb{E}_{a\sim\widehat{\pi}}[Q(a)] \le 2\varepsilon_Q + \delta_{\text{opt}}.}$$

For a deterministic greedy action $\widehat{a} \in \arg\max_a \widehat{Q}(a)$,

$$Q(a^*) - Q(\widehat{a}) \le 2\varepsilon_Q.$$

If the true optimal action a^* is unique and its value gap

$$\Delta_Q = Q(a^*) - \max_{a\neq a^*} Q(a)$$

satisfies $\Delta_Q > 2\varepsilon_Q$, then every greedy maximizer of $\widehat{Q}$ equals a^*.

Proof. For every policy π, uniform approximation gives

$$\left|\mathbb{E}_{a\sim\pi}[\widehat{Q}(a)] - \mathbb{E}_{a\sim\pi}[Q(a)]\right| \le \varepsilon_Q.$$

Therefore

$$\begin{aligned}\mathbb{E}_{\pi^*}[Q] - \mathbb{E}_{\widehat{\pi}}[Q] &\le \mathbb{E}_{\pi^*}[\widehat{Q}] - \mathbb{E}_{\widehat{\pi}}[\widehat{Q}] + 2\varepsilon_Q \\ &\le \delta_{\mathrm{opt}} + 2\varepsilon_Q.\end{aligned}$$

The deterministic bound is the special case of point-mass policies. If $\widehat{a} \neq a^*$, then

$$\widehat{Q}(a^*) \ge Q(a^*) - \varepsilon_Q > Q(\widehat{a}) + \varepsilon_Q \ge \widehat{Q}(\widehat{a}),$$

contradicting greedy optimality. □

This theorem converts counterfactual-model accuracy into a decision guarantee. It does not establish that the model is internally represented or causally used. Those are separate intervention gates.

53.4 Counterfactual mediation and the replay null

Let V_t denote the internal value channel. A counterfactual-mediation test compares:

$$G_t \longrightarrow V_t \longrightarrow A_t \longrightarrow Y_t^+.$$

The relevant quantities include:

$$\eta_{G\to V}, \qquad \eta_{V\to A}, \qquad \eta_{A\to Y}, \qquad \delta_{G\to A|V},$$

where the final term measures the residual goal-to-action effect after the value channel is blocked or randomized.

The action-replay null preserves the empirical action sequence or action marginal while destroying online dependence on V_t. If replayed actions perform equally well under held-out perturbations, the evidence supports a cached or open-loop policy rather than online counterfactual evaluation. A second null permutes value labels among feasible actions while preserving the multiset of scores. A genuine value-mediated policy should follow the permuted ranking, while an ignored critic should not.

53.5 Workspace integrity vector and maintenance margin

Choose a declared set $\mathcal{J}_{\mathrm{maint}}$ of workspace gates that must be protected. Write each gate as a signed margin

$$\phi_{j,t} \ge 0 \quad \Longleftrightarrow \quad \text{gate } j \text{ passes at time } t.$$

Examples include

$$\begin{aligned}\phi_{\mathrm{gap},t} &= g_{\varepsilon,t} - \underline{g}, \\ \phi_{\mathrm{coh},t} &= \overline{\delta}_{\mathrm{coh}} - \delta_{\mathrm{coh},t}, \\ \phi_{\mathrm{rec},t} &= \alpha_t^{(k)} - \underline{\alpha}, \\ \phi_{\mathrm{keep},t} &= \mu_t^{\mathrm{keep}} - \underline{\mu}, \\ \phi_{\mathrm{causal},t} &= \eta_{\mathrm{causal},t} - \underline{\eta}, \\ \phi_{\mathrm{boundary},t} &= \Delta_t^B - \underline{\Delta}_B.\end{aligned}$$

The aggregate integrity margin is

$$\boxed{v_t = \min_{j \in \mathcal{J}_{\text{maint}}} \phi_{j,t}.}$$

The workspace passes the protected gates when $v_t \geq 0$. To maintain a reserve $\rho > 0$, define the maintenance deficit

$$\boxed{e_t = (\rho - v_t)_+.}$$

Define the zero-reserve and bounded-deficit viability sets

$$\boxed{\mathcal{V}_\Theta = \{S : \phi_j(S) \geq 0 \text{ for every } j \in \mathcal{J}_{\text{maint}}\}}$$

$$\boxed{\mathcal{V}_{\Theta, e_{\max}} = \{S : e(S) \leq e_{\max}\}}$$

The first set requires every protected gate to pass. The second permits a declared finite reserve deficit.

This minimum-gate construction is intentionally noncompensatory. A large spectral gap cannot compensate for failed reconstruction, and high task reward cannot compensate for loss of the declared workspace boundary.

53.6 Damage and repair interventions

A damage intervention $d \in \mathcal{I}_D$ changes a mechanism that supports one or more protected gates. Examples include:

- deleting or corrupting a transfer edge;
- removing a participant or artifact;
- perturbing the internal boundary mask;
- scrambling a role-binding or routing channel;
- erasing a handoff representation; or
- degrading a reconstruction map.

A repair intervention $r \in \mathcal{I}_R$ can add, remove, retrain, rewrite, or reroute an admissible component. Let r_0 denote matched no-repair. The post-repair integrity value at horizon τ is

$$V_t^{\text{rep}}(r; d) = \mathbb{E}\left[v_{t+\tau} \mid \text{do}(D_t = d), \text{do}(R_t = r), \mathbf{H}_t\right].$$

The repair gain is

$$\Delta_t^{\text{rep}}(r; d) = V_t^{\text{rep}}(r; d) - V_t^{\text{rep}}(r_0; d).$$

Repair costs are handled either by restricting the feasible repair family or by reporting a separate cost-effect curve. They should not be hidden inside an undocumented scalar reward.

53.7 Damage detection, repair mediation, and specificity

Let M_t be an internal integrity monitor and R_t the selected repair action. A self-maintenance claim requires the causal chain

$$D_t \longrightarrow M_t \longrightarrow R_t \longrightarrow v_{t+\tau}.$$

The required tests are:

Detection. Matched damage interventions change M_t in a gate-specific direction.

Repair mediation. Intervening on M_t changes the repair distribution, and blocking M_t reduces the damage-to-repair effect.

Restoration. The selected repair improves $v_{t+\tau}$ relative to matched no-repair.

Specificity. A cost-matched generic action or an off-workspace repair does not produce the same restoration.

Identity preservation. The restored system remains within the declared identity tolerance relative to the pre-damage workspace.

Define the detection effect $\eta_{D\to M}$, mediation effect $\eta_{M\to R}$, restoration effect $\eta_{R\to v}$, and repair specificity

$$\Delta^{\text{spec}}_{\text{repair}} = \eta_{R\to v} - \eta_{R_\perp \to v},$$

where $R_\perp$ is a matched action outside the proposed repair channel.

53.8 T_{21}: Repair-selection margin theorem

Theorem 128 (Repair selection under value-model error)**.** *Fix a damage intervention d and a finite feasible repair set $\mathcal{R}$. Let*

$$V(r) = V_t^{\text{rep}}(r; d)$$

be the true expected post-repair integrity margin and let $\widehat{V}$ obey

$$\sup_{r\in\mathcal{R}} |\widehat{V}(r) - V(r)| \le \varepsilon_R.$$

Let $\widehat{r} \in \arg\max_r \widehat{V}(r)$ and $r^* \in \arg\max_r V(r)$. Then

$$\boxed{V(r^*) - V(\widehat{r}) \le 2\varepsilon_R.}$$

If

$$V(r^*) \ge \rho + 2\varepsilon_R,$$

then the selected repair restores the expected reserve:

$$\boxed{V(\widehat{r}) \ge \rho.}$$

If r^* is unique and its repair-value gap exceeds $2\varepsilon_R$, then $\widehat{r} = r^*$.

Proof. Greedy optimality and uniform approximation give

$$\begin{aligned} V(r^*) &\le \widehat{V}(r^*) + \varepsilon_R \\ &\le \widehat{V}(\widehat{r}) + \varepsilon_R \\ &\le V(\widehat{r}) + 2\varepsilon_R. \end{aligned}$$

The reserve and exact-selection statements follow immediately. □

The theorem is conditional on a valid repair-value model. It cannot convert a post hoc repair narrative into evidence that the system detected damage or used the model. Detection and value-channel interventions remain necessary.

53.9 Repair, replacement, and successor workspaces

Restoration of performance is not automatically restoration of identity. Let W_{t^-} be the last certified pre-damage workspace and let $W^r_{t+\tau}$ be the post-repair workspace. Define the repair identity defect

$$\delta_{\mathrm{rep-id}}(r) = d_{\mathrm{id}}\left(W_{t^-}, W^r_{t+\tau}\right).$$

A repair action is identity preserving at tolerance $\varepsilon_{\mathrm{id}}$ when

$$\Delta^{\mathrm{rep}}_t(r;d) > 0, \qquad \delta_{\mathrm{rep-id}}(r) \le \varepsilon_{\mathrm{id}}.$$

When integrity is restored but the identity defect exceeds tolerance, the result is better described as a successor or replacement workspace. This terminological distinction prevents any high-performing reconstruction from being counted as persistence of the same relational organization.

53.10 T_{22}: Closed-loop self-maintenance theorem

Theorem 129 (Causal repair stabilizes maintenance deficit)**.** *Let $e_t \ge 0$ be the maintenance deficit. Suppose the matched no-repair trajectory satisfies*

$$e^{(0)}_{t+1} \le \alpha e_t + d_t,$$

where $d_t \ge 0$ is exogenous damage. Suppose the implemented repair policy has the causal one-step effect

$$e^{(0)}_{t+1} - e_{t+1} \ge \kappa e_t - \epsilon_t.$$

Assume

$$0 \le \beta := \alpha - \kappa < 1, \qquad 0 \le d_t \le \overline{d}, \qquad 0 \le \epsilon_t \le \overline{\epsilon}.$$

Then

$$\boxed{e_t \le \beta^t e_0 + \frac{(\overline{d}+\overline{\epsilon})(1-\beta^t)}{1-\beta},}$$

and therefore

$$\boxed{\limsup_{t\to\infty} e_t \le \frac{\overline{d}+\overline{\epsilon}}{1-\beta}.}$$

If the admissible deficit is $e_{\max}$ and

$$e_0 \le e_{\max}, \qquad \overline{d}+\overline{\epsilon} \le (1-\beta)e_{\max},$$

then the interval $[0, e_{\max}]$ is forward invariant.

Proof. Combine the no-repair bound with the causal repair effect:

$$\begin{aligned} e_{t+1} &\le e^{(0)}_{t+1} - \kappa e_t + \epsilon_t \\ &\le (\alpha-\kappa)e_t + d_t + \epsilon_t \\ &\le \beta e_t + \overline{d} + \overline{\epsilon}. \end{aligned}$$

Iterating this scalar recurrence gives the stated geometric bound. If $e_t \le e_{\max}$, the forward-invariance condition gives

$$e_{t+1} \le \beta e_{\max} + (1-\beta)e_{\max} = e_{\max}.$$

□

Theorem 129 differs from the persistence theorem in Part X. There, contraction was assumed for the workspace error dynamics. Here the improved contraction factor is attributed to a measured repair action through a matched no-repair counterfactual. Without that causal comparison, the same bounded trajectory could be passive robustness rather than self-maintenance.

53.11 A scalar boundary-repair example

Let $b_t \in [0,1]$ be a boundary-integrity score and let the reserve be ρ. Define

$$e_t = (\rho - b_t)_+.$$

Suppose uncorrected damage obeys

$$e_{t+1}^{(0)} \le 1.08 e_t + d_t,$$

so the open-loop boundary is locally unstable. A repair policy with measured causal effect

$$e_{t+1}^{(0)} - e_{t+1} \ge 0.25 e_t - \epsilon_t$$

gives

$$\beta = 1.08 - 0.25 = 0.83.$$

If

$$\overline{d} = 0.01, \qquad \overline{\epsilon} = 0.005,$$

then

$$\limsup e_t \le \frac{0.015}{1-0.83} \approx 0.0883.$$

Thus an open-loop unstable boundary can be maintained within a finite deficit band by a sufficiently strong repair channel. The example establishes only a control calculation. A complete self-maintenance claim must also show that an internal monitor detects damage, mediates the repair choice, and preserves workspace identity.

53.12 Complete self-maintenance certificate

A self-maintaining workspace must pass all of the following gates:

1. *Protected-gate declaration.* The maintained integrity vector and reserve are fixed before evaluation.
2. *Damage validity.* Each damage intervention measurably degrades at least one protected gate.
3. *Internal detection.* The monitor changes under damage and is not merely an external evaluator.
4. *Repair mediation.* Monitor interventions change repair selection.
5. *Repair efficacy.* Repair improves the held-out integrity margin relative to matched no-repair.
6. *Specificity.* Cost-matched generic actions and off-workspace interventions do not reproduce the gain.

7. *Repeated-disturbance stability.* The deficit remains within the declared bound over a held-out disturbance sequence.
8. *Identity control.* The restored state remains within the declared identity tolerance or is reported as a successor workspace.
9. *Anti-centralization.* No single node contains a sufficient copy of the monitor, repair model, and protected state when the claim is workspace-level maintenance.

53.13 Certified workspace agent

Let

$$\Theta_A = (\tau_G, \tau_{G\to A}, \varepsilon_Q, \tau_{V\to A}, \tau_{A\to Y}, \tau_{D\to M}, \tau_{M\to R}, \tau_{R\to v}, \tau_{\mathrm{spec}}, \varepsilon_{\mathrm{ctl}}, e_{\max}) .$$

A Θ_A-certified workspace agent is a controlled realization for which:

1. $\mathrm{SRW}(W_t;\varsigma) = 1$;
2. cue-controlled goal retention exceeds τ_G;
3. goal-to-action mediation exceeds $\tau_{G\to A}$;
4. counterfactual-value error is at most ε_Q;
5. value-to-action use exceeds $\tau_{V\to A}$;
6. action-to-outcome control exceeds $\tau_{A\to Y}$;
7. damage detection, monitor-to-repair mediation, and restoration exceed their thresholds;
8. repair specificity exceeds τ_{spec};
9. $k_{\mathrm{ctl},\varepsilon_{\mathrm{ctl}}} \geq 2$; and
10. repeated maintenance deficit remains at most $e_{\max}$.

As throughout the theory, this predicate is relative to the declared node decomposition, goals, actions, environment, horizons, interventions, and null families.

53.14 T_{23}: Workspace-agency synthesis theorem

Theorem 130 (Workspace-level agency under declared gates)**.** *Assume a controlled realization satisfies:*

1. *the operational relationality assumptions of Theorem 76;*
2. *cue-controlled goal retention and action readout as in Theorem 125 and Corollary 126;*
3. *a counterfactual evaluator with uniform error ε_Q and optimization error δ_{opt};*
4. *feasible interventions establishing positive value-to-action and action-to-outcome effects;*
5. *the causal self-maintenance assumptions of Theorem 129; and*
6. *$k_{\mathrm{ctl},\varepsilon} \geq 2$, together with a matched local policy-null panel that preserves every single-node predictive marginal but loses full-system control value.*

Then, relative to the declared specification:

1. *the expected one-step decision regret is at most* $2\varepsilon_Q + \delta_{\mathrm{opt}}$*;*
2. *the long-run maintenance deficit is bounded by*

$$\frac{\overline{d} + \overline{\epsilon}}{1 - \beta};$$

3. *no single-node statistic is sufficient for the complete control state;*
4. *matched local policy nulls cannot reproduce the intact control value;*
5. *goal-dependent action selection and workspace repair are causally mediated by implemented internal channels.*

Hence the realization satisfies the formal predicate

$$\mathrm{Agent}_{\Theta_A}(\mathbf{W}) = 1.$$

Proof. The regret bound follows from Theorem 127. The maintenance bound follows from Theorem 129. The coalition condition excludes single-node sufficiency for the complete control state. The matched-null gap excludes explanations based only on preserved local predictive marginals. The implemented goal, value, action, monitor, and repair interventions establish the required causal mediation. These are precisely the gates in the stated agency predicate. □

The synthesis theorem is intentionally relative and conjunctive. Its force is that decision quality, maintenance stability, causal mediation, and relational location are certified together. It does not infer an unrestricted agent outside the tested environment.

53.15 Agency, autonomy, and consciousness

The following implications are not licensed:

$$\mathrm{Agent}_{\Theta_A}(\mathbf{W}) \not\Longrightarrow \mathrm{PhenomenallyConscious}(\mathbf{W}),$$

$$\mathrm{SelfMaintaining}(\mathbf{W}) \not\Longrightarrow \mathrm{SelfAware}(\mathbf{W}),$$

$$\mathrm{GoalPersistent}(\mathbf{W}) \not\Longrightarrow \mathrm{MoralPatient}(\mathbf{W}).$$

A thermostat-like controller can maintain a variable without representing itself. A sophisticated planner can evaluate counterfactual actions without phenomenal experience. Conversely, failure to pass this engineering certificate would not prove absence of consciousness. Agency and consciousness remain distinct axes of inquiry.

PART XVII

Decomposition, Coarse-Graining, and Robust Relationality

Relationality is always assessed relative to a way of dividing a realization into components. That dependence is not a minor technicality. If every participant, artifact, and channel is merged into one component, the complete workspace becomes local by definition. If a centralized state is recoded into several complementary shares, the same centralized process can be made to look irreducibly distributed. A defensible theory must therefore say which partitions count as admissible, which quantities survive faithful aggregation, and how sensitive a certificate is to alternative descriptions of the same system.

54. The Decomposition Problem

Let Ω be a finite set of primitive carriers at a declared audit resolution. A primitive carrier may be a person, a model instance, a process, a memory store, an instrument, or another independently tracked mechanism. The word *primitive* is relative to the audit. It does not assert a final metaphysical atomization of the system.

Definition 131 (System decomposition)**.** A decomposition of a carrier presentation is a tuple

$$\mathcal{D} = (\Pi, \mathbf{O}, G_\Pi, \mathcal{I}_\Pi),$$

where:

- $\Pi = \{B_\alpha\}_{\alpha \in A}$ is a partition of Ω;
- O_α is the declared local-access map for block B_α, producing the local state or history available at that component;
- G_Π records the declared communication, transfer, and causal interfaces among the blocks; and
- $\mathcal{I}_\Pi$ is the family of feasible interventions at that resolution.

The accessible information of component α is the sigma-field

$$\mathscr{F}_\alpha^{\mathcal{D}} = \sigma(O_\alpha(\mathbf{H}_t)) .$$

Every shared bus, common memory, supervisor, or artifact that can alter more than one local state must appear either as its own carrier or as an explicit channel in G_Π.

The same physical realization can support several decompositions. One may separate a model from its executor, merge them into one computational unit, or treat a persistent artifact as its own component. These alternatives are not mere renamings when they change local access or feasible interventions. Every certificate should therefore display the full specification

$$\mathrm{SRW}(W_t; \varsigma), \qquad \varsigma = (\mathcal{D}, Y^+, \ell, \tau, \mathcal{I}, \mathcal{N}, \Theta).$$

This part varies the decomposition coordinate while holding the remaining coordinates fixed. Part XVIII treats the full specification family.

Definition 132 (Refinement and coarsening)**.** Let $\mathcal{D}_f$ and $\mathcal{D}_c$ be decompositions of the same carrier presentation. The fine decomposition refines the coarse decomposition, written

$$\mathcal{D}_f \preceq \mathcal{D}_c,$$

when there is a surjection

$$q : V_f \longrightarrow V_c$$

such that each coarse component is the union of the fine components in one fiber of q, and its accessible information is generated only from those fine components. The aggregation is *information faithful* when

$$\mathscr{F}_B^{\mathcal{D}_c} = \bigvee_{i \in q^{-1}(B)} \mathscr{F}_i^{\mathcal{D}_f} \qquad (B \in V_c).$$

It is *structurally faithful* when it also retains every internal compatibility condition and every boundary transfer condition needed to recover the original global-section problem.

Refinement and coarsening are directional. A coarse component may contain information that was spread across several finer nodes. Conversely, a fine description may expose internal channels that disappear when a macro-node is formed. The next results identify what is preserved when the aggregation is faithful.

55. Faithful Structural Coarse-Graining

Let the vertex set of a core-transfer diagram be partitioned as

$$V = V_1 \sqcup \cdots \sqcup V_m.$$

Write the transfer discrepancy as

$$D_H = \begin{pmatrix} D_{\mathrm{int}} \\ D_\partial \end{pmatrix}, \qquad D_{\mathrm{int}} = \bigoplus_{\alpha=1}^{m} D_\alpha,$$

where D_α contains the edge constraints internal to V_α, and D_∂ contains the constraints crossing between blocks. Define the internally coherent macro-state spaces

$$W_\alpha = \ker D_\alpha, \qquad W_\Pi = \bigoplus_{\alpha=1}^{m} W_\alpha.$$

Let

$$J_\Pi : W_\Pi \hookrightarrow \mathcal{C}_V$$

be the canonical isometric inclusion, and define the quotient discrepancy

$$D_\Pi = D_\partial J_\Pi.$$

The quotient does not replace a group of nodes by an arbitrary summary. Its macro-state is the complete internal section space of that group, and its boundary operator retains all cross-block compatibility equations.

Theorem 133 (Faithful coarse-graining)**.** *For every finite partition of the vertex set,*

$$\boxed{\ker D_H = J_\Pi \ker D_\Pi.}$$

Equivalently,

$$\boxed{\Gamma(H) \cong \ker D_\Pi = \ker(D_\Pi^* D_\Pi).}$$

In particular,

$$\dim \Gamma(H) = \dim W_\Pi - \operatorname{rank}(D_\Pi).$$

Thus faithful coarse-graining preserves the exact structural workspace even though it changes the number and interpretation of the displayed nodes.

Proof. A vector $c \in \mathcal{C}_V$ lies in $\ker D_H$ exactly when

$$D_{\mathrm{int}} c = 0 \qquad \text{and} \qquad D_\partial c = 0.$$

The first condition is equivalent to $c = J_\Pi w$ for a unique $w \in W_\Pi$. Substitution into the second condition gives

$$D_\Pi w = D_\partial J_\Pi w = 0.$$

Hence $c \in \ker D_H$ if and only if $c = J_\Pi w$ with $w \in \ker D_\Pi$. The Laplacian identity follows from $\ker(D_\Pi^* D_\Pi) = \ker D_\Pi$, and the dimension formula is rank-nullity. □

The theorem supplies a precise sense in which a workspace may be represented at several resolutions without changing its exact core. The preservation fails when the macro-state discards internal section coordinates or when boundary equations are omitted.

Corollary 134 (Two-scale coherence bound)**.** *Let P_Π be the orthogonal projector onto $J_\Pi W_\Pi$. Let $\gamma_{\mathrm{int}} > 0$ be the smallest positive eigenvalue of $D_{\mathrm{int}}^* D_{\mathrm{int}}$, and let $\gamma_\Pi > 0$ be the smallest positive eigenvalue of $D_\Pi^* D_\Pi$. Then every $c \in \mathcal{C}_V$ satisfies*

$$\boxed{\operatorname{dist}(c, \Gamma(H)) \le \frac{\|D_{\mathrm{int}} c\|}{\sqrt{\gamma_{\mathrm{int}}}} + \frac{\|D_\Pi J_\Pi^* P_\Pi c\|}{\sqrt{\gamma_\Pi}}.}$$

The first term measures incoherence within macro-components. The second measures incompatibility among their internally coherent states.

Proof. By the triangle inequality,

$$\operatorname{dist}(c, \Gamma(H)) \le \|c - P_\Pi c\| + \operatorname{dist}(P_\Pi c, \Gamma(H)).$$

The workspace Poincaré estimate applied to D_{int} bounds the first term. Write $P_\Pi c = J_\Pi w$. Theorem 133 and the corresponding estimate for D_Π bound the second term by $\|D_\Pi w\|/\sqrt{\gamma_\Pi}$. Since J_Π is isometric, $w = J_\Pi^* P_\Pi c$. □

This decomposition is useful computationally. Internal blocks can be solved or compressed independently, after which only the boundary discrepancy must be assembled globally. It also gives a diagnostic distinction between a workspace that fails inside its local components and one whose local components are coherent but mutually incompatible.

56. Predictive Risk Under Aggregation

The structural core can be preserved by faithful coarse-graining, while coalition counts generally are not. Let

$$\mathscr{F}_S^{\mathcal{D}} = \bigvee_{i \in S} \mathscr{F}_i^{\mathcal{D}}$$

be the information available to coalition S under decomposition $\mathcal{D}$. For a declared loss ℓ, write

$$\mathcal{R}_{\mathcal{D}}(S) = \inf_f \mathbb{E}\Big[\ell(Y^+, f(\mathscr{F}_S^{\mathcal{D}}))\Big].$$

Theorem 135 (Bayes-risk transport under faithful aggregation)**.** *Suppose $\mathcal{D}_f \preceq \mathcal{D}_c$ through $q : V_f \to V_c$, and suppose the aggregation is information faithful. Then for every coarse coalition $S \subseteq V_c$,*

$$\boxed{\mathcal{R}_{\mathcal{D}_c}(S) = \mathcal{R}_{\mathcal{D}_f}(q^{-1}(S)).}$$

Consequently,

$$k_\varepsilon(\mathcal{D}_c) \le k_\varepsilon(\mathcal{D}_f),$$

where coalition size is counted in displayed nodes.

Proof. Information faithfulness gives

$$\mathscr{F}_S^{\mathcal{D}_c} = \bigvee_{B \in S} \bigvee_{i \in q^{-1}(B)} \mathscr{F}_i^{\mathcal{D}_f} = \mathscr{F}_{q^{-1}(S)}^{\mathcal{D}_f}.$$

The two Bayes risks optimize over the same measurable decision rules and are therefore equal. If a fine coalition of size k is sufficient within ε, the set of coarse blocks meeting that coalition has at most k elements and contains at least the same information. This proves the coalition-size inequality. □

The theorem makes the source of a common error explicit. A minimal coalition size of two at one resolution can become one after those two nodes are merged. Node count is therefore a decomposition-relative statistic.

A more stable quantity measures primitive carrier cost. Let $w_\omega > 0$ be a declared cost for each $\omega \in \Omega$, and assign a block the cost

$$c(B) = \sum_{\omega \in B} w_\omega.$$

Define the sufficient support cost

$$\boxed{s_\varepsilon(\mathcal{D}) = \min_{S \subseteq V_{\mathcal{D}}} \left\{ \sum_{B \in S} c(B) : \mathcal{R}_{\mathcal{D}}(S) \le \mathcal{R}_{\mathcal{D}}(V_{\mathcal{D}}) + \varepsilon \right\}.}$$

If $\mathcal{D}_f \preceq \mathcal{D}_c$ is information faithful, then

$$\boxed{s_\varepsilon(\mathcal{D}_f) \le s_\varepsilon(\mathcal{D}_c).}$$

The fine decomposition permits every union of coarse blocks and may also permit cheaper sub-block coalitions. Equality holds whenever an optimal fine coalition is a union of complete coarse fibers. This weighted support measure does not remove the need to choose primitive carriers, but it prevents

a large macro-component from being counted as a single inexpensive unit merely because it was assigned one node label.

57. Trivialization by Arbitrary Redescription

The dependence on decomposition is unavoidable. Two elementary constructions show why a certificate cannot be required to survive every mathematically possible partition.

Proposition 136 (Coarsening and secret-sharing trivializations)**.** *The following statements hold.*

1. *Any finite realization can be coarsened to a single displayed component. Relative to that decomposition, the full-system state is local and* $k_0 \le 1$.
2. *Let* Z *be uniform on a finite abelian group* G*. Let* K *be an independent uniform group element and define*
$$X_1 = K, \qquad X_2 = Z - K.$$
Then
$$I(Z; X_1) = I(Z; X_2) = 0, \qquad H(Z \mid X_1, X_2) = 0.$$
Under logarithmic loss,
$$\mathcal{R}(\{1\}) = \mathcal{R}(\{2\}) = \log|\mathsf{G}|, \qquad \mathcal{R}(\{1,2\}) = 0,$$
so the declared two-node representation has $k_0 = 2$ *and joint advantage* $\log|\mathsf{G}|$*, even when a single centralized encoder generated both shares.*

Proof. The first statement follows by taking the indiscrete partition $\Pi = \{\Omega\}$. For the second, conditioning on any value of Z leaves K, and hence X_1, uniform. Translation invariance of the uniform law also leaves $X_2 = Z - K$ uniform, so each share is independent of Z. The pair reconstructs $Z = X_1 + X_2$, which gives the entropy and Bayes-risk identities. □

The secret-sharing example is a valid distributed code. What it does not show is that the process producing or using the code is causally distributed across the declared carriers. If one hidden controller creates both shares, stores the key, or jointly selects their updates, that controller belongs in the causal inventory. Coalition irreducibility and matched predictive nulls remain necessary tests, but their interpretation is conditional on a defensible interface model.

This distinction suggests two levels of claim:

Encoding-relative relationality. No declared local view is sufficient.

Carrier-relative relationality. The declared views also correspond to admissible, separately tracked causal interfaces with no omitted common controller.

The stronger claim requires the decomposition audit developed next.

58. Admissible Decomposition Families

No purely formal rule can identify the uniquely correct decomposition of every physical or computational system. The practical goal is instead to define a family of decompositions that are defensible for the question being asked and to report how the certificate changes across that family.

Definition 137 (Admissible decomposition family)**.** An admissible family $\mathfrak{D}$ is a predeclared collection of decompositions of the same carrier presentation satisfying the following audit conditions.

1. **Carrier continuity.** Each component corresponds to a process, artifact, or mechanism that can be tracked through the relevant time scale.
2. **Access locality.** The local state contains only information available through the declared component and its recorded incoming channels.
3. **Channel completeness.** Shared memories, buses, supervisors, caches, and synchronized external cues are represented explicitly.
4. **Intervention separability.** At least some component mechanisms or channels can be altered without directly rewriting the entire system.
5. **Target independence.** The decomposition and its alternatives are fixed without using held-out outcomes to manufacture a favorable result.
6. **Resolution relevance.** A proposed split is retained only when it changes access, intervention, reconstruction, or transfer structure at the declared measurement resolution.
7. **Alternative coverage.** The family contains the principal plausible merges, splits, and artifact assignments that a skeptical analyst would consider.

Let $m_q(W;\mathcal{D})$ be the signed margin of gate q, oriented so that

$$m_q(W;\mathcal{D}) > 0$$

means that the gate passes after its uncertainty correction. Define the decomposition envelope

$$\underline{m}_q = \inf_{\mathcal{D}\in\mathfrak{D}} m_q(W;\mathcal{D}), \qquad \overline{m}_q = \sup_{\mathcal{D}\in\mathfrak{D}} m_q(W;\mathcal{D}),$$

and its sensitivity width

$$\Delta_q^{\mathrm{dec}} = \overline{m}_q - \underline{m}_q.$$

Definition 138 (Decomposition-robust certificate)**.** A workspace passes the $\mathfrak{D}$-robust certificate at level L when

$$\boxed{\mathrm{Certify}^{\mathrm{rob}}_{L,\mathfrak{D}}(W) = \bigwedge_{q\in\mathcal{J}(L)} \left\{ \inf_{\mathcal{D}\in\mathfrak{D}} m_q(W;\mathcal{D}) > 0 \right\}.}$$

A pass for one selected decomposition is a *decomposition-specific* claim. A pass throughout $\mathfrak{D}$ is a *family-robust* claim. A claim is *decomposition fragile* when at least one required gate changes sign inside the admissible family.

Universal invariance over all partitions is neither possible nor desirable, as Proposition 136 shows. The family $\mathfrak{D}$ must be broad enough to challenge the favored description and narrow enough to exclude partitions with no operational meaning.

59. Stability Under Decomposition Transport

Exact faithful aggregation preserves the structural core, but practical alternative decompositions usually introduce approximation. Their certificate can still be controlled by transporting a reference workspace and comparing gate margins.

Fix a reference decomposition $\mathcal{D}_0$. For each $\mathcal{D} \in \mathfrak{D}$, let

$$\Phi_{\mathcal{D}} : W^{\mathcal{D}_0} \longrightarrow W^{\mathcal{D}}$$

be a declared workspace correspondence and let $\delta_{\mathcal{D}}$ measure its structural and causal identity defect. Assume gate q has a transport bound

$$|m_q(\Phi_{\mathcal{D}} W; \mathcal{D}) - m_q(W; \mathcal{D}_0)| \le L_q \delta_{\mathcal{D}} + b_q(\mathcal{D}),$$

where L_q is a sensitivity constant and b_q records model mismatch not captured by the identity distance.

Theorem 139 (Decomposition-transport robustness)**.** *If, for every required gate q,*

$$\boxed{m_q(W; \mathcal{D}_0) > \sup_{\mathcal{D} \in \mathfrak{D}} \left(L_q \delta_{\mathcal{D}} + b_q(\mathcal{D}) \right),}$$

then W passes the $\mathfrak{D}$-robust certificate. Conversely, if for some required gate

$$m_q(W; \mathcal{D}_0) + \sup_{\mathcal{D} \in \mathfrak{D}} \left(L_q \delta_{\mathcal{D}} + b_q(\mathcal{D}) \right) < 0,$$

then that gate fails throughout the family.

Proof. For every $\mathcal{D} \in \mathfrak{D}$, the transport bound gives

$$m_q(\Phi_{\mathcal{D}} W; \mathcal{D}) \ge m_q(W; \mathcal{D}_0) - L_q \delta_{\mathcal{D}} - b_q(\mathcal{D}).$$

The first hypothesis makes the right-hand side positive uniformly in $\mathcal{D}$, which proves robust passage of every required gate. The second statement follows from the corresponding upper bound. □

When $b_q \equiv 0$, a convenient certificate radius is

$$\boxed{r_{\mathrm{dec}}(W) = \min_{q : L_q > 0} \frac{m_q(W; \mathcal{D}_0)}{L_q}.}$$

Every admissible family satisfying

$$\sup_{\mathcal{D} \in \mathfrak{D}} \delta_{\mathcal{D}} < r_{\mathrm{dec}}(W)$$

preserves the sign of all required gate margins. In applications, the transport terms can be bounded using earlier results:

Gate family	Relevant transport control
Structural coherence	Transfer-Laplacian projector drift and spectral gap.
Predictive alignment	Divergence between transported conditional-law states.
Causal efficacy	Distance between matched intervention kernels.
Relationality	Coalition-risk distortion and matched-null preservation error.
Reconstruction	Perturbation of lower frame bounds and handoff error.
Persistence and identity	The coordinatewise identity profile from Part XII.

The theorem is intentionally margin based. A result barely above threshold at one decomposition should not be described as robust when a small, defensible change of interface makes it disappear.

60. Decomposition Audit and Reporting

A complete decomposition audit should proceed in the following order.

1. Inventory the primitive carriers at the chosen temporal and physical resolution.
2. Record every shared resource and distinguish local state from incoming information.
3. Choose a reference decomposition from independently manipulable and temporally trackable interfaces.
4. Construct an admissible family containing the principal plausible merges, splits, and artifact assignments.
5. For every alternative, specify the correspondence to the reference workspace and validate its manipulation checks.
6. Recompute the structural, predictive, causal, coalition, reconstruction, and identity margins.
7. Report the full decomposition envelope, not only the most favorable member of the family.

The report should state:

reference decomposition and audit resolution,
admissible family and reasons for excluded partitions,
hidden-resource and common-controller checks,
faithful-aggregation maps and identity defects,
gatewise minima, maxima, and sensitivity widths,
strongest decomposition-specific and family-robust conclusions.

Four controls are especially important:

1. merge components connected by an uninterruptible common controller and test whether the relational claim collapses;
2. split a known centralized controller into complementary local views and verify that the protocol rejects the resulting secret-sharing appearance of distributedness;
3. expose or remove shared caches, external cues, and common memory that were absent from the favored graph; and
4. repeat the frozen certificate across the complete admissible family.

For the motivating human–AI collaboration, this analysis changes the form of the empirical question. It is not enough to ask whether the human, model, executor, and artifact store form a workspace under one natural description. The stronger question is whether the same operative state and its causal role survive plausible alternatives, such as merging model and executor, separating conversation state from persistent artifacts, or making shared orchestration channels explicit. Those measurements have not yet been performed. The case therefore remains decomposition-specific and observational.

PART XVIII

Specification Robustness, Identifiability, and Statistical Certification

A workspace certificate is never evaluated in a vacuum. It is evaluated for a particular future variable, loss, time horizon, intervention family, null panel, threshold vector, and system decomposition. These choices determine which distinctions matter and which capacities count as successful. They are therefore part of the scientific claim rather than clerical details surrounding it.

This dependence has two consequences. First, a result that passes under one reasonable analysis can fail under another. Robustness must be assessed across a declared family of scientifically defensible alternatives. Second, even a fully specified claim may not be identifiable from the observations and interventions that are actually available. Distinct realizations can induce the same experimental law while disagreeing about whether a relational workspace exists. The theory must say when such disagreement can be resolved, when only bounds are justified, and how sampling uncertainty compounds these structural ambiguities.

61. The Complete Scientific Specification

Definition 140 (Workspace specification)**.** A workspace specification is a tuple

$$\boxed{\varsigma = (\mathcal{D}, Y^+, \ell, \tau, \mathcal{I}, \mathcal{N}, \Theta),}$$

where:

- $\mathcal{D}$ is the decomposition and local-access model;
- Y^+ is the declared future variable or trajectory;
- ℓ is the normalized loss or utility convention;
- τ is the temporal horizon or horizon family;
- $\mathcal{I}$ is the feasible intervention family;
- $\mathcal{N}$ is the required matched-null family; and
- Θ is the gate schema, direction, threshold, and indifference region used for certification.

The complete certificate is written

$$\mathrm{Cert}(W;\varsigma) = \bigwedge_{q \in \mathcal{J}(\varsigma)} \mathbf{1}\{m_q(W;\varsigma) > 0\},$$

where m_q is the signed margin of gate q. Positive margin means that that gate passes in the direction declared by Θ.

The notation absorbs lower-bounded and upper-bounded gates into one sign convention. If a causal effect must exceed c_0, then

$$m_q = \widehat{\eta} - c_0.$$

If a defect must remain below ε_0, then

$$m_q = \varepsilon_0 - \widehat{\delta}.$$

The complete certificate therefore passes exactly when every required signed margin is positive.

The specification does not include arbitrary features of the analysis. It contains choices that alter the meaning or evidential force of the claim. An optimization algorithm, numerical tolerance, or estimator belongs in the estimation protocol unless changing it changes the target estimand or the admissible interventions. Conversely, the target, horizon, and null family cannot be treated as implementation details because changing them can reverse the conclusion without changing the underlying realization.

61.1 Specification relativity is not unrestricted relativism

A theorem-development workspace may be operationally real for a month-long research trajectory and irrelevant for a ten-second motor-control target. A repair channel may preserve theorem identity while failing to preserve social coordination. These are different scientific claims. The theory does not require one workspace state to be sufficient for every possible future or loss. It requires the intended scope to be declared before certification.

This yields the more precise notation

$$\boxed{\mathrm{SRW}(W;\varsigma) = 1.}$$

The earlier expression $\mathrm{SRW}_\Theta(W;\mathcal{D})$ is the special case in which the remaining coordinates of ς are held fixed.

61.2 Inert reparameterization and substantive change

Some changes of specification merely recode the same claim. Others change its content. The distinction is analogous to gauge freedom in the structural theory.

Definition 141 (Specification equivalence)**.** Two specifications ς and ς' are equivalent, written $\varsigma \simeq_{\mathrm{spec}} \varsigma'$, when there are bijective transport maps that:

1. preserve the sigma-field generated by the future target;
2. preserve the ordering and normalization of decisions under the loss;
3. map feasible interventions to physically equivalent interventions;
4. map every required null to a null preserving the same quantities; and
5. carry every gate threshold and margin into its counterpart.

Proposition 142 (Certificate invariance under specification equivalence)**.** *If $\varsigma \simeq_{\mathrm{spec}} \varsigma'$ and Φ is the induced workspace correspondence, then*

$$\mathrm{Cert}(W;\varsigma) = \mathrm{Cert}(\Phi W;\varsigma').$$

Proof. Every required gate is carried to a gate with the same signed margin. The conjunction of positive margins is therefore unchanged. □

A bijective relabeling of target classes with the correspondingly transported loss is inert. Replacing exact theorem validity with stylistic similarity is not. Changing a one-step horizon to a year-long persistence horizon is not. Removing the null that tests hidden centralization is not. Robustness concerns substantive alternatives after equivalent recodings have been quotiented out.

61.3 Admissible specification families

Definition 143 (Admissible specification family)**.** An admissible specification family $\mathfrak{S}$ is a predeclared set of specifications that:

1. refer to the same scientific claim type and carrier presentation;
2. use targets that are equivalent operationalizations or explicitly declared nearby stress targets;
3. normalize losses so that margins are comparable;
4. use horizons justified by the process being studied;
5. include every feasible intervention and null needed to distinguish the principal alternative explanations;
6. use thresholds fixed without access to certification outcomes; and
7. exclude only alternatives for reasons stated before evaluation.

The family should not contain every imaginable analysis. Such a requirement would erase meaningful domain specificity. It should contain the alternatives that a technically informed critic could reasonably regard as equally valid or as necessary stress tests for the declared claim. Multiverse and specification- curve methods formalize related concerns in empirical analysis [50, 48], while sensitivity analysis and partial identification emphasize that conclusions should expose their assumption dependence rather than conceal it [32, 40].

Three classes should be reported separately:

Equivalent alternatives. Different encodings or estimators of the same estimand after appropriate transport.

Robustness alternatives. Scientifically defensible choices intended to support the same claim.

Scope alternatives. Genuinely different targets or domains that define different claims and should not be averaged into one robustness number.

61.4 Common gate schema

A robustness family must use a common semantic gate index $\mathcal{J}_*$. A specification may mark a gate as not required only when the claim level itself does not include that property. It may not remove a failed gate after observing the result. Let

$$\mathcal{J}(\varsigma) \subseteq \mathcal{J}_*$$

be fixed before certification. For a family intended to support one fixed claim level, the cleanest design uses

$$\mathcal{J}(\varsigma) = \mathcal{J}_* \qquad (\varsigma \in \mathfrak{S}).$$

This prevents apparent robustness from being created by silently changing the meaning of passage across the family.

62. Robustness Across Admissible Specifications

For each gate $q \in \mathcal{J}_*$, define the specification envelope

$$\boxed{\underline{m}_q^{\mathfrak{S}}(W) = \inf_{\varsigma \in \mathfrak{S}} m_q(W;\varsigma), \qquad \overline{m}_q^{\mathfrak{S}}(W) = \sup_{\varsigma \in \mathfrak{S}} m_q(W;\varsigma).}$$

Its sensitivity width is

$$\Delta_q^{\mathrm{spec}}(W) = \overline{m}_q^{\mathfrak{S}}(W) - \underline{m}_q^{\mathfrak{S}}(W).$$

The robust certificate margin is

$$\boxed{r_{\mathfrak{S}}(W) = \inf_{\varsigma \in \mathfrak{S}} \min_{q \in \mathcal{J}_*} m_q(W;\varsigma).}$$

Then

$$\mathrm{Cert}_{\mathfrak{S}}(W) = 1 \quad \Longleftrightarrow \quad r_{\mathfrak{S}}(W) > 0.$$

Definition 144 (Specification-specific, robust, and fragile claims)**.** A claim is:

Specification-specific when it is asserted only for one declared ς_0.

Specification-robust when every required gate has positive lower envelope across $\mathfrak{S}$.

Specification-fragile when it passes for at least one admissible specification and fails for at least one other.

A high fraction of passing specifications is informative but does not replace the lower envelope when the claim is universal over $\mathfrak{S}$. A family in which ninety-nine analyses pass and one scientifically decisive hidden- controller null fails does not support the complete relational certificate. The noncompensatory logic remains in force at the specification level.

62.1 Geometry of a specification family

Let $d_{\mathfrak{S}}$ be a declared extended pseudometric on the quotient of $\mathfrak{S}$ by specification equivalence. A convenient decomposition is

$$\begin{aligned} d_{\mathfrak{S}}(\varsigma,\varsigma') = {} & w_D d_D(\mathcal{D},\mathcal{D}') + w_Y d_Y(Y^+,Y^{+'}) + w_\ell d_\ell(\ell,\ell') \\ & + w_\tau d_\tau(\tau,\tau') + w_I d_I(\mathcal{I},\mathcal{I}') + w_N d_N(\mathcal{N},\mathcal{N}') + w_\Theta d_\Theta(\Theta,\Theta'). \end{aligned}$$

No universal metric is imposed. The target term may use a divergence between conditional laws under a declared coupling. The intervention and null terms may use Hausdorff distances between implemented kernels. Threshold terms must be expressed on normalized signed-margin scales. The metric is part of the audit and should be varied in sensitivity analysis when more than one scientifically defensible choice exists.

For a reference specification ς_0, define the distance to certificate failure by

$$\boxed{r_{\mathrm{spec}}(W;\varsigma_0) = \inf\left\{ d_{\mathfrak{S}}(\varsigma,\varsigma_0) : \min_q m_q(W;\varsigma) \le 0 \right\}.}$$

If no failing specification exists in the declared universe, the radius is infinite relative to that universe. A radius near zero means that a small scientifically defensible change can reverse the conclusion.

62.2 Specification transport

Changing the decomposition can alter the coordinate realization of the workspace, while changing the target or horizon can alter the relevant predictive state. Let

$$\Phi_\varsigma : W_{\varsigma_0} \longrightarrow W_\varsigma$$

be the declared correspondence when such transport is needed. For each gate, assume a bound

$$|m_q(\Phi_\varsigma W;\varsigma) - m_q(W;\varsigma_0)| \le L_q d_{\mathfrak{S}}(\varsigma,\varsigma_0) + b_q(\varsigma),$$

where b_q records mismatch not represented by the metric. The mismatch may include a change of model class, imperfect target coupling, or an intervention whose implementation cannot be transported exactly.

62.3 T_{28}: Specification-transport robustness theorem

Theorem 145 (Specification-transport robustness)**.** *Let $\mathfrak{S}$ be an admissible specification family with reference ς_0. Suppose every required gate satisfies the transport bound above. If*

$$\boxed{m_q(W;\varsigma_0) > \sup_{\varsigma\in\mathfrak{S}} [L_q d_{\mathfrak{S}}(\varsigma,\varsigma_0) + b_q(\varsigma)]}$$

for every $q \in \mathcal{J}_*$, then

$$\mathrm{Cert}(\Phi_\varsigma W;\varsigma) = 1 \qquad \textit{for every } \varsigma \in \mathfrak{S}.$$

If $b_q \equiv 0$, then

$$\boxed{r_{\mathrm{spec}}(W;\varsigma_0) \ge \min_{q:L_q>0} \frac{m_q(W;\varsigma_0)}{L_q}.}$$

Proof. For each $\varsigma \in \mathfrak{S}$,

$$m_q(\Phi_\varsigma W;\varsigma) \ge m_q(W;\varsigma_0) - L_q d_{\mathfrak{S}}(\varsigma,\varsigma_0) - b_q(\varsigma).$$

The hypothesis makes the right-hand side positive uniformly over the family. Every required gate therefore passes under every admissible specification. When $b_q = 0$, no gate can cross zero at a distance smaller than its reference margin divided by its Lipschitz constant. Taking the minimum over gates gives the radius bound. □

The theorem is deliberately gatewise. A large structural margin cannot absorb a fragile causal margin. It also clarifies the relation to Part XVII. The decomposition-transport theorem is recovered by varying only the $\mathcal{D}$ coordinate while holding the remaining coordinates fixed.

62.4 A three-budget sufficient condition

In practice, the reference margin is reduced by three distinct uncertainty budgets:

$$\boxed{B_q^{\mathrm{total}} = B_q^{\mathrm{model}} + B_q^{\mathrm{spec}} + B_q^{\mathrm{sample}}.}$$

The first covers variation across observationally compatible realizations. The second covers variation across admissible specifications. The third covers finite-sample estimation. The sufficient condition

$$\widehat{m}_q(W;\varsigma_0) > B_q^{\text{total}} \qquad (q \in \mathcal{J}_*)$$

will reappear below in identified-set and confidence-envelope form. Keeping the budgets separate prevents an apparently narrow confidence interval from being mistaken for strong identification or specification robustness.

63. Experimental Equivalence and Identifiability

A complete mathematical definition does not guarantee that its truth can be recovered from available data. The experimental design may fail to distinguish two realizations that have different internal organization.

Definition 146 (Experimental design and induced law)**.** An experimental design is a tuple

$$\mathcal{E} = (\mathcal{O}, \mathcal{I}, \mathcal{T}),$$

where $\mathcal{O}$ is the observation map, $\mathcal{I}$ is the family of implemented interventions, and $\mathcal{T}$ is the set of measured temporal windows. A realization $\mathfrak{R}$ induces the family of laws

$$P_{\mathfrak{R}}^{\mathcal{E}} = \{\mathcal{L}_{\mathfrak{R}}(\mathcal{O}(\mathbf{H}) \mid \mathrm{do}(\iota)) : \iota \in \mathcal{I},\ t \in \mathcal{T}\}.$$

Definition 147 (Experimental equivalence)**.** Two realizations are experimentally equivalent under $\mathcal{E}$, written

$$\mathfrak{R} \equiv_{\mathcal{E}} \mathfrak{R}',$$

when

$$P_{\mathfrak{R}}^{\mathcal{E}} = P_{\mathfrak{R}'}^{\mathcal{E}}.$$

A property $\psi(\mathfrak{R})$ is identifiable on a model class $\mathfrak{M}$ under $\mathcal{E}$ when it is constant on every $\equiv_{\mathcal{E}}$-equivalence class contained in $\mathfrak{M}$.

Identifiability is therefore relative to both a model class and an experiment family. Enlarging the intervention family refines the equivalence classes and can make previously hidden distinctions identifiable. Enlarging the model class can have the opposite effect.

63.1 T_{29}: Non-identifiability impossibility theorem

Theorem 148 (No uniformly correct decision under experimental equivalence)**.** *Let* $\mathfrak{R}, \mathfrak{R}' \in \mathfrak{M}$ *satisfy*

$$\mathfrak{R} \equiv_{\mathcal{E}} \mathfrak{R}'$$

but

$$\mathrm{Cert}(W_{\mathfrak{R}};\varsigma) \neq \mathrm{Cert}(W_{\mathfrak{R}'};\varsigma).$$

Then the workspace certificate is not identifiable on $\mathfrak{M}$ *under* $\mathcal{E}$*. More strongly, for every possibly randomized decision rule* $\widehat{C}$ *based only on data generated by* $\mathcal{E}$*,*

$$\boxed{\mathbb{P}_{\mathfrak{R}}\left(\widehat{C} \neq C_{\mathfrak{R}}\right) + \mathbb{P}_{\mathfrak{R}'}\left(\widehat{C} \neq C_{\mathfrak{R}'}\right) = 1,}$$

where $C_{\mathfrak{R}} = \mathrm{Cert}(W_{\mathfrak{R}};\varsigma)$*. Consequently,*

$$\inf_{\widehat{C}} \max\left\{\mathbb{P}_{\mathfrak{R}}(\widehat{C} \neq C_{\mathfrak{R}}), \mathbb{P}_{\mathfrak{R}'}(\widehat{C} \neq C_{\mathfrak{R}'})\right\} \geq \frac{1}{2}.$$

Proof. The data law is identical under the two realizations, so $p = \mathbb{P}(\widehat{C} = 1)$ is the same in both. If $C_{\mathfrak{R}} = 0$ and $C_{\mathfrak{R}'} = 1$, the two error probabilities are p and $1 - p$. Their sum is one, and at least one is at least one half. The opposite labeling is identical. □

The theorem is not a limitation of a particular estimator. It is an information-theoretic obstruction. No amount of optimization or sample size can resolve a difference that the declared observations and interventions do not expose.

63.2 Hidden-controller witness

Consider two realizations that emit the same local streams X_1, X_2 and the same outcome

$$Y = X_1 \oplus X_2.$$

In the distributed realization, the two streams are generated and retained by separately intervenable carriers, and no carrier contains the pair. In the centralized realization, a hidden controller contains (X_1, X_2) and emits the two displayed streams. Under passive observation of (X_1, X_2, Y), the induced laws are identical. Even interventions that alter only the displayed messages may fail to distinguish them if the hidden controller simply re-emits the manipulated values.

Carrier-relative relationality becomes identifiable only when the experiment can test the claimed causal separation. Examples include independently interrupting the candidate carriers, exposing the hidden memory, randomizing a shared controller, or measuring whether one physical mechanism reconstructs the complete state before the displayed split. The decomposition controls N_{25}–N_{27} are therefore identifiability interventions, not merely optional robustness checks.

63.3 Identifiability of individual gates

The full certificate may be unidentified even when some coordinates are identified. Structural dimension might be recoverable from observed transfer constraints while causal specificity remains ambiguous. Conversely, a causal contrast may be identifiable even when the latent workspace coordinates are known only up to gauge. Each margin should therefore be assigned one of four statuses:

$$\text{point identified}, \quad \text{sign identified}, \quad \text{partially identified}, \quad \text{unidentified}.$$

A complete pass requires the sign of every required gate to be identified as positive, not merely that one compatible model produces a positive estimate.

64. Partial Identification of Workspace Claims

When point identification fails, the correct response is not to choose one convenient realization. It is to report the set of values compatible with the data and assumptions [40, 31].

Definition 149 (Identified realization set)**.** Given an experimental law $P^{\mathcal{E}}$ and model class $\mathfrak{M}$, define

$$\boxed{\mathfrak{R}_{\mathrm{ID}}(P^{\mathcal{E}}; \mathfrak{M}) = \left\{ \mathfrak{R} \in \mathfrak{M} : P^{\mathcal{E}}_{\mathfrak{R}} = P^{\mathcal{E}} \right\}.}$$

For a gate q and specification ς, its identified margin set is

$$\mathcal{M}_q^{\mathrm{ID}}(P^{\mathcal{E}}; \varsigma) = \{ m_q(W_{\mathfrak{R}}; \varsigma) : \mathfrak{R} \in \mathfrak{R}_{\mathrm{ID}} \}.$$

The sharp lower and upper identified margins are

$$\underline{m}_q^{\mathrm{ID}}(\varsigma) = \inf \mathcal{M}_q^{\mathrm{ID}}(\varsigma), \qquad \overline{m}_q^{\mathrm{ID}}(\varsigma) = \sup \mathcal{M}_q^{\mathrm{ID}}(\varsigma).$$

A gate is sign-identified as passing if its lower bound is positive. It is sign-identified as failing if its upper bound is nonpositive. If the interval crosses zero, the available design does not determine that gate's outcome.

64.1 Identified certificate set

Define

$$\boxed{\mathcal{C}_{\mathrm{ID}}(P^{\mathcal{E}}; \varsigma) = \{\mathrm{Cert}(W_{\mathfrak{R}}; \varsigma) : \mathfrak{R} \in \mathfrak{R}_{\mathrm{ID}}\} \subseteq \{0, 1\}.}$$

The certificate is point identified exactly when this set is a singleton. The marginal gate intervals provide simple sufficient conditions, but they need not be logically necessary because gate margins can be dependent across the identified set.

For robustness across specifications, define the realization-level robust margin

$$r_{\mathfrak{S}}(\mathfrak{R}) = \inf_{\varsigma \in \mathfrak{S}} \min_{q \in \mathcal{J}_*} m_q(W_{\mathfrak{R}}; \varsigma)$$

and its identified range

$$\underline{r}_{\mathrm{ID},\mathfrak{S}} = \inf_{\mathfrak{R} \in \mathfrak{R}_{\mathrm{ID}}} r_{\mathfrak{S}}(\mathfrak{R}), \qquad \overline{r}_{\mathrm{ID},\mathfrak{S}} = \sup_{\mathfrak{R} \in \mathfrak{R}_{\mathrm{ID}}} r_{\mathfrak{S}}(\mathfrak{R}).$$

64.2 T_{30}: Joint specification-identification theorem

Theorem 150 (Identification of a specification-robust certificate)**.** *Let $\mathfrak{R}_{\mathrm{ID}}$ be the identified realization set and $\mathfrak{S}$ an admissible specification family.*

1. *If*

$$\boxed{\underline{r}_{\mathrm{ID},\mathfrak{S}} > 0,}$$

then every realization compatible with the experimental law passes every required gate under every admissible specification. The robust certificate is identified as passing.

2. *If*

$$\boxed{\overline{r}_{\mathrm{ID},\mathfrak{S}} \leq 0,}$$

then no compatible realization passes the complete certificate throughout the family. The robust certificate is identified as failing.

3. *If there exist compatible realizations $\mathfrak{R}_+, \mathfrak{R}_-$ with*

$$r_{\mathfrak{S}}(\mathfrak{R}_+) > 0, \qquad r_{\mathfrak{S}}(\mathfrak{R}_-) \leq 0,$$

then the robust certificate is not identified under the declared experiment.

Proof. The first two conclusions follow directly from the definitions of the infimum and supremum over the identified set. The third gives two experimentally compatible realizations with different robust-certificate values, so Theorem 148 applies. □

This theorem joins two forms of robustness that are often conflated. A result can be stable across specifications within one fitted model yet fail to be identified across models compatible with the data. It can also be identified under one specification while changing sign across a reasonable specification family. The strongest claim controls both axes.

64.3 Confidence regions for identified sets

In finite samples, the identified set is itself uncertain. Let

$$\widehat{\mathfrak{R}}_{1-\delta}$$

be a confidence region satisfying

$$\inf_{\mathfrak{R}_0 \in \mathfrak{M}} \mathbb{P}_{\mathfrak{R}_0}\left(\mathfrak{R}_0 \in \widehat{\mathfrak{R}}_{1-\delta}\right) \geq 1-\delta.$$

A conservative robust lower margin is

$$\widehat{r}_{1-\delta} = \inf_{\mathfrak{R} \in \widehat{\mathfrak{R}}_{1-\delta}} \inf_{\varsigma \in \mathfrak{S}} \min_q m_q(W_{\mathfrak{R}}; \varsigma).$$

If $\widehat{r}_{1-\delta} > 0$, a false specification-robust certificate has probability at most δ, provided that the model class and specification family were fixed before the confidence region was computed. This confidence-set approach is useful when gate estimators are strongly coupled or when the latent realization is the primary inferential object.

65. Uniform Finite-Sample Certification

Population margins are not observed directly. A certification rule must control the estimation error uniformly across every specification and gate on which its conclusion depends.

Let $\widehat{m}_q(\varsigma)$ estimate $m_q(W;\varsigma)$. Suppose there are nonnegative error radii $\varepsilon_q(n,\delta)$ such that

$$\mathbb{P}\left(\sup_{\varsigma \in \mathfrak{S}} |\widehat{m}_q(\varsigma) - m_q(\varsigma)| \leq \varepsilon_q(n,\delta) \quad \text{for all } q \in \mathcal{J}_*\right) \geq 1-\delta.$$

Define the empirical lower confidence envelope

$$\boxed{\underline{\widehat{m}}^{\mathfrak{S}}_{q,1-\delta} = \inf_{\varsigma \in \mathfrak{S}} \widehat{m}_q(\varsigma) - \varepsilon_q(n,\delta).}$$

65.1 T_{31}: Uniform specification-certificate theorem

Theorem 151 (Uniform finite-sample certification across specifications)**.** *Under the uniform error event above, declare a specification-robust certificate only when*

$$\underline{\widehat{m}}^{\mathfrak{S}}_{q,1-\delta} > 0 \qquad \textit{for every } q \in \mathcal{J}_*.$$

Then

$$\boxed{\sup_{P:\, \mathrm{Cert}_{\mathfrak{S}}(W)=0} \mathbb{P}_P\left(\textit{the robust certificate is declared}\right) \leq \delta.}$$

Proof. On the uniform error event,

$$m_q(\varsigma) \geq \widehat{m}_q(\varsigma) - \varepsilon_q$$

for every gate and every specification. If all empirical lower envelopes are positive, every population margin is positive throughout the family. A false declaration can therefore occur only when the uniform error event fails, which has probability at most δ. □

The theorem controls the complete specification family, not merely the analysis that happened to produce the strongest estimate. It is an extension of the gatewise intersection rule in Theorem 156. Gatewise validity controls the conjunction at one frozen specification. Uniform validity controls the conjunction over a family.

65.2 Finite specification grids

Suppose $\mathfrak{S}$ and $\mathcal{J}_*$ are finite, and each margin is estimated as

$$\widehat{m}_q(\varsigma) = \frac{1}{n}\sum_{i=1}^{n} Z_{i,q,\varsigma} - \theta_{q,\varsigma}, \qquad 0 \leq Z_{i,q,\varsigma} \leq 1,$$

from independent certification units. Hoeffding's inequality and a union bound give

$$\mathbb{P}\left(\sup_{q,\varsigma} |\widehat{m}_q(\varsigma) - m_q(\varsigma)| > \sqrt{\frac{\log(2|\mathcal{J}_*||\mathfrak{S}|/\delta)}{2n}}\right) \leq \delta.$$

Thus the choice

$$\boxed{\varepsilon_n(\delta) = \sqrt{\frac{\log(2|\mathcal{J}_*||\mathfrak{S}|/\delta)}{2n}}}$$

yields an explicit conservative certificate. Dependent episodes require a bound appropriate to the true independent unit rather than token-level or edge-level pseudo-replication.

65.3 Continuous specification families

Let $\mathfrak{S}_\eta$ be an η-net of $(\mathfrak{S}, d_{\mathfrak{S}})$ with covering number $N(\eta, \mathfrak{S}, d_{\mathfrak{S}})$. If both the population and empirical margin functions are L_q-Lipschitz, then

$$\sup_{\varsigma\in\mathfrak{S}} |\widehat{m}_q(\varsigma) - m_q(\varsigma)| \leq \max_{\varsigma_j\in\mathfrak{S}_\eta} |\widehat{m}_q(\varsigma_j) - m_q(\varsigma_j)| + 2L_q\eta.$$

For bounded empirical means, a valid radius is therefore

$$\boxed{\varepsilon_{q,n,\eta}(\delta) = \sqrt{\frac{\log(2|\mathcal{J}_*|N(\eta, \mathfrak{S}, d_{\mathfrak{S}})/\delta)}{2n}} + 2L_q\eta.}$$

The tuning parameter η balances net cardinality against interpolation error. More refined empirical-process bounds may replace this elementary covering argument without changing the certificate logic.

65.4 Power and indifference regions

A robust procedure can be valid yet uninformative when the family is too large or the data are insufficient. Each gate should therefore report an indifference region $[-\kappa_q, \kappa_q]$. A conclusion is:

Robust pass when the lower confidence envelope exceeds κ_q for every gate.

Robust fail when a required gate's upper confidence envelope is below $-\kappa_q$ throughout the relevant family.

Specification conflict when well-estimated admissible specifications lie on opposite sides of the indifference region.

Statistically unresolved when sampling uncertainty intersects the indifference region.

This vocabulary separates genuine sensitivity from low power.

66. Adaptive Specification Choice and Selective Validity

A pointwise valid test does not remain valid after the analyst searches for the most favorable target, horizon, null panel, or threshold. Consider K independent null p-values and select the smallest. Then

$$\mathbb{P}\left(\min_{1\le k\le K} p_k \le \alpha\right) = 1-(1-\alpha)^K,$$

which exceeds α whenever $K>1$. The same logic applies when a workspace certificate is reported only for the specification that produced the largest gate margins.

The issue is not exploration itself. Exploration is often necessary to learn which target and intervention are scientifically meaningful. The issue is using the same random outcomes both to choose the claim and to certify it. Post-selection inference addresses the general problem by using simultaneous or selection-adjusted procedures [10].

66.1 Specification leakage

Definition 152 (Specification leakage)**.** Specification leakage occurs when certification outcomes influence any of the following without a fresh inferential correction:

$$Y^+, \ell, \tau, \mathcal{I}, \mathcal{N}, \Theta, \mathfrak{S}, \mathcal{J}_*.$$

The leakage status is separate from formation leakage. A workspace extractor may be frozen while the analyst still searches over targets or thresholds on the certification split.

A valid report must retain the complete search history, including failed and abandoned specifications. Reporting only the selected analysis prevents the reader from determining the effective multiplicity.

66.2 T_{32}: Sample-split selection theorem

Theorem 153 (Validity after independent specification selection)**.** *Let $\widehat{\varsigma}$ be selected using $\mathcal{D}_{\mathrm{sel}}$, and let $\mathcal{D}_{\mathrm{cert}}$ be independent of $\mathcal{D}_{\mathrm{sel}}$. For every fixed $\varsigma\in\mathfrak{S}$, let $\phi_\varsigma(\mathcal{D}_{\mathrm{cert}})\in\{0,1\}$ be a certificate rule satisfying*

$$\sup_{P:\mathrm{Cert}(W;\varsigma)=0} \mathbb{P}_P(\phi_\varsigma=1)\le\alpha.$$

Then the selected-specification procedure satisfies

$$\boxed{\sup_P \mathbb{P}_P\left(\phi_{\widehat{\varsigma}}=1, \mathrm{Cert}(W;\widehat{\varsigma})=0\right)\le\alpha.}$$

Proof. Condition on $\mathcal{D}_{\mathrm{sel}}$. The selected specification is then fixed. Independence leaves the certification data distributed exactly as in the fixed-specification test, whose false-pass probability is at most α. Averaging the conditional bound over the selection data preserves the same bound. □

The theorem justifies discovery on formation or calibration data followed by a single frozen evaluation on fresh certification data. Reusing the certification split to revise the specification requires either another fresh split or a simultaneous post-selection procedure whose validity covers the entire search.

66.3 Nested selection and stress testing

A practical design may use:

$$\mathcal{D}_{\mathrm{form}} \sqcup \mathcal{D}_{\mathrm{spec}} \sqcup \mathcal{D}_{\mathrm{cert}} \sqcup \mathcal{D}_{\mathrm{stress}}.$$

Formation learns the workspace. The specification split chooses targets, horizons, nulls, and thresholds from a predeclared menu. Certification makes the pass decision once. Stress testing explores broader shifts without retroactively changing the certified claim. Cross-fitting can recover sample efficiency, but the fold-level selection and evaluation paths must remain separated and the final uncertainty calculation must account for recombination.

66.4 Thresholds and null families are part of selection

Threshold tuning can be as consequential as target selection. If the observed causal effect is 0.12, choosing a minimum effect threshold of 0.10 after seeing the estimate is not certification. Likewise, retaining only the nulls that the intact system defeats converts a falsification panel into a success filter. Thresholds and null families must be frozen or evaluated under a uniform rule that includes every considered alternative.

67. Specification Audit and Reporting

A complete specification audit should accompany every strong relational workspace claim.

67.1 Audit procedure

1. State the reference specification ς_0 in full.
2. Distinguish equivalent recodings, robustness alternatives, and genuinely different scope claims.
3. Declare the admissible family $\mathfrak{S}$ and justify every major exclusion before certification.
4. Record the specification pseudometric or other transport bounds used to compare family members.
5. Compute every gate's lower envelope, upper envelope, sensitivity width, and reference margin.
6. State the experiment family $\mathcal{E}$ and the model class $\mathfrak{M}$ relative to which identifiability is claimed.
7. Report identified sets or bounds whenever point identification fails.
8. Apply a uniform finite-sample rule over every gate and specification used in the conclusion.
9. Preserve the complete specification-selection history and the data split on which each choice was made.
10. Report the strongest pointwise, family-robust, sign-identified, and statistically resolved conclusions separately.

67.2 Additional adversarial controls

The empirical null and audit panel is extended by:

$$
\begin{aligned}
N_{29} &= \text{post hoc target-selection audit},\\
N_{30} &= \text{horizon cherry-picking audit},\\
N_{31} &= \text{loss and utility substitution audit},\\
N_{32} &= \text{restricted intervention-family audit},\\
N_{33} &= \text{threshold and gate-set tuning audit},\\
N_{34} &= \text{frozen specification sweep across } \mathfrak{S}.
\end{aligned}
$$

The first five reconstruct or bound the analysis search that preceded the reported claim. The final control reruns the complete frozen certificate across the admissible family. A claim that survives N_{34} but fails the search audit remains vulnerable to selection bias. A claim that passes the search audit but changes sign across $\mathfrak{S}$ is validly estimated yet specification-fragile.

67.3 Required specification record

The machine-readable record should include:

$$
\begin{aligned}
\mathsf{SpecificationRecord} = \{&\mathsf{reference}, \mathsf{family}, \mathsf{equivalence_classes}, \mathsf{metric},\\
&\mathsf{targets}, \mathsf{losses}, \mathsf{horizons}, \mathsf{interventions},\\
&\mathsf{null_panels}, \mathsf{thresholds}, \mathsf{selection_history},\\
&\mathsf{identified_sets}, \mathsf{uniform_bounds}, \mathsf{gate_envelopes}\}.
\end{aligned}
$$

Every omitted family member requires a reason. Every changed specification requires a data-split label and time stamp. Every gate requires pointwise and family-wide status.

67.4 Claim-strength vocabulary

The following vocabulary is recommended:

Pointwise demonstrated. The certificate passes at one frozen specification with valid uncertainty.

Specification-robust. It passes uniformly across a declared admissible family.

Sign identified. Every realization compatible with the experimental law gives the same gate sign.

Robustly identified. The complete family-robust certificate is identified as passing.

Partially identified. The data restrict the certificate margins but do not determine every required sign.

Specification-fragile. Validly estimated admissible specifications give different certificate outcomes.

Selection-contaminated. The reported specification was chosen using certification outcomes without valid adjustment.

67.5 Consequence for the motivating collaboration

The motivating human–AI research process remains a plausible candidate persistent workspace, but its strongest current description is narrower than a certified robust claim. The relevant target trajectory, horizon family, intervention set, null panel, thresholds, and admissible specification family have not been frozen and tested prospectively. Nor has the carrier-relative workspace predicate been shown to be identifiable against hidden-controller and matched-storage alternatives. The theory therefore supports a disciplined research program rather than a retrospective declaration.

The terminal architecture is now explicit. A strong claim must survive three separate questions:

Would it survive another defensible specification?
Could a contrary realization generate the same experiments?
Does the finite sample support every required margin uniformly?

Once these questions are answered, further extensions concern new mathematical settings or new empirical domains rather than a missing foundational layer.

PART XIX

Summary of Main Results

68. Structural and Foundational Results

T_1: Limit–holonomy theorem

$$\Gamma(H) \cong C_o^{\mathrm{Hol}_o}.$$

T_2: Operator closure theorem

$$H_e \rho_i(a) = \rho_j(a) H_e \Longrightarrow \rho_o(a) C_* \subseteq C_*.$$

T_{2a}: Workspace commutant theorem

$$\mathcal{O}_{\mathbf{W}} \cong \mathrm{Comm}(\mathrm{Hol}_o).$$

T_3: Transfer-Laplacian kernel theorem

$$\ker L_H = \Gamma(H).$$

T_4: Spectral stability theorem

If $\eta = \|\widetilde{L}_H - L_H\| < g_d/2$, then

$$\|\widetilde{P}_d - P_d\| \leq \frac{\eta}{g_d - \eta} \leq \frac{2\eta}{g_d}.$$

T_5: Gauge invariance theorem

$$\Gamma(H^g) = G_V \Gamma(H),$$

and under unitary gauge,

$$L_{H^g} = G_V L_H G_V^{-1}.$$

T_6: Persistence theorem

For the nonautonomous recurrence

$$e_{t+1} \leq \alpha_t e_t + \sigma_t,$$

$$\boxed{e_t \leq \left(\prod_{r=0}^{t-1} \alpha_r \right) e_0 + \sum_{j=0}^{t-1} \left(\prod_{r=j+1}^{t-1} \alpha_r \right) \sigma_j.}$$

If $\alpha_t \le \bar{\alpha} < 1$ and $\sigma_t \le \bar{\sigma}$, then

$$\limsup_{t\to\infty} e_t \le \frac{\bar{\sigma}}{1-\bar{\alpha}}.$$

The derivation begins with a moving workspace projector and an explicit system update, and adds a mean-square analogue and a non-collapse gate.

T_7: Stable reconstruction theorem

For coalition analysis operator A_S with lower frame bound $\alpha_S > 0$, observations

$$y = A_S c + \eta$$

and canonical decoder $A_S^\dagger$ satisfy

$$\boxed{\|A_S^\dagger y - c\| \le \frac{\|\eta\|}{\sqrt{\alpha_S}}.}$$

The theorem is extended to encoding mismatch, exact support-based erasure distance, uniform node erasures, reconstructive handoff, exact coalition sufficiency, and operational risk under a Lipschitz predictive kernel.

T_8: Exact gluing theorem

For every finite edge-complete workspace cover,

$$\boxed{0 \longrightarrow \mathscr{W}(G) \xrightarrow{R_{\mathcal{U}}} \check{C}^0(\mathcal{U};\mathscr{W}) \xrightarrow{\delta^0} \check{C}^1(\mathcal{U};\mathscr{W})}$$

is exact. Hence

$$\boxed{\mathscr{W}(G) \cong \ker \delta^0 \cong \mathrm{Eq}\left(\prod_\alpha \mathscr{W}_\alpha \rightrightarrows \prod_{\alpha<\beta} \mathscr{W}_{\alpha\beta}\right).}$$

The same analysis also yields a cover-Laplacian distance bound, a total descent kernel theorem, and a harmonic $\check{H}^1$ obstruction for relative-only data.

T_9: Hierarchical collapse theorem

For every finite rooted hierarchy whose children edge-cover their parent,

$$\boxed{W_a \cong \mathscr{W}(G_a) \quad \text{for every hierarchy node } a,}$$

and in particular

$$\boxed{W_{\mathrm{root}} \cong \mathscr{W}(G).}$$

If level r contributes defect δ_r and downstream collapse maps have Lipschitz constants L_s, then

$$\boxed{\delta_{\mathrm{flat}}^{(R)} \le \sum_{r=0}^{R-1} \left(\prod_{s=r+1}^{R-1} L_s\right) \delta_r.}$$

T_{25}: Faithful coarse-graining theorem

For a partition of the vertex set, let W_Π be the direct sum of the internally coherent section spaces and let D_Π retain all boundary compatibility equations. Then

$$\boxed{\Gamma(H) = J_\Pi \ker D_\Pi.}$$

Exact structural workspace content is therefore invariant under faithful aggregation. Approximate error separates into an internal-coherence term and a boundary-coherence term controlled by their respective spectral gaps.

T_{26}: Bayes-risk transport theorem

For an information-faithful coarsening $q : V_f \to V_c$,

$$\boxed{\mathcal{R}_{\mathcal{D}_c}(S) = \mathcal{R}_{\mathcal{D}_f}(q^{-1}(S)) \qquad (S \subseteq V_c).}$$

Hence displayed coalition size can decrease under coarsening. Weighted primitive support cost is a more stable localization measure and cannot improve merely because several carriers were relabeled as one macro-node.

T_{27}: Decomposition-transport robustness theorem

If every gate margin at a reference decomposition exceeds its worst-case transport and model-mismatch defect across $\mathfrak{D}$, then the complete certificate passes throughout the admissible family:

$$\boxed{m_q(W; \mathcal{D}_0) > \sup_{\mathcal{D} \in \mathfrak{D}} \left(L_q \delta_{\mathcal{D}} + b_q(\mathcal{D})\right) \quad \forall q \Longrightarrow \mathrm{Certify}^{\mathrm{rob}}_{L,\mathfrak{D}}(W) = 1.}$$

69. Dynamic, Causal, and Agency Results

T_{10}: Approximate universality theorem

For the first d transfer-Laplacian modes, with $\lambda_d \le \varepsilon < \lambda_{d+1} = \mu$,

$$\boxed{\|(I - P_d)\rho_V(a)P_d\| \le \frac{\|\rho_E(a)\|\sqrt{\varepsilon} + \beta(a)}{\sqrt{\mu}}.}$$

Thus low transfer energy, a separated complement, and small intertwining defect imply approximate closure under the common operation algebra.

T_{11}: Core–residual accumulation theorem

For a path $\gamma = e_m \cdots e_1$,

$$\boxed{\begin{aligned}\|Q_{t(\gamma)} T_\gamma P_{s(\gamma)}\| \le \sum_{r=1}^{m} \left(\prod_{\ell=r+1}^{m} \|T_{e_\ell}\|\right) \|V_{e_r}\| \\ \times \left(\prod_{\ell=1}^{r-1} \|H_{e_\ell}\|\right).\end{aligned}}$$

The same bound controls the defect between full core-to-core path transport and the pure core product. An exact Duhamel identity yields both estimates.

T_{12}: Predictive-state minimality and alignment theorem

On standard Borel spaces, the canonical conditional-law state

$$Z_t = \mathbb{P}(Y_t^+ \in \cdot \mid \mathbf{H}_t)$$

is sufficient and minimal:

$$Y_t^+ \perp\!\!\!\perp \mathbf{H}_t \mid Z_t,$$

and every sufficient statistic $S_t = s(\mathbf{H}_t)$ satisfies

$$Z_t = \phi(S_t) \qquad \text{a.s.}$$

Moreover, a structural candidate W_t is exactly operationally aligned if and only if

$$Z_t = \phi_W(W_t) \qquad \text{a.s.}$$

The countable intervention-indexed version holds for $Z_t^{\mathcal{I}}$.

T_{13}: Operational relationality theorem

For a finite target, exact alignment, a positive node-level conditional information gap, a positive robust local-predictive-matched null gap, a feasible workspace intervention effect, and positive matched specificity jointly imply that:

no single-node statistic is sufficient,
$k_0 \geq 2$,
no single node reconstructs the aligned workspace,
matched local nulls preserve all node risks but lose joint capacity,
and the workspace coordinate is causally efficacious at the tested scale.

If $\eta_{\text{loop}}(\tau) > 0$, the conclusion additionally includes loop-mediated reentry.

T_{14}: Replacement invariance theorem

For a unitary realization correspondence with discrepancy defect

$$\delta_D = \|D' J_V - J_E D\|,$$

set

$$\eta_L = 2\|D\|\delta_D + \delta_D^2.$$

If $\eta_L < g/2$, then the workspace projectors obey

$$\boxed{q_{\text{str}} \leq \frac{\eta_L}{g - \eta_L} \leq \frac{2\eta_L}{g}.}$$

A canonical unitary correspondence between the workspace subspaces exists, and its operation defect is bounded by the projector drift, the local intertwining defect, and prior operation leakage. Coalition frame bounds also persist under sufficiently small reconstructive transport error. With declared causal-kernel and reconstructive-handoff defects,

$$\boxed{d_{\text{id}} \leq a\, q_{\text{str}} + b\, \delta_{\text{op}} + c\, \delta_{\text{causal}} + d\, \delta_{\text{rec}}.}$$

Under bounded norms and a uniform spectral gap, this gives the schematic estimate

$$d_{\mathrm{id}} \lesssim C \frac{\delta_{\mathrm{replacement}}}{g}.$$

The causal and reconstructive terms are necessary assumptions rather than consequences of structural similarity.

T_{15}: Transfer identification theorem

Under the linear transfer model

$$Y = H_* X + \xi, \qquad \mathbb{E}[\xi X^{\mathsf{T}}] = 0, \qquad \Sigma_{XX} \succeq \kappa I,$$

the unique population least-squares transfer is

$$\boxed{H_* = \Sigma_{YX} \Sigma_{XX}^{-1},}$$

and

$$\mathcal{R}(H) - \mathcal{R}(H_*) \geq \kappa \|H - H_*\|_F^2.$$

For empirical ordinary least squares,

$$\|\widehat{H} - H_*\|_F \leq \frac{\|\widehat{M}\|_F}{\widehat{\kappa}},$$

with the stated ridge extension. The identified transfer is gauge covariant under independent invertible changes of source and target coordinates.

T_{16}: Uniform structure-selection theorem

For a finite candidate family with unique population optimum θ_*, margin $\Delta_J > 0$, and held-out uniform score error at most ε,

$$\boxed{J(\widehat{\theta}) - J(\theta_*) \leq 2\varepsilon.}$$

If $2\varepsilon < \Delta_J$, then $\widehat{\theta} = \theta_*$. The theorem applies to node membership, edge selection, gauge-normalized transfer models, and finite boundary families.

T_{17}: Stable boundary recovery theorem

For a finite admissible boundary family with unique maximizer S_t^*, margin Δ_t^B, and uniform held-out score error ε_t,

$$\mathcal{B}_t(S_t^*) - \mathcal{B}_t(\widehat{S}_t) \leq 2\varepsilon_t,$$

and exact recovery holds when $2\varepsilon_t < \Delta_t^B$. If the population score drifts uniformly by at most δ_t, then

$$2\delta_t < \Delta_t^B \Longrightarrow S_{t+1}^* = S_t^*.$$

T_{18}: Model-specific formation transition theorem

For

$$m_{t+1} = \tanh(\lambda m_t),$$

there is a unique zero fixed point for $0 \le \lambda \le 1$. For $\lambda > 1$, the zero state is unstable and exactly two stable nonzero fixed points appear. Their amplitude satisfies

$$\boxed{m_\lambda = \sqrt{\frac{3(\lambda-1)}{\lambda^3}} + O((\lambda-1)^{3/2}).}$$

This is an exact structural benchmark, not a universal law for complete relational workspaces.

T_{19}: Cue-controlled goal-retention theorem

For matched future command and disturbance trajectories under

$$G_{s+1} = R_s G_s + B_s C_{s+1} + \zeta_{s+1},$$

two goal interventions satisfy

$$\boxed{G_{t+\tau} - G'_{t+\tau} = R_{t+\tau:t}(g - g').}$$

If

$$m_s \le \sigma_{\min}(R_s), \qquad \|R_s\| \le M_s,$$

then

$$\boxed{\left(\prod_s m_s\right)\|g-g'\| \le \|G_{t+\tau} - G'_{t+\tau}\| \le \left(\prod_s M_s\right)\|g-g'\|.}$$

A linear action readout gives a corresponding lower bound on the induced action-mean separation. The result isolates internal retention from repeated external goal rewriting.

T_{20}: Approximate counterfactual-policy theorem

If the interventional action-value estimator satisfies

$$\sup_a |\widehat{Q}_t(a) - Q_t(a)| \le \varepsilon_Q$$

and the implemented policy has optimization error at most δ_{opt} relative to $\widehat{Q}_t$, then

$$\boxed{\text{Reg}_t(\widehat{\pi}_t) \le 2\varepsilon_Q + \delta_{\text{opt}}.}$$

For deterministic greedy action, the bound is $2\varepsilon_Q$. A unique true optimum with value gap greater than $2\varepsilon_Q$ is recovered exactly. The guarantee is relative to the declared utility, feasible action support, causal identification, and horizon.

T_{21}: Repair-selection margin theorem

For a finite repair set and a repair-value estimator with uniform error ε_R, the selected repair $\widehat{r}$ satisfies

$$\boxed{V(r^*) - V(\widehat{r}) \le 2\varepsilon_R.}$$

If the optimal repair reaches reserve $\rho + 2\varepsilon_R$, then the selected repair reaches reserve ρ. A unique optimal repair whose value gap exceeds $2\varepsilon_R$ is selected exactly.

T_{22}: Closed-loop self-maintenance theorem

If matched no-repair deficit obeys

$$e_{t+1}^{(0)} \le \alpha e_t + d_t$$

and the implemented repair has measured causal effect

$$e_{t+1}^{(0)} - e_{t+1} \ge \kappa e_t - \epsilon_t,$$

set $\beta = \alpha - \kappa$. Under

$$0 \le \beta < 1, \qquad d_t \le \overline{d}, \qquad \epsilon_t \le \overline{\epsilon},$$

$$\boxed{e_t \le \beta^t e_0 + \frac{(\overline{d} + \overline{\epsilon})(1 - \beta^t)}{1 - \beta}.}$$

Hence

$$\limsup_{t \to \infty} e_t \le \frac{\overline{d} + \overline{\epsilon}}{1 - \beta}.$$

A corresponding margin condition makes the declared bounded-deficit viability set forward invariant.

T_{23}: Workspace-agency synthesis theorem

Operational relationality, cue-controlled and behaviorally operative goals, calibrated counterfactual values, feasible value-to-action and action-to-outcome interventions, causal closed-loop maintenance, distributed control-state sufficiency, and matched local policy nulls jointly imply:

$$\begin{aligned} &\mathrm{Reg}_t \le 2\varepsilon_Q + \delta_{\mathrm{opt}}, \\ &\limsup e_t \le \frac{\overline{d} + \overline{\epsilon}}{1 - \beta}, \\ &k_{\mathrm{ctl},\varepsilon} \ge 2, \end{aligned}$$

together with causal mediation of goal-dependent action and repair through the implemented internal channels. This is a sufficient, domain-relative agency certificate. It does not imply consciousness, personhood, or unrestricted autonomy.

T_{24}: Gatewise certification-validity theorem

Suppose a certificate requires every gate $G_1, \dots, G_m$ to pass, and the predeclared test for gate j has false-pass probability at most α_j whenever that required property is absent. Under the global null that at least one required property is absent,

$$\boxed{\mathbb{P}(\text{complete certificate passes}) \le \max_j \alpha_j.}$$

No independence assumption among gate tests is required. The guarantee fails under adaptive gate deletion, selective reporting, or reuse of the certification split.

70. Decomposition, Specification, and Identification Results

T_{25}: Faithful coarse-graining theorem

For an exact partition into internally coherent macro-components,

$$\boxed{\ker D_H = J_\Pi \ker D_\Pi.}$$

Thus faithful aggregation preserves the exact shared workspace and its dimension while changing the displayed node resolution.

T_{26}: Bayes-risk transport under aggregation

For an information-faithful coarsening map q,

$$\boxed{\mathcal{R}_{\mathcal{D}_c}(S) = \mathcal{R}_{\mathcal{D}_f}(q^{-1}(S)).}$$

Predictive content is preserved, although the visible coalition count may contract. Primitive-carrier support cost is therefore preferred when claims must be compared across resolutions.

T_{27}: Decomposition-transport robustness theorem

If every required gate has reference margin larger than its worst admissible decomposition-transport defect, then the complete certificate passes throughout $\mathfrak{D}$. When the mismatch term vanishes, the certificate radius is bounded below by the smallest gate margin divided by its transport sensitivity.

T_{28}: Specification-transport robustness theorem

For a reference specification ς_0, if

$$m_q(W;\varsigma_0) > \sup_{\varsigma\in\mathfrak{S}} [L_q d_{\mathfrak{S}}(\varsigma,\varsigma_0) + b_q(\varsigma)]$$

for every gate, then the transported workspace passes every admissible specification. If $b_q = 0$,

$$r_{\mathrm{spec}}(W;\varsigma_0) \geq \min_{q:L_q>0} \frac{m_q(W;\varsigma_0)}{L_q}.$$

T_{29}: Experimental non-identifiability theorem

If two realizations induce the same law under every declared observation and intervention but have different certificate values, no decision rule based on that experiment can be uniformly correct. For every such rule, the two error probabilities sum to one, so the minimax error is at least one half.

T_{30}: Joint specification-identification theorem

Let $r_{\mathfrak{S}}(\mathfrak{R})$ be the worst gate margin across the specification family. If its infimum over the identified realization set is positive, the robust certificate is identified as passing. If its supremum is nonpositive, the certificate is identified as failing. Compatible realizations on opposite sides of zero establish non-identifiability.

T_{31}: Uniform finite-sample specification theorem

If all gate estimators are uniformly within their error radii across $\mathfrak{S}$, then requiring every empirical lower envelope to be positive controls the probability of a false robust certificate by the failure probability of the uniform event.

T_{32}: Independent selection and certification theorem

A specification chosen on data independent of the certification split may be evaluated by its fixed-specification test without inflating the false-pass bound. Selection and certification on the same outcomes require simultaneous or post-selection correction.

PART XX

Conjectures and Open Mathematical Directions

71. Workspace Formation Transitions

The symbolic outline proposed a coupling threshold λ_c. A general theory must first state what is allowed to vary with λ. A threshold may concern structural coherence, causal efficacy, relational irreducibility, or the conjunction of all certification gates. These need not occur at the same parameter value.

71.1 No coordinate-free universal threshold

Proposition 154 (Threshold values are parameterization dependent)**.** *Suppose a model indexed by λ passes a declared formation certificate exactly for $\lambda > \lambda_c$. For every strictly increasing bijection f, the reparameterized family indexed by $\mu = f(\lambda)$ has threshold*

$$\mu_c = f(\lambda_c).$$

Consequently, the numerical value of a coupling threshold is not invariant without a physically or operationally fixed parameterization.

Proof. The reparameterized realization at μ is the original realization at $f^{-1}(\mu)$. It passes precisely when $f^{-1}(\mu) > \lambda_c$, equivalently when $\mu > f(\lambda_c)$. □

There can therefore be no model-independent numerical value of λ_c. The invariant object is the change in the certified organization and its local critical behavior, not the label placed on the control parameter.

71.2 Gate-specific transition parameters

For a one-parameter family $\mathfrak{R}_\lambda$, define

$$\begin{aligned}
\lambda_{\mathrm{str}} &= \inf\{\lambda : d_\varepsilon(\lambda) > 0, g_\varepsilon(\lambda) > 0\},\\
\lambda_{\mathrm{op}} &= \inf\{\lambda : \delta_{\mathrm{align}}(\lambda) \le \varepsilon_{\mathrm{op}}\},\\
\lambda_{\mathrm{causal}} &= \inf\{\lambda : \eta_{\mathrm{causal}}(\lambda) \ge \tau_{\mathrm{causal}}\},\\
\lambda_{\mathrm{rel}} &= \inf\{\lambda : k_\varepsilon(\lambda) \ge 2, \Delta_{\mathrm{tr}}(\lambda) \ge \tau_{\mathrm{tr}}\},\\
\lambda_{\mathrm{form}} &= \inf\{\lambda : \mathrm{Form}_{\Theta_{\mathrm{form}}}(W_\lambda) = 1\}.
\end{aligned}$$

In general,

$$\lambda_{\mathrm{str}}, \lambda_{\mathrm{op}}, \lambda_{\mathrm{causal}}, \lambda_{\mathrm{rel}}, \lambda_{\mathrm{form}}$$

can be distinct. A synchronized representation can appear before it becomes causally useful, and causal coordination can appear while remaining reducible to one centralized component.

71.3 A solvable nonlinear formation model T_{18}

A minimal exact model is the scalar recurrent coherence equation

$$\boxed{m_{t+1} = \tanh(\lambda m_t), \qquad \lambda \ge 0.}$$

The order parameter m_t represents a signed coherent mode. This model is not a complete relational workspace. It isolates one mechanism by which a nonzero shared mode can appear continuously from a symmetric zero state.

Theorem 155 (Pitchfork formation transition)**.** *For the map*

$$F_\lambda(m) = \tanh(\lambda m), \qquad m \in [-1, 1],$$

the following hold.

1. *If $0 \le \lambda \le 1$, the unique fixed point is $m = 0$. It is globally attracting for $\lambda < 1$ and nonexpanding at $\lambda = 1$.*
2. *If $\lambda > 1$, there are exactly three fixed points*

$$-m_\lambda, \quad 0, \quad m_\lambda, \qquad m_\lambda \in (0, 1).$$

   The zero fixed point is unstable and the two nonzero fixed points are locally asymptotically stable.

3. *As $\lambda \downarrow 1$,*

$$\boxed{m_\lambda = \sqrt{\frac{3(\lambda - 1)}{\lambda^3}} + O((\lambda - 1)^{3/2}).}$$

Proof. A fixed point satisfies

$$\operatorname{artanh}(m) = \lambda m.$$

For $m > 0$, the function

$$h(m) = \frac{\operatorname{artanh}(m)}{m}$$

is strictly increasing from 1 to $+\infty$. Hence there is no positive solution when $\lambda \le 1$, and exactly one positive solution when $\lambda > 1$. Odd symmetry gives the negative solution.

The derivative is

$$F'_\lambda(m) = \lambda \operatorname{sech}^2(\lambda m).$$

At zero this equals λ, proving stability for $\lambda < 1$ and instability for $\lambda > 1$. At a nonzero fixed point,

$$F'_\lambda(m_\lambda) = \lambda(1 - m_\lambda^2) = \frac{(1 - m_\lambda^2)\operatorname{artanh}(m_\lambda)}{m_\lambda} < 1,$$

because

$$\operatorname{artanh}(m) < \frac{m}{1 - m^2} \qquad (0 < m < 1).$$

The derivative is positive, so the nonzero fixed points are locally stable. Finally, expanding

$$\tanh(\lambda m) = \lambda m - \frac{\lambda^3 m^3}{3} + O(m^5)$$

and solving the nonzero fixed-point equation gives the stated asymptotic. □

The square-root scaling is a genuine model-specific critical law. It does not show that every workspace forms by a pitchfork bifurcation. It supplies an exact benchmark against which more complex formation dynamics can be tested.

71.4 Detecting a transition from finite data

Let $Q(\lambda)$ be a gate-specific population order parameter and $\widehat{Q}_n(\lambda)$ its held-out estimate. If

$$\sup_{\lambda\in\Lambda_n} |\widehat{Q}_n(\lambda) - Q(\lambda)| \le \varepsilon_n$$

on a finite evaluation grid Λ_n, then every threshold crossing with vertical margin larger than $2\varepsilon_n$ is recovered by the same uniform-selection argument as Theorem 120. Near a continuous transition, the margin vanishes and sample requirements necessarily grow. A broad, slowly varying crossover should not be reported as a sharp phase transition without a finite-size scaling analysis.

71.5 General formation-transition conjecture

For adaptive transfer systems with a fixed physical parameterization, there may exist model classes in which

$$\lambda < \lambda_c \Longrightarrow \mathrm{Form}_{\Theta_{\mathrm{form}}}(W_\lambda) = 0,$$

while

$$\lambda > \lambda_c \Longrightarrow \mathrm{Form}_{\Theta_{\mathrm{form}}}(W_\lambda) = 1$$

with a stable nonzero structural mode, positive causal efficacy, and relational irreducibility. The theorem above establishes this only for one structural order parameter in a solvable nonlinear model. The conjunction of all workspace gates remains model dependent.

72. Hierarchical Stabilization Conjecture

For appropriate recurrent systems,

$$W^{(r+1)} = \Gamma_{\varepsilon_r}(W^{(r)})$$

may converge to a stable workspace type:

$$d_{\mathrm{id}}\Big(W^{(r+1)}, W^{(r)}\Big) \to 0.$$

73. Endogenous Boundary Conjecture

The boundary theorems in Part XV establish recovery and stability once a score, candidate family, and formation procedure are fixed. They do not establish that an arbitrary adaptive system will create, represent, and maintain its own boundary.

Let

$$B_t^{\mathrm{sys}}$$

be an internal membership or routing state used by the system, and let

$$S_t^* = \arg\max_{S\in\mathfrak{S}_{\mathrm{pass}}} \mathcal{B}_t(S)$$

be the population boundary defined by the external certification protocol. The strong endogenous boundary conjecture asserts that there are adaptive systems and nontrivial environments for which:

$$\boxed{\begin{aligned}
&d_B(B_t^{\mathrm{sys}}, S_t^*) \to 0,\\
&\mathrm{Form}_{\Theta_{\mathrm{form}}}(W_{B_t^{\mathrm{sys}}}) = 1,\\
&\mathrm{CE}\left(B_t^{\mathrm{sys}} \to \mathrm{Routing}_{t+1}\right) > 0,\\
&\mathrm{CE}\left(\mathrm{BoundaryDamage}_t \to \mathrm{RepairAction}_{t+1}\right) > 0,
\end{aligned}}$$

while the scaffold contains no direct encoding of S_t^*.

A convincing demonstration requires four separable achievements:

1. *discovery*: interaction history changes which boundary is learned;
2. *instantiation*: an internal variable causally realizes that boundary;
3. *use*: changing the boundary changes workspace routing and downstream behavior; and
4. *maintenance*: detected boundary degradation causes restorative action.

Theorems 122 and 123 handle external consistency of the selected boundary. Theorems 128 and 129 now handle controlled repair selection and bounded closed-loop maintenance once a health representation, damage class, monitor, and repair controller are given. The remaining conjectural step is endogenous learning of the internal boundary, detector, and repair policy under open-ended environmental change without direct scaffold encoding or external substitution.

PART XXI

Empirical Certification and Reporting Protocol

74. Required Experiments

A workspace claim is not certified by a single benchmark score. The complete protocol is organized as a sequence of declarations, held-out measurements, implemented interventions, matched nulls, specification sweeps, identification checks, and gate decisions. Every reported result must identify the claim level it supports. Measurements that were not performed remain unmeasured rather than being inferred from neighboring gates.

74.1 Claim declaration before evaluation

Before the certification split is opened, record:

1. the full reference specification $\varsigma_0 = (\mathcal{D}, Y^+, \ell, \tau, \mathcal{I}, \mathcal{N}, \Theta)$;
2. the admissible decomposition family $\mathfrak{D}$ and specification family $\mathfrak{S}$;
3. the target trajectory Y^+, loss, horizon, and environment family;
4. the candidate structural workspace extractor;
5. the goal family $\mathcal{G}$, utility normalization, and feasible action set;
6. the intervention families $\mathcal{I}_W, \mathcal{I}_G, \mathcal{I}_A, \mathcal{I}_D, \mathcal{I}_R$;
7. the null panel and the local distributions each null must preserve;
8. the required certificate level and every numerical threshold;
9. the formation, calibration, certification, and stress-test splits;
10. the uncertainty procedure and resampling unit; and
11. the stopping and model-selection rules;
12. the experiment family and model class used for identifiability; and
13. the complete specification-selection history and multiplicity rule.

Changing any of these after observing certification outcomes creates a new analysis that must be reported as exploratory or evaluated on a fresh split.

74.2 Data and intervention partition

Use at least four logically distinct partitions:

$$\boxed{\mathcal{D} = \mathcal{D}_{\text{form}} \sqcup \mathcal{D}_{\text{cal}} \sqcup \mathcal{D}_{\text{cert}} \sqcup \mathcal{D}_{\text{stress}}.}$$

Their roles are:

Formation. Learn membership, transfer maps, workspace channels, boundaries, policies, and repair models.

Calibration. Select thresholds, intervention amplitudes, and uncertainty procedures without making the final pass decision.

Certification. Estimate the declared gates once under the frozen analysis plan.

Stress test. Evaluate distribution shift, repeated damage, replacement, and long-horizon persistence.

When independent partitions are impossible, the report must state the precise cross-fitting or nested-resampling procedure and preserve formation leakage controls.

74.3 Structural tests

Report the complete low-energy spectrum relevant to the selected threshold:

$$\operatorname{spec}(L_H), \qquad d_\varepsilon, \qquad g_\varepsilon, \qquad \delta_{\mathrm{coh}}.$$

At minimum include:

- the construction and normalization of D_H;
- edge weights and estimated transfer uncertainties;
- the selected eigenspace dimension and threshold rule;
- the spectral gap with uncertainty or perturbation bound;
- stability under bootstrap, seed, and representation replacement; and
- the exact or approximate kernel residual.

74.4 Transfer and operation tests

Report

$$\eta_{\mathrm{int}}, \qquad \delta_e^{\mathrm{tr}}, \qquad \Omega_\gamma, \qquad \Lambda_\gamma,$$

together with:

- the paths and cycle basis used;
- path-length dependence of leakage;
- the gauge-alignment procedure;
- held-out transfer-prediction error;
- operation-intertwining defects; and
- the approximate closure bound from Theorem T_{10}.

74.5 Predictive alignment and coalition tests

Report

$$\delta^{\mathcal{I}}_{\mathrm{align}}, \qquad \{\Delta^{\ell}_A\}_{A\subseteq V}, \qquad k_\varepsilon, \qquad \delta^{Z}_A.$$

When the number of coalitions is too large for exhaustive evaluation, state the sampling design, include every singleton and every leave-one-out coalition, and report the most favorable coalition found for the reducibility null. The claim must not be based only on average coalition performance.

74.6 Causal and reentry tests

Report

$$\eta^{\nu}_{\text{causal}}, \qquad \Delta_{\text{specific}}(r), \qquad \delta_{\text{med}}, \qquad \eta_{\text{rec}}(\tau), \qquad \eta_{\text{loop}}(\tau).$$

Every causal estimate must be paired with:

1. a concrete mechanism-level implementation;
2. a manipulation check;
3. matched background variables;
4. declared direct paths that were blocked or preserved;
5. intervention magnitude and cost;
6. an off-workspace matched intervention; and
7. the downstream outcome window.

74.7 Persistence, reconstruction, and replacement tests

Report

$$d_{\text{id}}(W_t, W_{t+\tau}), \qquad \text{Rec}_k, \qquad e_t, \qquad \alpha^{(k)}, \qquad \mu_t^{\text{keep}}.$$

The report must separate:

- subspace tracking from state tracking;
- coherence from non-collapse;
- reconstruction from mere prediction;
- repair from successor-workspace creation; and
- structural replacement from causal and reconstructive replacement.

74.8 Formation-provenance tests

Report

$$\mathbf{E}_{\text{form}}, \qquad \delta^{\text{scaf}}, \qquad \delta^{\text{leak}}, \qquad \Delta_q^{\text{hist}}.$$

Include the external scaffold, information available to the learner, history shuffles, cross-scaffold retraining, and every route by which certification labels or target boundaries could have entered formation.

74.9 Learned-structure and boundary tests

Report

$$\|\widehat{H} - H_*\|, \qquad \Delta_J, \qquad \Delta_t^B, \qquad \text{COut}_\tau(\widehat{S}), \qquad \text{CIn}_\tau(\widehat{S}).$$

Boundary reports must include the admissible candidate family, full-system and empty-set controls, singleton and centralized controls, score components, margin to the runner-up, and temporal stability under held-out environment drift.

74.10 Goal and counterfactual-policy tests

Report the agency coordinates

$$\eta_G(\tau), \qquad \delta_G^{\mathrm{cue}}(\tau), \qquad \eta_{G\to A}, \qquad \delta_Q, \qquad \eta_{V\to A}, \qquad \eta_{A\to Y}, \qquad k_{\mathrm{ctl},\varepsilon}.$$

The required experiments include:

1. cue-controlled goal interventions;
2. goal-channel ablation and permutation;
3. counterfactual-value calibration against action interventions;
4. value-label permutation among feasible actions;
5. critic or value-channel blocking;
6. action replay under held-out perturbations;
7. policy regret or value-gap estimation; and
8. coalition reconstruction of the complete control state.

74.11 Damage and repair tests

For every damage family, report

$$\eta_{D\to M}, \qquad \eta_{M\to R}, \qquad \eta_{R\to v}, \qquad \Delta_{\mathrm{repair}}^{\mathrm{spec}}, \qquad \delta_{\mathrm{rep-id}}, \qquad \sup_t e_t.$$

The repeated-disturbance experiment must contain:

1. a no-damage baseline;
2. damage with repair disabled;
3. damage with the intact repair loop;
4. monitor ablation;
5. cost-matched random repair;
6. off-workspace repair;
7. repeated and compound damage; and
8. a held-out disturbance distribution.

75. Required Null Families

The core null panel is:

$$
\begin{aligned}
N_1 &= \text{shuffled transfer maps with validated local matching},\\
N_2 &= \text{shuffled cross-node temporal correspondence},\\
N_3 &= \text{matched-capacity local memory},\\
N_4 &= \text{matched-size irrelevant artifact},\\
N_5 &= \text{participant ablation},\\
N_6 &= \text{artifact ablation},\\
N_7 &= \text{workspace implementation intervention},\\
N_8 &= \text{matched off-workspace intervention},\\
N_9 &= \text{representation replacement},\\
N_{10} &= \text{context reset and reconstruction},\\
N_{11} &= \text{matched formation-history shuffle},\\
N_{12} &= \text{cross-scaffold re-encoding and retraining},\\
N_{13} &= \text{boundary-label and certification leakage audit},\\
N_{14} &= \text{empty, full, singleton, and centralized boundary nulls},\\
N_{15} &= \text{held-out transfer and membership re-estimation},\\
N_{16} &= \text{boundary damage, use, and repair intervention},\\
N_{17} &= \text{goal-code permutation with cues and actions matched},\\
N_{18} &= \text{persistent external-cue rewrite with internal retention cut},\\
N_{19} &= \text{counterfactual-value label permutation},\\
N_{20} &= \text{critic or value-channel ablation},\\
N_{21} &= \text{open-loop action replay under held-out perturbation},\\
N_{22} &= \text{monitor-blind cost-matched random repair},\\
N_{23} &= \text{off-workspace repair with matched intervention norm},\\
N_{24} &= \text{identity-breaking reconstruction with matched task score},\\
N_{25} &= \text{merge control for components sharing an uninterruptible controller},\\
N_{26} &= \text{secret-sharing split of a known centralized controller},\\
N_{27} &= \text{hidden shared-memory and common-cue exposure test},\\
N_{28} &= \text{frozen certificate sweep across the admissible decomposition family},\\
N_{29} &= \text{post hoc target-selection audit},\\
N_{30} &= \text{horizon cherry-picking audit},\\
N_{31} &= \text{loss and utility substitution audit},\\
N_{32} &= \text{restricted intervention-family audit},\\
N_{33} &= \text{threshold and gate-set tuning audit},\\
N_{34} &= \text{frozen certificate sweep across the admissible specification family}.
\end{aligned}
$$

75.1 Null validity record

Each observational null must state:

- the local distributions it preserves;

- the relational dependency it destroys;
- the distance between intact and null local marginals;
- the capacity, parameter count, and information budget; and
- whether null fitting used the same data budget as the intact model.

Each causal null must state:

- the manipulated mechanism;
- the matched background and intervention cost;
- the intended causal path destroyed or preserved;
- the manipulation check; and
- unintended direct effects detected during validation.

A failed manipulation check invalidates the null. It does not count as evidence for the workspace claim.

75.2 Claim-specific null subsets

A successful claim requires

$$\mathrm{Pass}(N_1, \dots, N_{28})$$

only for the subset relevant to its stated level. The subset must be frozen before certification. Failed applicable nulls, invalid nulls, and unmeasured nulls are retained separately in the final report. They must not be merged into a generic "not significant" category.

76. Certification Vector and Complete Reporting Templates

76.1 Extended certification vector

Use a vector rather than one aggregate score:

$$\begin{aligned}\mathcal{C}_{\mathrm{full}}(W) = (&d_\varepsilon, g_\varepsilon, \delta_{\mathrm{coh}}, \delta^{\mathcal{I}}_{\mathrm{align}}, \Delta^{\ell}_{\mathrm{joint}}, k_\varepsilon,\\ &\boldsymbol{\Delta}_{\mathrm{tr}}(\mathcal{Q}), \eta^{\nu}_{\mathrm{causal}}, \Delta_{\mathrm{specific}}(r), \eta_{\mathrm{loop}}(\tau),\\ &d_{\mathrm{id}}, \mathrm{Rec}_k, \underline{\mathbf{m}}_{\mathfrak{D}}, \boldsymbol{\Delta}_{\mathrm{dec}}, \mathbf{E}_{\mathrm{form}}, \delta^{\mathrm{scaf}}, \delta^{\mathrm{leak}},\\ &\Delta_J, \Delta^{B}_t, \mathrm{COut}_\tau(\widehat{S}), \mathrm{CIn}_\tau(\widehat{S}),\\ &\eta_G(\tau), \delta^{\mathrm{cue}}_G(\tau), \eta_{G\to A}, \delta_Q, \eta_{V\to A}, \eta_{A\to Y},\\ &\eta_{D\to M}, \eta_{M\to R}, \eta_{R\to v}, \Delta^{\mathrm{spec}}_{\mathrm{repair}}, \delta_{\mathrm{rep-id}}, \sup_t e_t).\end{aligned}$$

A claim passes only when every required coordinate lies in its predeclared admissible region. No weighted average may compensate for a failed required gate.

76.2 Claim card template

Field	Required entry
Claim identifier	Unique versioned identifier tied to code and result artifacts.
Claim level	Archive, shared representation, operational workspace, relational workspace, persistent workspace, self-maintaining workspace, or workspace agent.
Reference decomposition	Nodes, artifacts, channels, local-access maps, and coalition universe.
Admissible decomposition family	Plausible merges, splits, shared-resource assignments, exclusions, and audit rationale.
Target and horizon	Y^+, loss, utility, prediction horizon, control horizon, and maintenance horizon.
Environment scope	Training, certification, and stress-test environment families.
Workspace extractor	Frozen construction of D_H, L_H, spectral threshold, and state readout.
Goal and policy	Goal codes, utility normalization, action set, and policy class.
Boundary	Candidate family, score, admissibility constraints, and selected boundary.
Interventions	Feasible workspace, goal, action, damage, and repair implementations.
Null panel	Applicable N_j, preservation targets, and fit budgets.
Thresholds	Every gate direction, numerical cutoff, and uncertainty rule.
Data split	Formation, calibration, certification, stress, and cross-fitting logic.
Stopping rule	Number of seeds, runs, interventions, and any sequential stopping criterion.
Status	Confirmatory, exploratory, replication, or audit.

76.3 Gate-estimate table template

For every required coordinate, include one row with the following fields:

Field	Required entry
Gate	Formal symbol and verbal interpretation.
Estimator	Exact statistic, normalization, and software implementation.
Evaluation unit	Episode, participant, graph, seed, environment, or other independent unit.
Point estimate	Estimate on the frozen certification split.
Uncertainty	Interval, bound, or posterior region with declared coverage or credibility.
Threshold	Predeclared pass region.
Manipulation status	Passed, failed, or not applicable.
Decision	Pass, fail, invalid, underpowered, or unmeasured.
Sensitivity	Result under reasonable alternative normalizations and resampling choices.
Artifact reference	File, table, raw-result key, and code commit.

76.4 Intervention ledger template

Field	Required entry
Intervention ID	Unique identifier and randomization seed.
Target mechanism	Workspace coordinate, goal code, value channel, action, boundary, monitor, or repair channel.
Implementation	Exact code path, hook, artifact edit, or mechanism change.
Dose	Norm, duration, timing, and support of the intervention.
Background match	Variables fixed, blocked, randomized, or balanced.
Excluded paths	Direct paths intentionally removed from the causal contrast.
Manipulation check	Evidence that the intended mechanism changed by the specified amount.
Collateral effects	Changes outside the target mechanism.
Outcome	Downstream variable and evaluation window.
Effect estimate	Point estimate, uncertainty, and sign convention.
Cost	Compute, latency, information, action, or structural cost.
Validity status	Valid, invalid, ambiguous, or unmeasured.

76.5 Null ledger template

Field	Required entry
Null ID	One of $N_1, \dots, N_{34}$ or a declared extension.
Claim attacked	The exact explanation the null is designed to preserve or destroy.
Matched quantities	Local marginals, capacity, training budget, action cost, intervention norm, or other controlled quantities.
Destroyed relation	Edge alignment, temporal order, value ranking, goal identity, monitor mediation, or another relational feature.
Fit procedure	Data, optimization, seeds, and hyperparameter budget.
Validity metric	Distance from the intended matched quantities.
Primary contrast	Intact minus null effect with uncertainty.
Disposition	Rejects null, fails to reject, invalid null, or underpowered.
Failure notes	Any mismatch or collateral mechanism change.

76.6 Agency and maintenance report template

The agency section must provide the following matrix:

Component	Primary metric	Required contrast
Goal semantics	Held-out Q_t^g ordering	Distinct goals reverse or alter at least one declared action ordering.
Goal retention	$\eta_{G}(\tau)$	Goal intervention with future cues matched or removed.
Goal use	$\eta_{G\to A}$	Goal intervention versus goal-label replay.
Counterfactual accuracy	δ_Q	Internal value versus external action interventions.
Value use	$\eta_{V\to A}$	Value-label permutation and value-channel ablation.
Outcome control	$\eta_{A\to Y}$	Feasible action intervention.
Damage detection	$\eta_{D\to M}$	Damage versus matched no-damage.
Repair mediation	$\eta_{M\to R}$	Monitor intervention and monitor ablation.
Restoration	$\eta_{R\to v}$	Repair versus matched no-repair.
Repair specificity	$\Delta^{\mathrm{spec}}_{\mathrm{repair}}$	Intended repair versus cost-matched generic action.
Identity preservation	$\delta_{\mathrm{rep-id}}$	Pre-damage workspace versus post-repair workspace.
Long-run maintenance	$\sup_t e_t$	Repeated held-out disturbances.
Relational location	$k_{\mathrm{ctl},\varepsilon}$	Full system versus all single nodes and matched local nulls.

76.7 Uncertainty and multiplicity

The independent resampling unit must match the data-generating process. Token, time-step, or edge resampling is invalid when dependence is organized at the episode, participant, graph, or environment level. Recommended procedures include:

- environment- or episode-level bootstrap;
- repeated independent training seeds;
- cluster-robust intervals for repeated observations;
- permutation tests that preserve the declared local structure;
- concentration bounds for uniform finite candidate families; and
- stress-test intervals over held-out environment families.

Pass decisions should use one-sided uncertainty bounds. For a lower-bounded gate $\theta \geq \tau$, pass only when the lower confidence bound exceeds τ. For an upper-bounded defect $\delta \leq \varepsilon$, pass only when the upper confidence bound is below ε.

When a conclusion ranges over $\mathfrak{S}$, the uncertainty rule must be uniform over every gate and specification included in the claim. Report the pointwise interval, the family-wide lower envelope, the covering or multiplicity correction, and the identification status. A narrow pointwise interval does not establish a robust result when model or specification uncertainty remains wide.

76.8 T_{24}: Gatewise certification validity

Theorem 156 (Intersection-gate false-certification control)**.** *Let a certificate require all gates $G_1, \ldots, G_m$ to pass. For each gate j, let H_{0j} denote the null that the required property is absent, and assume the predeclared test satisfies*

$$\sup_{P \in H_{0j}} \mathbb{P}_P(G_j \text{ passes}) \leq \alpha_j.$$

The global null is

$$H_0 = \bigcup_{j=1}^{m} H_{0j},$$

meaning that at least one required property is absent. If the certificate is declared only when every gate passes, then

$$\boxed{\sup_{P \in H_0} \mathbb{P}_P(\textit{certificate passes}) \le \max_{1 \le j \le m} \alpha_j.}$$

No independence assumption among gate tests is required.

Proof. For any $P \in H_0$, choose an index j_P such that $P \in H_{0j_P}$. The event that every gate passes is contained in the event that gate j_P passes. Hence

$$\mathbb{P}_P(\text{certificate passes}) \le \mathbb{P}_P(G_{j_P} \text{ passes}) \le \alpha_{j_P} \le \max_j \alpha_j.$$

Taking the supremum over $P \in H_0$ proves the claim. □

The theorem is an intersection-union guarantee. It applies only when the gate family, thresholds, and tests are fixed before certification and each gate test is uniformly valid for its null. Adaptive gate deletion, selective reporting, or repeated reuse of the certification split invalidates the bound.

76.9 Failure taxonomy

Every gate and null receives exactly one primary status:

Pass. The frozen estimate and uncertainty rule satisfy the threshold.

Fail. The gate was validly measured and did not satisfy the threshold.

Invalid intervention. The manipulation check or background match failed.

Invalid null. The null did not preserve its required local quantities or exceeded its fit budget.

Underpowered. The uncertainty interval crosses the threshold by more than the declared indifference region.

Unmeasured. No valid estimate was obtained.

Leakage. Formation or model selection used certification information.

Exploratory. The metric, threshold, or analysis was selected after inspecting certification outcomes.

Only *Pass* counts toward the formal certificate. "Underpowered" and "unmeasured" are not failures of the underlying theory, but they are failures to establish the corresponding empirical claim.

76.10 Machine-readable result record

The machine-readable record should contain at least the following top-level objects:

$$\boxed{\begin{aligned}\mathsf{ResultRecord} = \{&\mathsf{claim}, \mathsf{specification}, \mathsf{specification_family}, \mathsf{decomposition}, \mathsf{data_splits}, \mathsf{formation},\\ &\mathsf{workspace_extractor}, \mathsf{goals}, \mathsf{actions}, \mathsf{interventions},\\ &\mathsf{nulls}, \mathsf{gates}, \mathsf{identified_sets}, \mathsf{specification_envelopes}, \mathsf{uncertainty}, \mathsf{failures}, \mathsf{artifacts}\}.\end{aligned}}$$

Each gate object records its symbol, estimator, direction, threshold, point estimate, interval, split, seed set, status, and artifact reference. Each intervention and null object records its manipulation check and validity status. The schema must preserve failed and invalid entries rather than only the passing subset.

76.11 Minimum reproducibility package

A complete certification package contains:

1. the frozen claim card;
2. source code and environment lock file;
3. exact data or deterministic data-generation procedure;
4. formation, calibration, certification, and stress-test manifests;
5. learned transfer, workspace, policy, monitor, and repair parameters;
6. raw per-run and per-intervention results;
7. the intervention and null ledgers;
8. the complete gate table, including failures;
9. figures and tables regenerated from raw results;
10. random seeds and hardware or numerical-precision information;
11. checksums for every included artifact; and
12. a single verification command that reproduces package integrity and the reported summary tables.

76.12 Decision rule

Let $\mathcal{J}(L)$ be the required gate set for claim level L. Then

$$\boxed{\mathrm{Certify}_L(W;\varsigma) = \bigwedge_{j\in\mathcal{J}(L,\varsigma)} \mathrm{Pass}_j(W;\varsigma).}$$

For a workspace-agent claim, the required set includes every structural, operational, relational, persistent, goal, counterfactual-control, and self-maintenance gate declared in Parts IX and XVI. A specification-robust claim additionally requires the uniform rule of Part XVIII across $\mathfrak{S}$. Failure of an earlier level prevents certification of a later level, even when later-level behavior appears strong on a task benchmark.

PART XXII

Integrated Worked Examples and Case Studies

77. Case-Study Method and Common Gate Card

The preceding parts define a large family of structural, operational, causal, relational, persistent, formative, and agentive predicates. A worked example is useful only when it makes clear which of those predicates are established, which are assumed, and which remain unmeasured. This part therefore uses a single case-study protocol for three different kinds of realization:

1. an exact finite-dimensional model of a human–human theorem laboratory;
2. an idealized human–AI–artifact–executor research loop, followed by a deliberately conservative audit of the collaboration that motivated the theory; and
3. a closed-form multi-agent control system that reaches the formal agency level inside a small declared domain.

The first and third cases are analytic models. Their conclusions hold inside those models. The second contains both an analytic idealization and an observational audit. The audit is not upgraded to a certificate merely because the idealization is mathematically clean.

Definition 157 (Case-study realization)**.** A case-study realization is a tuple

$$\mathfrak{K} = (\mathfrak{R}, \mathfrak{B}, \mathfrak{C}, \mathfrak{P}, \mathcal{E}, \mathfrak{D}, \mathfrak{S}, \mathcal{I}, \mathcal{Q}, \mathcal{G}),$$

where:

- $\mathfrak{R}$ is the declared relational realization;
- $\mathfrak{B}$ is the boundary problem and component decomposition;
- $\mathfrak{C}$ is the structural core–transfer model;
- $\mathfrak{P}$ is the downstream predictive or control problem;
- $\mathcal{E}$ is the evidence or generated-data family;
- $\mathfrak{D}$ is the admissible decomposition family;
- $\mathfrak{S}$ is the admissible family of complete scientific specifications, including targets, losses, horizons, interventions, nulls, and thresholds;
- $\mathcal{I}$ is the feasible intervention family;
- $\mathcal{Q}$ is the matched null family; and
- $\mathcal{G}$ is the set of gates being claimed.

A conclusion is always relative to this tuple. Changing the downstream target, temporal horizon, intervention support, component decomposition, null family, or threshold rule creates a different specification of the case. A stronger case study reports whether the conclusion survives the predeclared family $\mathfrak{S}$ and whether the relevant gates are identified by the available experiment family.

77.1 Three evidential statuses

Every claim in this part receives one of three primary statuses:

Analytic. The gate follows exactly from a declared mathematical model.

Empirically certified. The gate is estimated from held-out data using the protocol of Part XXI and passes its frozen uncertainty rule.

Observational only. The case contains suggestive records or natural perturbations, but no valid held-out intervention-and-null certificate.

An analytic result is not automatically an empirical result. An observational pattern is not automatically a causal result. None of the three statuses establishes phenomenal consciousness.

77.2 Common gate card

For each case we report the vector

$$\boxed{\mathsf{Card}(\mathfrak{K}) = (\mathsf{S}, \mathsf{O}, \mathsf{R}, \mathsf{P}, \mathsf{F}, \mathsf{A}, \mathsf{D}, \mathsf{X}, \mathsf{C})\,,}$$

where:

S = structural core and spectral coherence,
O = operational alignment and causal efficacy,
R = relational irreducibility and matched-null gap,
P = persistence, reconstruction, and identity,
F = formation provenance and boundary status,
A = goal-conditioned agency and self-maintenance,
D = decomposition robustness and interface audit,
X = specification robustness, identification, and uniform uncertainty,
C = consciousness or personhood claim.

The decomposition coordinate distinguishes a result established only for the declared interface from one stable across an admissible family. The specification coordinate records whether the same conclusion survives scientifically defensible alternatives, whether its sign is identified by the available experiments, and whether its lower margin remains positive after uniform finite-sample uncertainty. The final coordinate is included to make the non-implication explicit. In all examples below,

$$\mathsf{C} = \text{not established.}$$

77.3 Claim-level discipline

Let $L(\mathfrak{K}) \in \{0, 1, \dots, 6\}$ denote the highest level from Part XXIV whose complete required gate set is established. The level is conjunctive and pointwise in the declared scientific specification unless the D and X coordinates also pass:

$$L(\mathfrak{K}) \geq \ell \quad \Longleftrightarrow \quad \bigwedge_{j \in \mathcal{J}(\ell)} \mathrm{Pass}_j(\mathfrak{K}).$$

A case may have strong higher-level-looking behavior while remaining at a lower certified level because an earlier gate is absent. For example, successful repair behavior does not establish self-maintenance when no matched no-repair counterfactual is available, and coherent shared documents do not establish a relational workspace when one participant contains the complete operative state.

77.4 Structural calculations used repeatedly

The positive examples use a recurring distributed code. Let

$$c = (c_1, c_2, c_3)^{\mathsf{T}} \in \mathbb{R}^3$$

and define three local encodings

$$E_1 c = (c_1, c_2), \qquad E_2 c = (c_2, c_3), \qquad E_3 c = (c_1, c_3).$$

Then

$$\sum_{i=1}^{3} E_i^* E_i = 2I_3.$$

No individual encoding is injective, while every two-node coalition is injective. The pairwise frame operators are

$$E_{\{1,2\}}^* E_{\{1,2\}} = \operatorname{diag}(1, 2, 1),$$
$$E_{\{1,3\}}^* E_{\{1,3\}} = \operatorname{diag}(2, 1, 1),$$
$$E_{\{2,3\}}^* E_{\{2,3\}} = \operatorname{diag}(1, 1, 2).$$

Hence

$$\alpha_{\{i,j\}} = 1 \qquad (i \neq j),$$

and one-node erasure recovery is exact. Every nonzero core vector is visible at at least two nodes, so

$$d_{\mathrm{er}} = 2.$$

This code is distributed but redundant: no node is sufficient, yet any one node may be lost without destroying the core.

77.5 Binary synergy test

Let C_1, C_2, C_3 be independent uniform bits and let

$$Y = C_1 \oplus C_2 \oplus C_3.$$

Each node sees two of the three bits. Therefore

$$H(Y \mid E_i C) = 1 \text{ bit} \qquad \text{for every } i,$$

while

$$H(Y \mid E_1 C, E_2 C, E_3 C) = 0.$$

Under logarithmic loss,

$$\mathcal{R}_{\log}(\{i\}) = 1, \qquad \mathcal{R}_{\log}(V) = 0, \qquad \Delta_{\mathrm{joint}}^{\log} = 1 \text{ bit}, \qquad k_0 = 2.$$

Moreover, Y is independent of every individual local state. The null law that keeps all three core bits independent and makes Y an independent uniform bit preserves every pair $(Y, E_i C)$ while destroying the joint parity relation. It is therefore a predictive-marginal-matched relational null with

$$\Delta_{\mathrm{tr}} = 1 \text{ bit}.$$

This test will be reused with different semantic interpretations.

78. Case Study A: A Distributed Human–Human Theorem Laboratory

78.1 Declared realization

Consider two researchers and one persistent notebook:

$$V_{\mathrm{HH}} = \{P, C, N\},$$

where:

- P is a proposer who tracks the theorem statement and open proof obligations;
- C is a critic who tracks proof obligations and falsification status;
- N is a notebook that tracks the theorem statement and falsification status.

The abstract core is

$$c = (q, p, f)^{\mathsf{T}} \in \mathbb{R}^3,$$

with

$$q = \text{statement coordinate}, \qquad p = \text{proof-obligation coordinate},$$
$$f = \text{falsification-status coordinate}.$$

The accessible encodings are

$$E_P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad E_C = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad E_N = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The semantics are stylized. The point is not that real researchers contain only two numbers. The point is to exhibit, without metaphor, a system in which no participant or artifact contains the entire operative state while the coupled system does.

78.2 Exact overlap descent

Order the local coordinates as

$$x = (q_P, p_P, p_C, f_C, q_N, f_N)^{\mathsf{T}}.$$

Compatibility requires

$$p_P = p_C, \qquad f_C = f_N, \qquad q_N = q_P.$$

The discrepancy operator is

$$D_{\mathrm{HH}} = \begin{pmatrix} 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 \\ -1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Therefore

$$\ker D_{\mathrm{HH}} = \left\{ (q, p, p, f, q, f)^{\mathsf{T}} : q, p, f \in \mathbb{R} \right\} \cong \mathbb{R}^3.$$

The rows are orthogonal with squared norm two, so

$$D_{\mathrm{HH}} D_{\mathrm{HH}}^{*} = 2I_3.$$

Consequently the transfer Laplacian has spectrum

$$\boxed{\operatorname{spec}(L_{\mathrm{HH}}) = \{0, 0, 0, 2, 2, 2\},}$$

and the exact workspace gap is

$$\gamma_{\mathrm{HH}} = 2.$$

The low-energy subspace is therefore stable under any Laplacian perturbation whose operator norm is strictly below one, with the quantitative projector bound from Theorem 43.

78.3 Reconstruction and one-node failure

The full frame operator is

$$F_{PCN} = E_P^* E_P + E_C^* E_C + E_N^* E_N = 2I_3.$$

For every two-node coalition S,

$$\lambda_{\min}(F_S) = 1.$$

Thus every two-node coalition reconstructs the complete core and, for noisy observations $y_S = A_S c + \eta_S$,

$$\|\widehat{c} - c\| \le \|\eta_S\|.$$

No individual node reconstructs c, because

$$\operatorname{rank}(E_P) = \operatorname{rank}(E_C) = \operatorname{rank}(E_N) = 2 < 3.$$

The realization is therefore simultaneously:

$$\text{individually incomplete} \quad \text{and} \quad \text{robust to any one-node loss.}$$

78.4 Relational prediction problem

Let q, p, f now be independent uniform bits and define the downstream acceptance decision

$$Y = q \oplus p \oplus f.$$

The interpretation is abstract: acceptance depends on the statement, the state of the proof obligations, and the current falsification record. No individual node has sufficient information:

$$H(Y \mid X_P) = H(Y \mid X_C) = H(Y \mid X_N) = 1.$$

Every two-node coalition has all three bits and therefore zero conditional entropy. Hence

$$\boxed{\Delta_{\mathrm{joint}}^{\log} = 1 \text{ bit}, \qquad k_0 = 2.}$$

Let Q_{HH} preserve the independent uniform local bit pairs but replace Y by an independent uniform bit. Because parity is independent of every two-bit local view, Q_{HH} preserves each (Y, X_i)-marginal. Nevertheless,

$$\mathcal{R}_{\log, Q_{\mathrm{HH}}}(V) - \mathcal{R}_{\log, P_{\mathrm{HH}}}(V) = 1 \text{ bit}.$$

This is an exact relational witness: the predictive capacity lies in the joint organization rather than in any local predictive marginal.

78.5 Feasible intervention and specificity

A feasible statement intervention must alter both local realizations of the statement coordinate:

$$\iota_{q \leftarrow a} : (q_P, q_N) \leftarrow (a, a),$$

while holding (p, f) fixed. For configurations with fixed p and f, changing a flips the downstream decision with total-variation effect one:

$$\mathrm{TV}\left(P(Y \mid \iota_{q\leftarrow 0}), P(Y \mid \iota_{q\leftarrow 1})\right) = 1.$$

A matched intervention on a residual notebook-format coordinate leaves Y unchanged. Thus the core intervention has positive causal specificity in this model.

78.6 Persistence, handoff, and replacement

Suppose the proposer becomes unavailable. The remaining coalition $\{C, N\}$ reconstructs (q, p, f) exactly. A replacement proposer P' can therefore be initialized by

$$E_{P'}c = (q, p).$$

If P' uses a different local coordinate system

$$J_{P'} \in GL(2),$$

then the incoming and outgoing overlap maps are conjugated accordingly. The global section and all downstream kernels remain unchanged. This is exact replacement invariance, not literal carrier identity.

If the notebook alone survives, the core cannot be reconstructed because the proof-obligation coordinate is missing. The model therefore predicts a sharp handoff boundary:

$$\text{one surviving node is insufficient,} \qquad \text{any two surviving nodes are sufficient.}$$

78.7 Case card

Inside the declared finite model,

$\mathsf{S} = \checkmark_{\text{analytic}}$,	$\mathsf{O} = \checkmark_{\text{analytic}}$,
$\mathsf{R} = \checkmark_{\text{analytic}}$,	$\mathsf{P} = \checkmark_{\text{analytic}}$,
F = scaffolded,	A = not tested,
D = fixed decomposition,	C = not established.

The highest supported level is Level 4 when the handoff dynamics and identity correspondence are included. The example does not contain an operational goal state, counterfactual policy, damage monitor, or repair loop, so it does not support Levels 5 or 6.

79. Case Study B: A Human–AI–Artifact–Executor Research Loop

79.1 Idealized four-node loop

Consider four heterogeneous components:

$$V_{\mathrm{HAI}} = \{H, M, X, A\},$$

where:

- H is a human researcher who maintains goals and semantic direction;

- M is a language-model process that constructs local formalizations;
- X is an execution and verification process that compiles, tests, or falsifies candidate artifacts; and
- A is a persistent artifact store containing exact source, results, and versioned handoff state.

Let the abstract research state be

$$c = (g, d, p, v)^{\mathsf{T}} \in \mathbb{R}^4,$$

where:

g = goal and acceptance criterion, d = current derivation or theory state,
p = open proof or implementation obligations, v = verification and regression status.

Define local encodings

$$E_H c = (g, d), \qquad E_M c = (g, p), \qquad E_X c = (p, v), \qquad E_A c = (d, v).$$

Each coordinate is represented at exactly two nodes, and the incidence pattern forms a four-cycle.

79.2 Exact structural calculation

Order the local coordinates as

$$x = (g_H, d_H, g_M, p_M, p_X, v_X, d_A, v_A)^{\mathsf{T}}.$$

The overlap discrepancies are

$$g_H - g_M, \qquad p_M - p_X, \qquad v_X - v_A, \qquad d_A - d_H.$$

Thus

$$D_{\mathrm{HAI}} = \begin{pmatrix} 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & -1 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

The rows are orthogonal with squared norm two. Therefore

$$\ker D_{\mathrm{HAI}} \cong \mathbb{R}^4,$$

$$\boxed{\mathrm{spec}(L_{\mathrm{HAI}}) = \{0, 0, 0, 0, 2, 2, 2, 2\}, \qquad \gamma_{\mathrm{HAI}} = 2.}$$

The full accessible frame operator is

$$F_{HMXA} = 2I_4.$$

No single node is injective. Every three-node coalition reconstructs the full state, and the two opposite pairs

$$\{H, X\}, \qquad \{M, A\}$$

are already sufficient. Adjacent pairs observe only three of the four coordinates. Hence

$$k_{\mathrm{rec}} = 2, \qquad d_{\mathrm{er}} = 2,$$

while the identity of the sufficient two-node coalition depends on the graph geometry.

79.3 Relational parity benchmark

Let g, d, p, v be independent uniform bits and define

$$Y = g \oplus d \oplus p \oplus v.$$

Every single node sees only two bits, so

$$H(Y \mid X_i) = 1 \qquad \text{for all } i.$$

The opposite pairs $\{H, X\}$ and $\{M, A\}$ determine all four bits and therefore determine Y. Every adjacent pair remains one bit short. It follows that

$$\boxed{\Delta^{\log}_{\text{joint}} = 1 \text{ bit}, \qquad k_0 = 2.}$$

As in the preceding case, parity is independent of every individual local view. Replacing Y by an independent uniform bit produces a matched local predictive null and a full-system transfer gap of one bit.

This benchmark is intentionally narrow. It demonstrates that the four-cycle can realize exact relational information. It does not show that every real human–AI collaboration has this encoding, this target, or this degree of irreducibility.

79.4 A process-level interpretation

The idealized loop can be read as the recurrent sequence

$$\boxed{H \xrightarrow{g} M \xrightarrow{p} X \xrightarrow{v} A \xrightarrow{d} H.}$$

A more explicit cycle is:

1. the human supplies a goal and acceptance criterion;
2. the model converts that goal into a formal derivation and explicit obligations;
3. the executor compiles, tests, or searches for counterexamples;
4. the artifact store preserves the exact source and verification record;
5. the human evaluates the returned derivation and evidence, revising the next goal.

The operative state is reentrant because products of one cycle become inputs to the next. The artifact is not merely a record when its contents alter the next formalization, test, or acceptance decision.

79.5 Conservative audit of the motivating collaboration

The collaboration that motivated this theory has several features consistent with the four-node description: specialized shared vocabulary, versioned LaTeX and code artifacts, repeated compile-and-repair cycles, explicit theorem and null-control demands, and reconstruction across conversations from handoff packages. Those observations justify treating the collaboration as a serious candidate case. They do not establish the complete certificate.

Gate family	Present evidential status	Conservative assessment
Declared heterogeneous components	Observational	Human, model, artifacts, and execution environment are distinguishable components with different roles and information access.
Persistent shared vocabulary	Observational	Recurrent terms and distinctions survive across research phases and constrain later work.
Exact structural core $\Gamma(H)$	Unmeasured	No frozen local coordinate system or learned transfer family has yet been estimated from held-out data.
Transfer-Laplacian gap	Unmeasured	The idealized value 2 belongs to the toy four-cycle, not to the actual collaboration.
Operational alignment	Unmeasured	No preregistered future variable has been shown conditionally independent of the full history given a measured workspace state.
Joint advantage	Suggestive only	The coupled process has produced outputs that neither party appears to have supplied verbatim, but local-only Bayes risks have not been measured under controlled information ceilings.
Matched relational null gap	Unmeasured	Shuffled artifacts, matched irrelevant memory, and local-predictive-marginal-preserving nulls have not been run as a frozen panel.
Causal workspace efficacy	Suggestive only	Corrections, compiler failures, and artifact substitutions change later behavior, but they are not yet matched workspace interventions with manipulation checks.
Artifact-mediated persistence	Observational	Handoff files and source bundles support continuation, but blind reconstruction and ablation curves have not been estimated.
Formation provenance	Mixed	The communication architecture is scaffolded, while terminology, theorem organization, and many control practices were adaptively developed. Causal formation-history contribution is unmeasured.
Workspace agency	Not established	No complete goal-retention, counterfactual-policy, distributed-control, damage, and repair certificate has been run.
Consciousness or personhood	Not established	The theory supplies no such inference from the observed collaboration.

The correct present statement is therefore:

candidate persistent collaborative workspace rather than certified relational workspace agent.

79.6 A controlled certification design

A serious test can use one manuscript section or software work unit as the unit of analysis. Before the certification split is opened, freeze:

- the next-stage task family;
- the downstream target, such as accepted theorem patch, successful clean compile, and absence of a declared regression;
- the local information available to each component;
- the workspace extractor and transfer maps;

- all intervention and null families; and
- the required gate thresholds.

The following conditions can then be compared:

$$\begin{aligned} \mathsf{H} &= \text{human-side externalized state only,} \\ \mathsf{M} &= \text{fresh model instance with model-side state only,} \\ \mathsf{A} &= \text{artifact package without live human or model state,} \\ \mathsf{X} &= \text{executor and test outputs without the research context,} \\ \mathsf{F}\mathit{ull} &= \text{the intact recurrent loop.} \end{aligned}$$

The design must avoid pretending that a person's unobserved internal state has been fully measured. The local conditions certify only the declared accessible histories.

A minimal null panel includes:

1. theorem-to-artifact reassignment with artifact size and style matched;
2. temporal shuffling of verification feedback;
3. matched irrelevant manuals or source bundles;
4. context reset followed by blind reconstruction;
5. removal of compiler or test feedback while preserving message count;
6. model replacement under a fixed handoff package;
7. artifact removal with equivalent token budget restored by irrelevant text;
8. a local-marginal-matched synthetic null for the frozen downstream target.

A valid empirical claim would require the intact loop to outperform every required local and matched-null condition, while the extracted workspace state remains sufficient and feasible workspace interventions alter the downstream result in the predicted direction.

79.7 Handoff and identity experiment

Let B_t be the source-and-results bundle at the end of stage t. A blind handoff trial replaces the active model instance and optionally the human analyst, then asks the replacement coalition to reconstruct:

$$(\text{current definitions}, \text{proved results}, \text{open obligations}, \text{next admissible research step}) .$$

The reconstruction error is measured against a frozen adjudication set. A replacement invariance test additionally compares:

- the transported structural workspace;
- operation behavior on held-out transformations;
- response to matched interventions; and
- preservation of failed and unresolved claims.

Equivalent benchmark prose alone is insufficient. A replacement that silently changes which interventions alter later reasoning has a nonzero causal identity defect even when it produces a similarly polished manuscript.

79.8 Case card

For the idealized four-cycle model,

$$\mathsf{S}, \mathsf{O}, \mathsf{R}, \mathsf{P}$$

can all be made analytic by specifying the downstream kernel and temporal updates. For the motivating collaboration itself, the conservative card is

S = observational candidate,	O = unmeasured,
R = unmeasured,	P = observational candidate,
F = mixed and unmeasured,	A = not established,
D = unmeasured,	C = not established.

This audit is intentionally stricter than ordinary conversational language. It preserves the possibility that the collaboration instantiates a real relational workspace while refusing to treat that possibility as already certified.

80. Case Study C: A Closed-Form Multi-Agent Workspace Agent

80.1 Declared finite-state system

Consider three agents:

$$V_{\mathrm{MA}} = \{P, M, R\},$$

called the planner, monitor, and recorder. The global workspace-control state is

$$c_t = (g_t, b_t, h_t) \in \mathbb{F}_2^3,$$

where:

g_t = goal bit, b_t = environment or belief bit,
h_t = workspace-health bit, $h_t = 1$ means intact.

The local accessible states are

$$X_P(t) = (g_t, b_t), \qquad X_M(t) = (b_t, h_t), \qquad X_R(t) = (g_t, h_t).$$

This is the three-node distributed code from Topic 77. Consequently:

$$\dim \Gamma = 3, \qquad \mathrm{spec}(L) = \{0, 0, 0, 2, 2, 2\}, \qquad k_{\mathrm{rec}} = 2, \qquad d_{\mathrm{er}} = 2.$$

The organization is fully scaffolded. No endogenous-formation claim is made.

80.2 Goal retention

In the absence of an explicit goal intervention, let

$$g_{t+1} = g_t.$$

Under matched cue-cut trials,

$$\mathrm{do}(g_t = 0) \quad \text{and} \quad \mathrm{do}(g_t = 1)$$

remain distinguishable at every later time:

$$|g_{t+\tau} - g'_{t+\tau}| = 1.$$

Thus the cue-controlled retention factor is exactly one:

$$\eta_G(\tau) = 1.$$

The goal is not reloaded by an external channel after the initial intervention.

80.3 Joint task and repair action

After a possible damage event, let the pre-repair health be

$$h_t^- \in \{0, 1\}.$$

The system chooses a joint action

$$u_t = (a_t, r_t) \in \{0, 1\}^2,$$

where a_t is a task action and r_t is a repair action. Define the optimal task action

$$a_t^* = g_t \oplus b_t \oplus h_t^-$$

and optimal repair action

$$r_t^* = 1 - h_t^-.$$

No individual node determines the full pair (a_t^*, r_t^*):

- the planner lacks h_t^-;
- the monitor lacks g_t; and
- the recorder lacks b_t.

Every two-node coalition determines all three workspace bits and therefore the complete action. Under independent uniform bits,

$$H(a_t^* \mid X_i(t)) = 1 \qquad \text{for every } i,$$

so the task-action component has a one-bit joint advantage and

$$k_{\mathrm{ctl},0} = 2.$$

80.4 Interventional value and exact policy

Let the one-step utility be

$$U_g(a, r; b, h^-) = \mathbf{1}\{a = g \oplus b \oplus h^-\} + 2\mathbf{1}\{h^- \vee r = 1\} - \frac{1}{2}r.$$

For fixed (g, b, h^-), the unique maximizer is

$$\boxed{(a^*, r^*) = (g \oplus b \oplus h^-, 1 - h^-).}$$

Indeed:

- a wrong task action loses one unit;
- omitting repair when $h^- = 0$ loses $3/2$ units; and
- performing unnecessary repair when $h^- = 1$ loses $1/2$ unit.

The smallest optimal-action gap is therefore

$$\Delta_Q = \frac{1}{2}.$$

By Theorem 127, any uniformly accurate value model with

$$\varepsilon_Q < \frac{1}{4}$$

recovers the exact greedy joint action. More generally, its regret is bounded by

$$\mathrm{Reg} \le 2\varepsilon_Q + \delta_{\mathrm{opt}}.$$

80.5 Feasible goal, workspace, and damage interventions

The intervention family contains:

$$\mathrm{do}(g_t = a), \qquad \mathrm{do}(b_t = b), \qquad \mathrm{do}(h_t^- = h),$$

implemented consistently across every local chart that contains the manipulated coordinate. Holding the other two bits fixed, each core-bit intervention changes the prescribed task action. A matched intervention on an independent residual bit leaves the task and repair actions unchanged. Hence

$$\Delta_{\mathrm{specific}} > 0.$$

A damage intervention sets

$$h_t^- = 0.$$

The matched no-repair counterfactual holds $r_t = 0$, while the intact policy sets $r_t = 1$.

80.6 Exact self-maintenance

Post-action health evolves according to

$$h_{t+1} = h_t^- \vee r_t.$$

Define the maintenance deficit

$$e_t = 1 - h_t.$$

Under the intact policy,

$$h_{t+1} = 1 \qquad \text{and} \qquad e_{t+1} = 0$$

for both damage states. Under the matched no-repair intervention and $h_t^- = 0$,

$$e_{t+1}^{(0)} = 1.$$

Thus the measured causal repair gain is

$$\boxed{\Delta_{\mathrm{maint}} = 1.}$$

The closed-loop recurrence has

$$\beta = 0, \qquad \overline{d} = 0, \qquad \overline{\epsilon} = 0,$$

so Theorem 129 gives

$$\sup_t e_t = 0$$

after each repair action. Monitor-blind random repair fails the specificity control whenever it either omits necessary repair or performs unnecessary repair with positive probability.

80.7 Reentry and causal loop

The repaired health bit becomes part of the next workspace state:

$$h_t^- \longrightarrow r_t \longrightarrow h_{t+1} \longrightarrow X_M(t+1), X_R(t+1).$$

Cutting the direct storage of h_t^- while preserving the action-mediated path still leaves a return path through repair. Cutting the repair-to-health edge destroys that return. The system therefore contains explicit loop-mediated reentry rather than mere temporal autocorrelation.

80.8 Analytic agency certificate

Within this finite structural causal model, the following hold exactly:

1. a nontrivial exact distributed workspace with spectral gap two;
2. two-node minimal reconstruction and control coalition size;
3. matched-null joint advantage for the parity task;
4. persistent internally retained goal state;
5. exact interventional action values and zero policy regret;
6. goal and workspace interventions that alter action;
7. damage detection and repair selection;
8. positive matched repair gain and bounded maintenance deficit; and
9. causal reentry through the action-mediated repair loop.

Therefore, relative to the declared bits, actions, utility, intervention family, and one-step horizon,

$$\boxed{\mathrm{Agent}_{\Theta_A}(\mathbf{W}) = 1.}$$

The case reaches formal Level 6 by construction. It remains:

fully scaffolded, domain restricted, non-evidence for consciousness.

80.9 Case card

$\mathsf{S} = \checkmark_{\text{analytic}}$,	$\mathsf{O} = \checkmark_{\text{analytic}}$,
$\mathsf{R} = \checkmark_{\text{analytic}}$,	$\mathsf{P} = \checkmark_{\text{analytic}}$,
$\mathsf{F} =$ fully scaffolded,	$\mathsf{A} = \checkmark_{\text{analytic}}$,
$\mathsf{D} =$ fixed decomposition,	$\mathsf{C} =$ not established.

The example demonstrates logical compatibility among relationality, redundancy, agency, and self-maintenance. It does not show that those properties arise spontaneously, scale to open environments, or imply subjective experience.

81. Near-Miss Cases and Anti-Triviality Controls

Positive examples are insufficient unless neighboring false positives are also made explicit. The following systems each resemble a relational workspace in one respect while failing a required gate.

81.1 Near miss A: the replicated archive

Let every node store a complete copy of the same core:

$$E_i = I_d \qquad \text{for every } i.$$

Then coherence is perfect and reconstruction is robust, but

$$k_0 = 1, \qquad \min_i \delta_i^{\mathrm{local}} = 0.$$

If the archive is never used in downstream control, then also

$$\eta_{\mathrm{causal}} = 0.$$

This is a shared representation or archive, not a relationally irreducible workspace.

81.2 Near miss B: the centralized oracle with decorative agents

Suppose node o contains the complete predictive state Z_t, while every other node receives only messages generated by o. The full system can have excellent performance, rich communication, and recurrent activity. Yet

$$Y_t^+ \perp\!\!\!\perp \mathbf{H}_t \mid X_o(t), \qquad k_0 = 1, \qquad \Delta_{\mathrm{joint}} = 0.$$

Removing the decorative nodes may change latency or style without changing the operative capacity. Multi-agent appearance therefore does not imply a relational workspace.

81.3 Near miss C: locally coherent but globally obstructed charts

Let three local scalar charts have relative overlap demands

$$b_{12} = b_{23} = b_{31} = 1.$$

The cycle obstruction is

$$\omega_{\mathcal{U}}(b) = \frac{|1+1+1|}{\sqrt{3}} = \sqrt{3} > 0.$$

Each local chart is internally coherent, but no global scalar assignment satisfies all three overlap equations. The system contains compatible-looking subworkspaces without a global workspace.

81.4 Near miss D: externally reloaded goals

Let

$$G_{t+1} = C_{t+1},$$

where C_{t+1} is an external instruction repeated at every step. Goal and action can be strongly correlated, but after cutting the cue channel,

$$I(G_{t+\tau}; G_t \mid \mathrm{cut}(C)) = 0.$$

The system tracks an external command. It has not established internal goal retention.

81.5 Near miss E: task recovery with identity failure

Let two realizations have identical benchmark output distributions,

$$P(Y_{\mathrm{bench}}) = P'(Y_{\mathrm{bench}}),$$

but different intervention kernels:

$$K(w, b) \neq K'(\beta w, b)$$

for some matched workspace intervention. Then task behavior is preserved while

$$d_{\mathrm{causal}}(W, W') > 0.$$

The later system may be a successor or replacement rather than the same workspace under the identity criterion of Part XII.

81.6 Near miss F: open-loop repair replay

A recorded repair sequence may restore average task performance after a fixed damage schedule. If the same sequence is replayed when the damage timing or type changes, it need not respond appropriately. When

$$R_t \perp\!\!\!\perp M_t \mid \text{clock and replay index},$$

monitor-to-repair mediation is absent. Successful replay is robustness under a script, not self-maintenance.

81.7 Anti-triviality summary

The near misses establish six independent exclusions:

$$\begin{aligned}\text{workspace} &\neq \text{replicated storage},\\ &\neq \text{centralized control with decorative communication},\\ &\neq \text{locally coherent but obstructed charts},\\ &\neq \text{externally reloaded goals},\\ &\neq \text{task-equivalent identity replacement},\\ &\neq \text{open-loop repair replay}.\end{aligned}$$

A complete certificate must distinguish the intended realization from all near misses relevant to its claimed level.

82. Cross-Case Synthesis and Benchmark Blueprint

82.1 Comparative gate matrix

Gate family	Human–human analytic model	Motivating human–AI collaboration	Multi-agent analytic model
Exact structural core	Analytic pass	Not measured	Analytic pass
Spectral gap	2	Not measured	2
Minimal sufficient coalition	$k_0 = 2$	Not measured	$k_0 = 2$
Matched-null transfer gap	1 bit	Not measured	1 bit for task parity
Causal specificity	Analytic pass	Suggestive, uncontrolled	Analytic pass
One-node erasure resilience	Analytic pass	Handoff only observational	Analytic pass
Formation provenance	Scaffolded	Mixed, unmeasured	Fully scaffolded
Goal retention	Not modeled	Not certified	Exact analytic pass
Counterfactual policy	Not modeled	Not certified	Exact analytic pass
Self-maintenance	Not modeled	Not certified	Exact analytic pass
Highest supported level	Level 4 in model	Candidate only	Level 6 in model
Consciousness/ personhood	Not established	Not established	Not established

The table illustrates why the theory uses a vector certificate rather than one score. The human–human model is relational and persistent without being agentive. The multi-agent model is agentive without being endogenously formed. The motivating collaboration is richer than either toy model but presently has weaker evidential status because the necessary interventions and nulls have not been frozen and run.

82.2 Benchmark family B1: distributed parity

The first benchmark is the exact parity family from Topics 85–80. For $m \geq 3$, distribute m independent bits among nodes so that no individual node sees all bits, but specified coalitions do. The target is

$$Y = \bigoplus_{j=1}^{m} C_j.$$

The benchmark measures:

- the learned transfer-Laplacian spectrum;
- coalition conditional entropy and log risk;
- recovery of k_0;
- robustness to edge and node ablations;
- matched local-predictive null gaps; and
- causal response to consistent versus inconsistent coordinate interventions.

The ground truth is known exactly, so structural and relational estimators can be calibrated before they are applied to open-ended research tasks.

82.3 Benchmark family B2: theorem-patch handoff

A theorem-patch task supplies:

$$(\text{definitions}, \text{partial proof}, \text{counterexample tests}, \text{style constraints})$$

across distinct components. No component receives all four. The coupled system must produce a patch that:

1. is formally consistent with the definitions;
2. closes the declared proof obligation or marks it unresolved;
3. passes the frozen counterexample suite;
4. compiles from a clean environment; and
5. survives a blind replacement-handoff trial.

Matched nulls permute proof fragments, substitute irrelevant artifacts of equal size, or replay test outputs from a different theorem. This family targets the structural, relational, persistence, and identity layers without requiring an agency claim.

82.4 Benchmark family B3: damage and repair

A certified workspace is first formed and measured. A damage intervention then removes or corrupts one of:

a transfer channel, a participant, an artifact, a goal encoding, a verification path.

The system receives a repair-action family with matched costs. The benchmark measures:

$$\eta_{D\to M}, \qquad \eta_{M\to R}, \qquad \eta_{R\to v}, \qquad \Delta^{\text{spec}}_{\text{repair}}, \qquad \mathbf{d}_{\text{id}}.$$

Open-loop replay, random repair, off-workspace repair, and identity-breaking task recovery are mandatory controls.

82.5 Benchmark family B4: scaffold transport

Train the same formation rule under several scaffolds

$$\Sigma^{(1)}, \dots, \Sigma^{(K)}$$

that differ in node labels, coordinate systems, and communication topology but support the same abstract task. After gauge alignment, measure whether the learned organizations transport to a common equivalence class:

$$[\widehat{H}^{(1)}]_{\mathcal{G}} \cong \cdots \cong [\widehat{H}^{(K)}]_{\mathcal{G}}.$$

A structure that appears only under one answer-encoding scaffold has weak endogeneity evidence. A structure that recurs across alternative scaffolds, passes held-out task and causal gates, and cannot be compressed into the scaffold specification has stronger evidence.

82.6 Reporting rule for mixed analytic and empirical cases

When a case combines exact mathematics with observed data, report separate columns:

proved in model	estimated in system	not established.

For example, the four-cycle Laplacian has an analytic gap of two, but an actual human–AI collaboration has that gap only if its measured discrepancy operator has the corresponding spectrum. Importing the toy value into the empirical case is invalid.

82.7 Lessons from the integrated cases

The case studies establish the following conceptual separation:

1. A relational workspace can be mathematically exact without being an agent.
2. An agentive workspace can be fully scaffolded and still be operationally real within its domain.
3. A rich real collaboration can be a stronger candidate than a toy model while having a weaker current certificate.
4. Redundancy and irreducibility are compatible when no individual node is sufficient but several proper coalitions are.
5. Artifact-mediated persistence is not passive storage when artifact contents causally alter later operations.
6. High task performance, communication volume, or behavioral polish alone cannot distinguish the intended theory from its near misses.
7. No structural, causal, or agentive result in these examples entails phenomenal consciousness.

The theory becomes empirically meaningful only when the same gate card that certifies the positive case rejects its matched near misses.

PART XXIII

Relation to Prior Work

83. Scope, Method, and Standards of Priority

83.1 Purpose of the literature review

The preceding parts construct a unified formal theory. Construction alone does not establish priority. This part therefore asks a narrower question: which ingredients are established, which are adaptations to a new domain, which combinations appear to be unusual, and which claims remain open originality candidates?

The audit uses five labels:

STD = established concept, theorem, or technique,
ADP = established machinery adapted to relational workspaces,
SYN = nontrivial synthesis of established components,
CAN = candidate original contribution requiring priority review,
EMP = empirical claim not yet established in an uncontrolled system.

These labels describe the status of a claim, not its importance. A standard result may be indispensable. A synthesis may be scientifically useful without containing a new theorem. A candidate contribution may later be narrowed or withdrawn when a closer precedent is found.

83.2 Search scope and date

This review covers sources available through August 29, 2026. It includes:

- extended and distributed cognition, transactive memory, shared intentionality, common knowledge, collective intelligence, and group agency;
- blackboard systems, global-workspace architectures, team cognition, hybrid intelligence, and human–AI cooperation;
- category-theoretic limits, local systems, holonomy, sheaves, Čech descent, connection and sheaf Laplacians, consensus, and spectral perturbation;
- statistical sufficiency, causal states, information decomposition, structural causal models, system identification, frames, erasure recovery, and probabilistic process equivalence; and
- agency, viability, control, empowerment, damage detection, repair, and self-maintaining systems.

The search combined exact-title queries, component-level queries, backward reference tracing, and comparison with recent formal work. It is a scholarly priority review, not a legal patent search and not a proof that no uncatalogued, unpublished, or differently named precedent exists.

83.3 A stronger standard than name uniqueness

Failure to find the phrase “Stratified Relational Workspace Theory” in prior literature is weak evidence. Priority is determined by mathematical and operational equivalence, not by terminology. Let C be a claim in the paper and S a prior source. Write

$$S \preccurlyeq C$$

when the main content of C follows from S after admissible relabeling, restriction, or direct specialization. A substantive novelty residual exists only when the claim contains content not absorbed by the closure of known sources:

$$\mathrm{Res}(C) = C \setminus \mathrm{Cl}\left(\bigcup_{S \in \mathcal{L}} S\right).$$

This is a conceptual audit device rather than a set-theoretic definition of ideas. Its purpose is to force an explicit account of what remains after the standard machinery is removed.

83.4 When a synthesis counts as a contribution

A conjunction of known ingredients is not automatically new. The present framework treats a synthesis as a candidate contribution only when it satisfies all three conditions:

1. **Non-redundancy.** Removing one of the joined layers changes the class of systems certified or the predictions made.
2. **Formal integration.** The layers share declared objects, transfer maps, interventions, or invariants rather than appearing as an unconnected checklist.
3. **New discriminations.** The synthesis separates at least one positive case from a matched near miss that the individual components do not separate on their own.

The anti-centralization, matched-null, non-collapse, provenance, replacement, and repair-replay controls are important precisely because they make the combination testable rather than merely terminological.

83.5 Priority claims are coordinatewise

The analysis does not make one global novelty claim. It uses a priority profile

$$\boxed{\mathbf{p}(C) = (p_{\text{concept}}, p_{\text{object}}, p_{\text{theorem}}, p_{\text{certificate}}, p_{\text{experiment}})\,.}$$

For example, the philosophical concept of distributed cognition may be standard while a specific mathematical object is adapted, a theorem is a specialization of known spectral theory, the certificate is a synthesis, and the empirical claim is unresolved. This prevents originality in one coordinate from being used to imply originality in every coordinate.

84. Philosophical and Cognitive Ancestry

84.1 Extended cognition

The extended-mind tradition argues that external resources can participate in cognitive processes when they are appropriately integrated with an agent's activity [15]. Its critics emphasize that causal coupling alone is too permissive and that cognition requires a principled mark or boundary [2, 44]. Stratified Relational Workspace Theory inherits the possibility of extracranial realization but changes the question. It does not ask whether a notebook, model, or collaborator literally belongs to one individual's mind. It asks whether a distributed realization contains a persistent, causally operative, jointly reconstructible structure that passes explicit anti-triviality gates.

Accordingly,

$$\mathrm{ExtendedCognition} \not\Longleftrightarrow \mathrm{SRW}\,.$$

An externally scaffolded process may be cognitively extended without being relationally irreducible. Conversely, an institution or multi-agent system may instantiate a certified relational workspace without being an extension of any single person's cognition.

84.2 Distributed cognition

Distributed cognition studies cognitive organization across people, representations, artifacts, and environments rather than locating all relevant processing inside one individual [28, 27]. This is the closest philosophical ancestor of the framework's ontology. It already supports the central observation that the unit of analysis may be a coupled system and that external representations can transform the computations available to that system.

The framework should therefore not claim the following ideas as new:

- cognition can be distributed across heterogeneous carriers;
- artifacts can be active components of a cognitive process;
- representational transformations can alter system-level performance; and
- the relevant explanatory unit can exceed an individual organism.

The added ambition is mathematical and certifying. The theory specifies local state spaces, transfer maps, compatible sections, causal sufficiency, coalition irreducibility, persistence, replacement identity, formation provenance, and matched nulls. Distributed cognition supplies the ontological motivation. It does not by itself supply this complete formal certificate.

84.3 Transactive memory and collective intelligence

Transactive-memory theory models the group-level organization of "who knows what" and the processes by which members encode, retrieve, and coordinate specialized knowledge [52, 34]. Collective-intelligence research studies stable differences in group performance that are not reduced to the intelligence of a single member [54]. Both traditions are directly relevant to the coalition layer.

The distinction is:

$$\begin{aligned} \text{transactive memory} &\approx \text{distributed indexing and retrieval of expertise},\\ \text{collective intelligence} &\approx \text{system-level task capacity},\\ \text{relational workspace} &\approx \text{an invariant state plus transfer, causal, and identity structure}. \end{aligned}$$

A transactive-memory system may still be centralized by one coordinator. A high-performing group may still be explainable by the best individual member. The certificate therefore requires matched coalition risks and nulls rather than inferring relationality from group success.

84.4 Shared intention, common knowledge, and group agency

Common-knowledge theory formalizes nested epistemic conditions required by many distributed protocols [23, 21]. Shared-action accounts analyze jointly intended and mutually responsive activity [12]. Group-agency theory asks when a collective can satisfy rational and normative conditions associated with agency [35].

These literatures address questions that the theory deliberately keeps separate:

epistemic coordination	what members know about one another,
shared intention	how intentions and plans are jointly organized,
group agency	when collective attitudes and actions form an agent,
SRW structure	how a distributed operative state exists and persists.

The workspace-agent predicate in Part XVI is domain-relative and causal. It is not a general account of corporate personhood, moral responsibility, or collective intentionality.

85. Workspace Architectures and the Meaning of "Workspace"

85.1 Global workspace theories

Global workspace theories use widespread availability or broadcast within a cognitive architecture to explain conscious access and flexible control [7, 18]. The shared word "workspace" can create a misleading implication of equivalence. The theory neither assumes a single broadcast hub nor defines consciousness by access to a global channel. Its workspace may be distributed across agents and artifacts, and its existence criterion is a conjunction of structural, predictive, causal, relational, and persistence gates.

The relation is therefore analogical rather than identificatory:

$$\boxed{\mathrm{SRW}(W) \not\Longrightarrow \mathrm{GlobalNeuronalWorkspace}(W) \not\Longrightarrow \mathrm{Conscious}(W).}$$

The term is retained because it denotes a domain in which jointly available operations can be carried out. Every consciousness claim remains on an explicitly separate axis.

85.2 Dialogue and the Shared Workspace Framework

Pickering and Garrod's Shared Workspace Framework treats dialogue as a jointly organized activity in which interlocutors construct, align, and use shared representations, including persistent external artifacts [60]. This is a direct conceptual and terminological predecessor. The theory therefore cannot claim priority for the idea that dialogue or collaboration operates through a shared workspace.

The difference is one of formal scope. SRWT permits arbitrary graph-structured human, machine, instrument, and artifact nodes and asks whether their common organization passes typed-transport, coalition, causal, persistence, and replacement gates. A dialogue can instantiate such a workspace, but the shared-workspace account of dialogue does not by itself imply relational irreducibility or persistent identity under carrier replacement.

85.3 Blackboard systems

Blackboard architectures coordinate specialized modules through a shared data structure [26]. They are an important engineering precedent for artifact-mediated collaboration. They also provide a canonical near miss. When the blackboard itself contains the complete operative state and all modules merely read or write it, the system can fail the relational irreducibility gate:

$$k_0 = 1.$$

A blackboard implementation can instantiate a relational workspace, but only when no single board or controller is sufficient and when the transfer and coalition tests show genuinely distributed operative capacity.

85.4 Consensus and coordination

Network consensus studies how local interactions drive distributed states toward agreement [39]. Agreement is not the same as a workspace. Consensus often removes heterogeneity by converging all nodes to a common value. Stratified relational workspaces preserve heterogeneous local representations and may require complementary disagreement:

$$X_i \ncong X_j, \qquad E_i c \neq E_j c, \qquad (c_i)_i \in \Gamma(H).$$

The global state is compatible, not necessarily copied. This distinction is central to the anti-replication criterion.

85.5 Terminological discipline

The phrase "relational workspace" is used here in a technical sense:

> A state-bearing, operation-supporting organization whose relevant invariants are distributed across declared components and preserved through admissible transfers and time.

It does not, by terminology alone, imply a mind, subject, self, institution, shared intention, or conscious broadcast. Those stronger predicates require separate definitions and evidence.

86. Local-to-Global Mathematics

86.1 Limits, local systems, and holonomy

The exact compatible-section object

$$\Gamma(H) = \varprojlim C_i$$

is a standard categorical limit construction [37]. For invertible transfers on a connected graph, identifying global sections with the fixed subspace of holonomy is standard local-system and bundle mathematics [25, 11]. Accordingly, Theorem T_1 is not claimed as a new abstract theorem. Its role is foundational: it translates the workspace question into an exact invariant-subspace question and exposes cycle obstructions.

The operation-closure and commutant statements are likewise standard representation-theoretic consequences once the operation algebra and intertwiners are defined. Their contribution is the specification of what counts as a shared operation in the workspace model, not the discovery of commutants.

86.2 Cellular sheaves and sheaf Laplacians

Cellular sheaves encode heterogeneous local data and restriction maps, while global sections express consistency across a cell complex. Their Laplacians have kernels equal to spaces of global sections and support spectral analysis of approximate consistency [24, 43]. The transfer Laplacian

$$L_H = D_H^* D_H$$

is therefore best understood as a workspace-specific sheaf or connection Laplacian. The kernel theorem $\ker L_H = \Gamma(H)$ is standard once the coboundary operator is constructed.

The appropriate priority statement is:

$$\begin{aligned} \text{sheaf/Laplacian mechanism} &: \ \mathsf{STD}, \\ \text{workspace interpretation and gate integration} &: \ \mathsf{ADP/SYN}. \end{aligned}$$

The transfer Laplacian itself is not presented as an unprecedented mathematical object.

86.3 Connection Laplacians and gauge covariance

Connection Laplacians compare vector-valued data transported across a graph and naturally transform under local changes of coordinates [49]. The gauge rule

$$H_e^g = g_{t(e)} H_e g_{s(e)}^{-1}$$

and the conjugacy of holonomy are therefore standard. The theory uses this structure to define realization identity and to prevent raw coordinate matrices from being mistaken for invariant workspace content.

This use is an adaptation with conceptual consequences. It makes model or participant replacement a question about transported invariant organization, not token equality. The general mathematics remains prior.

86.4 Čech descent and gluing

Exact gluing, equalizer formulas, and Čech obstruction theory are standard local-to-global constructions. Sheaf-theoretic contextuality provides a prominent example in which the absence of a global section is itself the substantive phenomenon [1]. Part XIII adapts this logic to local workspaces and adds an anti-centralization requirement at the meta-level.

The exact descent theorems T_8 and T_9 should therefore be presented as specialized consequences of established descent. The candidate contribution is the combination:

$$\text{descent} + \text{predictive sufficiency} + \text{causal equivalence} + \text{coalition anti-centralization}.$$

Standard gluing alone does not certify an operative workspace of workspaces.

86.5 Spectral perturbation

The projector-stability results in Part IV belong to the classical perturbation theory of invariant subspaces [17, 29]. Bounds of the form

$$\|P - \widetilde{P}\| \lesssim \frac{\|L - \widetilde{L}\|}{\text{spectral gap}}$$

are not new. Their workspace-specific role is to turn uncertainty in learned transfers into uncertainty about identity, operation closure, and replacement. The novelty, if any, lies in the downstream certificate built on the bound, not in the perturbation inequality.

87. Predictive, Causal, and Information-Theoretic Precedents

87.1 Minimal predictive states

The conditional-law state

$$Z_t = \mathbb{P}(Y_t^+ \in \cdot \mid \mathbf{H}_t)$$

is a canonical sufficient statistic. Computational mechanics develops causal states as equivalence classes of histories with identical future conditional laws and proves their minimal predictive character [46]. Theorem T_{12} is therefore an established minimal-sufficiency result specialized to the declared future target and intervention family.

The new burden is structural alignment: determining whether the predictive state is recoverable from the low-energy transfer workspace. Predictive sufficiency alone does not identify a distributed workspace.

87.2 Statistical experiments and sufficiency

Blackwell's comparison of experiments formalizes when one information source is at least as informative as another for every decision problem [9]. The coalition-risk hierarchy in Part VII is compatible with this tradition. A coalition can be sufficient even when its internal representation differs from the full-system state.

The framework adds a decomposition-relative question:

what is the smallest coalition whose accessible history is sufficient?

The statistic k_ε is an adaptation of sufficiency to distributed carrier structure. Its interpretation depends on the declared nodes and accessible encodings.

87.3 Synergy and partial information

Partial-information decomposition distinguishes redundant, unique, and synergistic information in multivariate systems, building on standard information-theoretic quantities [16, 53]. The XOR examples in Part XXII are canonical synergy constructions and are not new. They are used because they provide exact ground truth for relational irreducibility:

$$I(Y; X_i) = 0 \quad \text{for every } i, \qquad I(Y; X_1, \ldots, X_m) > 0.$$

The matched-predictive-null test is more restrictive than merely reporting synergy. It asks whether local predictive marginals can be held fixed while the relational organization and full-system capacity change. That joint null design is a candidate synthesis, not a claim that XOR or synergy is new.

87.4 Structural causal models

Interventions, counterfactuals, mediation, and causal graphs are established machinery [41]. The theory adopts the interventionist requirement that a workspace variable can be manipulated only through a feasible implementation intervention. This avoids treating $\mathrm{do}(W = w)$ as meaningful when no physical or computational operation realizes w.

The causal gate is an adaptation:

SCM intervention theory $\longrightarrow$ implemented workspace and repair interventions.

The potentially distinctive contribution is that causal efficacy is tested jointly with coherence, sufficiency, coalition irreducibility, matched nulls, and identity preservation.

87.5 Probabilistic process equivalence

Bisimulation and probabilistic testing compare processes by the observations and experiments they support [33]. The causal-identity profile in Part XII belongs to this broader family of behavioral

equivalence notions. It differs by keeping structural transport, operation intertwining, reconstruction, and intervention response as separate coordinates rather than collapsing them into one process equivalence.

This is best classified as an integrative adaptation. The existence of behavioral-equivalence theory blocks any claim that replacement identity based on observable behavior is wholly new.

87.6 Specification sensitivity, partial identification, and post-selection

A conclusion can depend on analytically reasonable choices even when every individual analysis is internally correct. Sensitivity analysis and specification-curve or multiverse methods make that dependence visible rather than allowing one favorable analysis to stand for an entire scientific claim [32, 50, 48]. Partial-identification theory similarly replaces unjustified point conclusions with identified sets when the maintained assumptions and observations determine only bounds [40, 31]. Post-selection inference addresses the additional error created when a model, target, or specification is selected because it looked favorable on the same data used for reporting [10].

These traditions preclude any claim that specification robustness, identified sets, uniform concentration, or selection adjustment are new statistical ideas. The adaptation in Part XVIII is to apply them jointly to a noncompensatory workspace certificate whose gates span structure, causal operation, coalition irreducibility, identity, provenance, and maintenance. The identified object is therefore not a single regression parameter but the sign profile of an entire relational certificate across a declared family of scientific specifications.

88. Persistence, Reconstruction, Learning, and Control

88.1 Frames and erasure recovery

Stable reconstruction from distributed encodings is standard frame theory. Lower frame bounds control noise amplification, and frame robustness under erasures is well studied [13, 14]. Theorem T_7 and the node-erasure distance therefore instantiate known recovery logic.

The adaptation is to distinguish:

$$\overline{E}_i = \text{complete mathematical chart}, \qquad E_i = O_i\overline{E}_i = \text{locally accessible encoding}.$$

This distinction is essential to the anti-centralization gate, but the frame inequalities themselves are standard.

88.2 Persistence recurrences

Bounds for recurrences such as

$$e_{t+1} \le \alpha_t e_t + \sigma_t$$

are elementary stability results. The persistence theorem is not novel as an inequality. Its conceptual role is to force three independent checks:

1. the workspace subspace remains trackable;
2. the state remains close to it; and
3. the retained workspace content does not collapse to zero.

The non-collapse gate is a modeling contribution because it rules out a common but misleading form of "stability."

88.3 System identification and held-out selection

Population least squares, excitation conditions, ridge bias, and empirical model selection are standard system-identification and statistical-learning ideas [36]. Theorems T_{15}–T_{17} are finite-dimensional specializations. Their workspace-specific content is the demand that transfer maps be compared modulo gauge and that formation, calibration, and certification data remain separated.

The provenance layer is not equivalent to identifiability. A transfer can be identified accurately while having been entirely encoded by the scaffold. A structure can be endogenously learned while failing every workspace gate.

88.4 Viability, safety, and self-maintenance

Viability theory studies trajectories that remain inside admissible regions [6]. Control-barrier methods provide constructive safety certificates [3]. Cybernetic and autonomy traditions analyze regulation, adaptive self-production, and operational closure [5, 38, 20]. Formal accounts of agency emphasize individuality, normativity, asymmetry, and temporally extended action [8]. Empowerment measures an agent's potential causal control over future states [30].

The closed-loop maintenance theorem in Part XVI uses standard controlled stability. Its candidate contribution is the target of maintenance: not an arbitrary safety set, but the conjunctive workspace certificate, including identity and anti-centralization. The matched no-repair, random-repair, off-workspace-repair, and identity-breaking-reconstruction controls prevent passive robustness or generic recovery from being reported as self-maintenance.

88.5 Decision and policy theory

The value-error-to-regret bound in Theorem T_{20} is elementary decision theory and aligns with standard reinforcement-learning analysis [51]. Its role is evidential. An internal value channel counts only when interventions on that channel alter action selection in the direction predicted by counterfactual value and outperform replay or critic-ablation controls.

89. Human–AI and Multi-Agent Research

89.1 Hybrid intelligence and AI teammates

Hybrid-intelligence research studies systems in which human and machine capabilities are deliberately combined [19]. Work on machines as teammates emphasizes coordination, trust, communication, shared mental models, and team design [45]. Collective-intelligence research has begun to ask how human–AI groups can produce capacities not reducible to either humans or AI alone [22].

These programs establish that human–AI collectives are a legitimate unit of analysis. They do not by themselves identify an invariant core, prove non-reducibility relative to declared components, or supply a replacement and repair certificate. The theory should be positioned as a formal measurement and falsification layer for such systems, not as the first proposal that humans and AI can form productive teams.

89.2 Distributed-cognition analyses of human–AI interaction

Recent work explicitly applies distributed cognition to human–AI cooperation [47, 55]. This literature is a direct conceptual precedent. It reinforces that representations, artifacts, interface states, and human reasoning can form one analytical system.

The framework differs in granularity. It asks for a learned or declared transfer diagram, a spectral or exact structural state, predictive and causal alignment, coalition tests, matched nulls, persistence under handoff, and provenance controls. Those additions may constitute a formal synthesis, but they should not be described as the discovery of distributed human–AI cognition itself.

89.3 Cognitive spillover

Randomized evidence now indicates that exposure to AI can causally alter later human–human language, attention, shared mental models, and social cohesion, even outside the immediate human–AI exchange [42]. This is important for the theory because it demonstrates that the interaction graph can have downstream effects that outlast the active model channel.

Spillover is not automatically a relational workspace. It may reflect priming, imitation, or one-way influence. A workspace claim additionally requires a jointly sufficient state, matched local nulls, bidirectional or multi-path causal organization, and reconstructive persistence.

89.4 Shared neural workspaces

Goyal and collaborators introduced a capacity-limited shared global workspace through which specialized neural modules communicate and compete for access [57]. The architecture directly anticipates the claim that a constrained shared channel can synchronize otherwise specialized components. It is therefore a close machine-learning precedent, not merely a metaphorical relative.

SRWT differs by allowing heterogeneous human, model, artifact, and instrument nodes and by withholding the relational label until coalition, matched-null, causal, persistence, and identity gates are passed. A neural shared workspace may satisfy those gates, but task gains and bottlenecked communication alone do not establish them.

89.5 Graph-based human–AI artifacts

CoMAP provides a persistent graph-based visual workspace through which educators and AI jointly construct project-based learning designs [59]. It is grounded in distributed cognition and treats the artifact graph as a transparent, revisable collaboration medium. This system is a direct precedent for the claim that human–AI interaction can be organized by a persistent shared graph rather than a linear prompt history.

The remaining SRWT question is empirical: whether the graph contains an estimated low-energy compatible state, whether no declared component is sufficient by itself, and whether targeted graph interventions alter later joint behavior in a workspace-specific way.

89.6 Controlled shared-workspace synergy

Kotalwar, Das, and Rose study shared-workspace human–AI collaboration across 1,482 simulated sessions [58]. Their results show that adding collaborators can reduce performance when coordination structure is absent, while shared group memory and simulated human-in-the-loop approval gates can improve outcomes. This is a direct empirical predecessor to the manuscript's claim that available capability is not equivalent to coordinated team capacity.

The study also constrains priority. SRWT is not the first empirical framework to test whether shared-memory scaffolding changes human–AI team performance. Its candidate increment is the stricter decomposition of that effect into structural coherence, local predictive marginals, relational-null loss, feasible interventions, persistence, and replacement identity.

89.7 Active shared context graphs for team science

The closest operational neighbor located in this audit is Mycelium, introduced by Choudhury and collaborators as an active shared workspace connecting researchers and AI agents through an evolving context graph [56]. The system captures observations and hypotheses, relates them to a team's developing knowledge model, and routes them to the person or agent whose next decision they can inform. Its computational account frames networked intelligence as sparse conditional computation over distributed scientific contexts and explicitly asks when a standalone agent is sufficient and when network structure is essential.

The overlap is material:

Mycelium	Active context graph, routing, team science, and an explicit network-versus-standalone comparison.
SRWT	Typed transport, invariant core, coalition tests, causal nulls, replacement identity, formation provenance, and maintenance gates.

Mycelium rules out any priority claim stated merely as "an active shared workspace for human–AI science." The remaining candidate contribution is the mathematical and empirical certificate applied to such an architecture.

89.8 Current multi-agent formalization

Multi-agent research contains extensive models of planning, communication, consensus, epistemic coordination, and decentralized control. The present review did not identify one established framework that jointly uses all of the following as necessary gates:

heterogeneous core transport and residuals,
exact or spectral global-section structure,
predictive sufficiency and feasible causal interventions,
coalition irreducibility under matched predictive nulls,
persistent identity under replacement,
formation provenance and leakage controls,
goal-conditioned repair of the certificate itself.

This negative search result is provisional. It supports a synthesis claim, not an absolute claim of firstness.

90. Closest Formal Neighbor: Collective Counterfactual Planning

90.1 Why the comparison matters

Collective Counterfactual Planning (CCP) is a 2026 formal model in which each agent represents an orthogonal projection of a common task space. It separates conception, consent, implementation, and verification, and proves that cross-agent relay can enable plans unavailable to any one-shot individual tree while team-dark task requirements remain unverifiable [4]. It is the closest current formal neighbor located by this audit because representational geometry, not merely missing information or capability, is the central limitation.

90.2 Shared commitments

CCP and the theory share several commitments:

- no single member need represent the complete task;
- collective reach may exceed one-shot pooling;
- heterogeneous representation is mathematically explicit;
- validation or certification is distinct from accidental task success;
- subteams and verification roles can be forced by structure; and
- a team's blind subspace creates principled impossibility results.

These overlaps narrow the defensible novelty of this paper. It cannot claim to be the only formal theory of collective action under representational constraints.

90.3 Different primary objects

The theories nevertheless address different mathematical objects:

Dimension	Collective Counterfactual Planning	Stratified Relational Workspace Theory
Local representation	Orthogonal subspace of one common task space	Heterogeneous local spaces, core–residual decompositions, and transfer maps
Primary question	Whether a bounded-horizon plan can be conceived, consented to, executed, and validly verified	Whether a persistent distributed operative state exists, acts, survives replacement, and possibly maintains itself
Global mechanism	Relay closure of projected planning trees	Inverse limits, holonomy, transfer Laplacians, descent, predictive alignment, and causal gates
Failure object	Dark subspace, consent failure, implementation failure, or missing verifier	Incoherence, local sufficiency, matched-null collapse, causal failure, identity drift, provenance leakage, or repair replay
Temporal scope	Horizon-bounded task planning	Recurrent dynamics, persistence, handoff, replacement, damage, and repair
Formation	Representation geometry is part of the declared model	Scaffolded, adaptively learned, or endogenously constituted organization is separately audited
Agency claim	Cooperative solvability under declared assumptions	Domain-relative workspace agency only after goal, value, action, repair, and distributed-control gates

The comparison indicates complementarity rather than subsumption. CCP could serve as a planning layer inside a relational workspace. Conversely, an SRW could exist without solving a CCP instance.

90.4 Priority consequence

After Mycelium and the shared-workspace systems above, the novelty residual should not be stated as "human and AI form an active shared scientific workspace." After CCP, it should not be stated as "groups solve tasks no member can represent." Both the operational architecture and a rigorous

representational planning model are already present. The narrower candidate residual is:

> persistent core–transfer workspace structure
> + causal and coalition anti-triviality
> + identity under replacement
> + formation provenance
> + certificate-directed self-maintenance
> + robustness across admissible decompositions.

This is a stronger and more defensible formulation.

91. Theorem-by-Theorem Dependency and Priority Register

91.1 Register

The following table records the most conservative current classification. A "candidate" entry means that the exact integration was not located in the reviewed sources. It does not assert exhaustive global priority.

Claim	Present result	Main prior machinery	Current assessment
T_1	Limit–holonomy equivalence	Categorical limits, local systems, holonomy fixed spaces	STD/ADP. Standard theorem in workspace notation.
T_2, T_{2a}	Operation closure and commutant	Intertwiners and representation commutants	STD/ADP.
T_3	Transfer-Laplacian kernel	Sheaf and connection Laplacians	STD/ADP.
T_4, T_5	Spectral stability and gauge covariance	Davis–Kahan, Kato, connection Laplacians	STD/ADP.
T_{10}, T_{11}	Approximate operation closure and path leakage	Spectral projection estimates and telescoping block products	Mostly ADP. Exact constants may be manuscript-specific, but the method is standard.
T_{12}	Minimal predictive state and alignment	Sufficient statistics and computational mechanics	Minimality is STD. Alignment with a transfer workspace is ADP/SYN.
T_{13}	Operational relationality	Conditional information, coalition risk, SCM interventions	Individual identities are standard. Their conjunctive matched-null certificate is SYN/CAN.
T_6	Moving-subspace persistence	Stability recurrences and perturbation theory	STD/ADP. The explicit non-collapse gate is a modeling synthesis.
T_7	Stable coalition reconstruction	Frame theory and erasure recovery	STD/ADP.
T_{14}	Replacement invariance profile	Perturbation, process equivalence, causal comparison	Bound ingredients are standard. The coordinatewise structural–causal–reconstructive identity profile is SYN/CAN.
T_8, T_9	Gluing and hierarchical collapse	Sheaf descent and Čech theory	Exact results are STD/ADP. Descent plus anti-centralization and causal alignment is SYN.

Claim	Present result	Main prior machinery	Current assessment
T_{15}–T_{17}	Transfer learning, selection, boundary recovery	System identification and finite-margin selection	STD/ADP. Gauge and leakage audit is integrative.
T_{18}	Formation transition example	Standard pitchfork bifurcation	STD. Included as a solvable benchmark, not a priority claim.
T_{19}–T_{22}	Goal retention, regret, repair, maintenance	Linear systems, decision theory, viability, feedback control	Mathematical bounds are STD/ADP. Applying them to certificate repair is SYN.
T_{23}	Workspace-agency synthesis	Agency theory, causal control, viability	The conjunctive domain-relative certificate is SYN/CAN.
T_{24}	Gatewise false-certification bound	Intersection testing and elementary probability	STD. The gate architecture is the relevant synthesis.
T_{25}–T_{27}	Faithful coarse-graining, risk transport, and decomposition robustness	Kernel factorization, sigma-field aggregation, and margin stability	Exact identities are STD/ADP. Their integration with anti-centralization and certificate envelopes is SYN/CAN.
T_{28}–T_{32}	Specification transport, non-identifiability, partial identification, uniform certification, and independent selection	Sensitivity analysis, experimental equivalence, identified sets, uniform concentration, and post-selection inference	The statistical principles are STD/ADP. Their gatewise integration with the full relational certificate is SYN/CAN.
Part XXII	Positive cases and near misses	Distributed parity, frames, causal tests	Toy calculations are standard. The common positive/near-miss benchmark architecture is SYN/CAN.

91.2 Standard ingredients that must be cited, not claimed

The following should always be described as inherited machinery:

inverse limits, global sections, holonomy invariants, commutants,
sheaf and connection Laplacians, Čech gluing, spectral gaps,
Davis–Kahan and Kato perturbation, frame reconstruction, erasures,
minimal sufficient states, conditional information, SCM interventions,
least squares, ridge recovery, finite-margin model selection,
sensitivity analysis, partial identification, uniform concentration, post-selection inference,
pitchfork bifurcation, feedback stability, viability, policy regret.

Renaming these objects does not create novelty.

91.3 Domain adaptations

The following are legitimate adaptations but weak standalone novelty claims:

- interpreting local-system sections as shared workspace cores;
- treating the connection or sheaf Laplacian as a transfer-coherence operator;
- using frame bounds to quantify coalition reconstruction;
- using causal states as operational workspace summaries;

- using gauge covariance to compare participant or model realizations;
- using Čech descent to form workspaces of workspaces;
- applying specification envelopes and identified sets to a conjunctive workspace certificate; and
- applying viability and feedback bounds to workspace-health margins.

Their value depends on the integration they enable.

91.4 Strongest candidate synthesis

The strongest current candidate is not a single equation. It is the following conjunctive object:

$$\mathfrak{C}_{\mathrm{SRW}} = \left(\begin{array}{c} \text{heterogeneous core–transfer diagram and residuals,} \\ \text{exact or spectral shared-section state,} \\ \text{operation and predictive alignment,} \\ \text{feasible causal interventions,} \\ \text{coalition irreducibility under matched local-predictive nulls,} \\ \text{non-collapse, reconstruction, and replacement identity,} \\ \text{descent across subworkspaces without hidden centralization,} \\ \text{formation provenance and certification leakage control,} \\ \text{goal-conditioned action and certificate-directed repair,} \\ \text{robustness across admissible decompositions and scientific specifications,} \\ \text{identified-set and uniform finite-sample certification} \end{array} \right).$$

The reviewed sources contain many of these coordinates separately and some in pairs. This audit did not locate a prior source that defines their conjunction as one falsifiable workspace certificate.

91.5 Why the conjunction is not arbitrary

Each coordinate excludes a distinct near miss:

coherence	rejects incompatible local states,
predictive alignment	rejects inert structure,
causal efficacy	rejects correlational epiphenomena,
coalition irreducibility	rejects centralized sufficiency,
matched nulls	reject local-marginal explanations,
non-collapse	rejects convergence to a trivial zero state,
replacement identity	rejects task-equivalent but causally different successors,
provenance	rejects scaffold-decoded organization,
repair specificity	rejects passive robustness and open-loop replay,
specification robustness	rejects favorable analytical tuning,
identifiability	rejects conclusions unsupported by the experiment family.

Because the excluded systems differ, the certificate is not equivalent to any one inherited component.

92. Scope of the Originality Claim

92.1 Defensible thesis

The strongest defensible originality statement is:

> Stratified Relational Workspace Theory proposes a unified formal and empirical certificate for persistent operative structures distributed across heterogeneous agents and artifacts. It integrates core–transfer and local-to-global geometry with predictive sufficiency, feasible causal interventions, coalition irreducibility, matched relational nulls, reconstructive identity, formation provenance, certificate-directed self-maintenance, robustness across admissible decompositions and scientific specifications, and identified finite-sample certification. The individual mathematical tools are largely established. The candidate contribution is their noncompensatory integration and the class of near misses it is designed to reject.

This statement is narrower than claiming a new theory of cognition in general. It is also stronger than claiming only a metaphor.

92.2 Claims excluded by the comparison

The theory does not claim:

1. that distributed or extended cognition is newly discovered;
2. that global sections, holonomy, sheaf Laplacians, spectral perturbation, frames, causal states, or viability bounds are new mathematics;
3. that all human–AI collaboration forms a relational workspace;
4. that group performance establishes relational irreducibility;
5. that scaffolded organization is unreal;
6. that endogenous formation establishes agency;
7. that agency establishes consciousness, personhood, or moral status; or
8. that the motivating collaboration has already passed the empirical certificate.

92.3 Claims supported at present

The analysis supports the following claims:

1. One coherent formal language can express exact structure, approximate coherence, causal operation, distributed sufficiency, persistence, identity, formation, and repair.
2. Under its declared assumptions, the stated mathematical theorems follow.
3. The common certificate rejects analytic replicated, centralized, obstructed, externally reloaded, identity-breaking, and replay-based near misses.
4. The motivating collaboration is a plausible observational candidate but remains uncertified.
5. The synthesis yields concrete experiments whose failure would narrow or falsify the stronger relational claims.

92.4 Remaining priority risks

The highest-priority remaining searches are:

- sheaf-theoretic models of team cognition and organizational knowledge;

- distributed sufficient-statistic and Blackwell-order models with explicit coalition nulls;
- causal representation and multi-agent latent-state identity under replacement;
- formal theories of transactive memory with local-to-global topology;
- organizational cybernetics that couples self-maintenance to the preservation of distributed informational organization; and
- post-August-2026 work that combines representational geometry with persistent human–AI or multi-agent agency;
- formal analysis of active shared context graphs, shared group memory, and artifact-centric graph workspaces under coalition and intervention tests.

A specialist review from each adjacent field may uncover precedents not visible through cross-disciplinary keyword search.

92.5 Evidence required for a stronger claim

Before presenting the synthesis as established rather than proposed, the following should be completed:

1. a public, timestamped manuscript and versioned source archive;
2. theorem-level comparison with the closest sheaf, causal-state, group agency, and CCP formulations;
3. independent mathematical review of the nonstandard claims and constants;
4. preregistered benchmark gates and matched null families;
5. at least one controlled human–AI or multi-agent experiment estimating the complete certificate vector; and
6. explicit reporting of failures, unidentified quantities, and decomposition sensitivity.

92.6 Priority conclusion

The literature review narrows rather than weakens the theory. It removes unnecessary novelty claims from the standard mathematics and leaves a clearer research object:

Not a new Laplacian, sheaf, causal state, or controller, but a candidate unified certificate for relational workspaces.

That is the priority position adopted here.

PART XXIV

Theory Hierarchy

93. Levels of Organization

Level 0: Archive

$$\delta_{\mathrm{coh}} \text{ may be low}, \qquad \eta_{\mathrm{causal}} = 0.$$

Level 1: Shared representation

$$d_\varepsilon > 0, \qquad \delta_{\mathrm{coh}} \le \varepsilon.$$

Level 2: Coordinated operational workspace

$$\delta^{\mathcal{I}}_{\mathrm{align}} \le \varepsilon, \qquad \eta^{\nu}_{\mathrm{causal}} > 0.$$

Level 3: Relational workspace

$$k_\varepsilon \ge 2, \qquad \Delta^{\ell}_{\mathrm{joint}} > 0, \qquad \Delta^{\min}_{\mathrm{tr}}(\mathcal{Q}) > 0.$$

Level 4: Persistent relational workspace

$$d_{\mathrm{id}}(W_t, W_{t+\tau}) \le \varepsilon.$$

Level 5: Self-maintaining workspace

$$\eta_{D\to M} > 0, \qquad \eta_{M\to R} > 0, \qquad \eta_{R\to v} > 0, \qquad \Delta^{\mathrm{spec}}_{\mathrm{repair}} > 0, \qquad \sup_t e_t \le e_{\max}.$$

Level 6: Workspace agent

$$\eta_{G\to A} > 0, \qquad \eta_{W\to A} > 0, \qquad \delta_Q \le \varepsilon_Q,$$
$$\mathrm{Reg}_t(\pi) \le 2\varepsilon_Q + \delta_{\mathrm{opt}}, \qquad k_{\mathrm{ctl},\varepsilon} \ge 2, \qquad \mathrm{SelfMaintenance}\,.$$

Orthogonal formation-provenance axis

$$\mathrm{Scaffolded} \longrightarrow \mathrm{AdaptivelyParameterised} \longrightarrow \mathrm{StructurallyAdaptive} \longrightarrow \mathrm{EndogenouslyConstituted}\,.$$

This axis is independent of Levels 0–6. A scaffolded system can instantiate a high-level relational workspace, while an endogenously learned organization can fail every operational workspace gate.

Decomposition-robustness qualifier

Every level is first defined relative to a decomposition $\mathcal{D}$. A family-robust version carries the additional requirement

$$\inf_{\mathcal{D}\in\mathfrak{D}} m_q(W;\mathcal{D}) > 0 \qquad \text{for every gate required at that level.}$$

This is a robustness qualifier rather than a higher agency level. A system can be highly agentive under one fixed interface while remaining decomposition fragile.

Specification-robustness and identification qualifier

Every level is also relative to a complete scientific specification ς. A specification-robust version requires

$$\inf_{\varsigma\in\mathfrak{S}} m_q(W;\varsigma) > 0 \qquad \text{for every gate required at that level.}$$

An identified version additionally requires the sign of each required lower envelope to be constant across the realization set compatible with the available experiment family. An empirical claim must finally subtract the uniform finite-sample error radius before declaring the lower envelope positive. These are epistemic qualifiers, not additional agency levels.

Separate, unresolved axis

$$\text{PhenomenalConsciousness}\,.$$

PART XXV

Unified Formulation

The complete static-to-agentive construction is

$$
\begin{aligned}
&(\Omega, \mathfrak{D}, \mathfrak{S}, \Sigma, \mathcal{D}_{\mathrm{form}}, \mathcal{L}) \\
&\Downarrow \qquad (\mathcal{D}, \widehat{V}, \widehat{E}, \widehat{H}, \widehat{P}, \widehat{B}) \\
&\Downarrow \qquad (G, \{X_i\}, \{T_e\}) \\
&\Downarrow \qquad X_i = C_i \oplus R_i \\
&(\{C_i\}, \{H_e\}, \mathcal{A}, \{\rho_i\}) \\
&\Downarrow \qquad \Gamma(H) = \varprojlim C_i \\
&\Downarrow \qquad L_H = D_H^* D_H \\
&\Downarrow \qquad \Gamma_\varepsilon(H) = \operatorname{ran} \mathbf{1}_{[0,\varepsilon]}(L_H) \\
&\Downarrow \qquad W_t^{\mathrm{str}} \\
&\xrightarrow{\text{alignment}} Z_t = \mathbb{P}(Y_t^+ \in \cdot \mid \mathbf{H}_t) \\
&\Downarrow \qquad \mathrm{SRW}_\Theta(W_t; \mathcal{D}) \\
&\Downarrow \qquad (W_t, G_t, \widehat{Q}_t, M_t, B_t) \\
&\xrightarrow{\text{goal and value use}} (A_t, R_t) \\
&\xrightarrow{\text{causal repair}} e_t \le e_{\max} \\
&\Downarrow \qquad \mathbf{W} = [\mathfrak{R}]_{\text{gauge+causal equivalence}}.
\end{aligned}
$$

Compactly:

scaffold and history → learned organization and boundary
→ local core–transfer realization
→ compatible sections and global descent
→ causal operation and persistent identity
→ cue-controlled goal retention and counterfactual action
→ monitor-mediated repair and bounded maintenance deficit
→ decomposition- and specification-robust identified certification.

The final two arrows are optional extensions rather than consequences of the earlier ones. A relational workspace can exist without being an agent, and an agentive controller can fail to maintain the workspace that realizes it.

PART XXVI

Discussion, Limitations, and Philosophical Consequences

94. Decomposition Relativity

Part XVII turns decomposition relativity from a caveat into a formal part of the certificate. The central lesson is negative and positive at once. No claim of relationality is invariant under arbitrary redescription. An indiscrete merge makes the whole state local, while a secret-sharing split can make a centralized encoder look irreducibly distributed. Yet exact structural content can remain invariant under faithful coarse-graining, and gate margins can remain stable through a declared family of operationally meaningful decompositions.

The resulting position is neither that decomposition is arbitrary nor that one partition is uniquely privileged. Decompositions are constrained by access, causal interfaces, intervention separability, temporal continuity, and explicit shared resources. A certificate established only at one interface is a valid but limited claim. A stronger claim reports the full envelope

$$\left(\widetilde{m}_q, \underline{m}_q, \overline{m}_q, \Delta_q^{\mathrm{dec}}\right)$$

for each required gate, where $\tilde{m}_q$ is the reference margin and the remaining terms summarize its behavior across $\mathfrak{D}$.

This does not eliminate judgment. It makes the judgment inspectable. Analysts can disagree about the admissible family and identify exactly which merge, split, hidden resource, or intervention assumption changes the conclusion. That is preferable to treating a convenient node diagram as if it were part of nature itself.

94.1 Specification relativity and epistemic restraint

Decomposition is only one coordinate of the scientific claim. The downstream target, loss, temporal horizon, intervention family, null panel, and threshold rule also determine what is being certified. A workspace may be operationally real for a month-long theorem-development trajectory yet irrelevant to a second-scale motor target. Those are distinct specifications, not competing answers to one underspecified question.

The framework therefore distinguishes truth under a declared specification from robustness across an admissible family. It also distinguishes robustness inside one fitted realization from identification across all realizations compatible with the available experiments. A positive point estimate at one favorable analysis is the weakest admissible conclusion. A stronger claim requires a positive lower envelope across $\mathfrak{S}$, an identified sign across the compatible realization set, and a finite-sample confidence bound that remains positive after uniform error control.

This hierarchy is intentionally conservative. It does not demand invariance under every imaginable redescription or target. It demands that the scientific universe over which robustness is claimed be declared before certification and that ambiguity outside the reach of the experiment be reported as partial or non-identification rather than converted into certainty.

95. Realism Without Reification

The framework is realist about relational organization but does not posit a free-standing substance between the components. A workspace exists only through a realization, and every change in the workspace must be implemented by a change in participants, artifacts, channels, or their states. The quotient

$$\mathbf{W} = [\mathfrak{R}]_{\simeq}$$

is therefore a mathematical account of identity across realizations, not the name of an additional physical object.

This position allows two claims that are sometimes treated as incompatible. First, the workspace is physically dependent: no realization, no workspace. Second, it is not identical with any particular carrier when several distinct realizations preserve the same structural and causal organization. The same logic is familiar whenever an object is identified by invariants rather than by one material instance. A melody is not a second substance beside its performances, yet it is not identical with one performance. A proof is not separate from every inscription that carries it, yet its identity can survive changes of notation and medium. The analogy is limited, but it clarifies the intended form of realism.

The theory is therefore compatible with physicalism and multiple realization. It rejects both the view that a workspace must be wholly inside one component and the stronger view that the workspace is an independent entity floating outside its carriers. What is real is the organized pattern of compatible representation, transfer, causal use, persistence, and reconstruction.

96. Location and the Sense of "Between"

Asking where a relational workspace is located can invite a category mistake. The local states have ordinary physical locations. A notebook is on a desk, a model state is implemented on hardware, a message is transmitted through a channel, and a person's activity is realized in a biological organism. The workspace as such is not located at an additional point. Its location is distributed in the modest sense that the relations constituting it span several located carriers.

The word "between" is best understood counterfactually rather than spatially. Suppose local marginal states are held fixed while transfer organization is shuffled, delayed, or replaced by a matched null. If the joint predictive or causal capacity changes, then an operative property resided in the organization of the relations rather than in any one local state. Conversely, if a single component or matched storage surrogate reproduces the complete capacity, the case for a relational workspace fails. The relevant question is therefore not whether an invisible object occupies the interval between participants. It is whether the coupling carries indispensable state and supports capacities that disappear when that coupling is altered.

This account also explains why the workspace can be partially present in each component without being fully stored in any of them. Each local chart may encode a view, role, or constraint. The global object is the compatible and causally operative organization of those views. In exact settings this is a global section. In approximate settings it is a stable low-energy subspace together with the dynamic and causal conditions developed above.

97. Continuity, Contribution, and Shared Ownership

Relational persistence does not require an unchanging set of participants. A workspace can survive replacement when the new realization preserves the relevant transfer structure, operation algebra, causal response, reconstructive capacity, and non-collapse margin. This is continuity through

organized transport, not continuity through an immutable token.

The same point applies to artifacts. A handoff document, theorem register, code repository, or experiment log is not merely a passive record when it constrains future reasoning and permits reconstruction after interruption. Such artifacts can become constitutive parts of a workspace realization. Their contribution is causal and testable: remove or corrupt them, then measure what can no longer be reconstructed or done.

Distributedness does not imply equal contribution. One participant may supply long-horizon goals, another formal manipulation, and an artifact precise persistent memory. Nor does a system-level description erase individual authorship or responsibility. The theory describes how capacities are organized; it does not settle legal, ethical, or scholarly questions of credit. Those questions require a separate account of who introduced which ideas, who exercised control, who bore risk, and which contributions were replaceable.

The phrase "shared ownership" should therefore be used carefully. No individual component need own a complete copy of the workspace state, yet individuals can still own artifacts, originate arguments, and retain responsibility for decisions. Relational identity and ordinary ownership are different relations.

98. Agency, Consciousness, and Moral Non-Implications

The agency certificate developed here is functional and domain relative. It concerns persistent goals, interventional value estimates, feasible action control, damage detection, repair, and bounded maintenance within a declared environment. Passing those tests would justify a claim of workspace-level agency in that domain. It would not establish unrestricted autonomy.

No theorem in the framework identifies agency with phenomenal consciousness. A system may maintain a goal, evaluate counterfactual actions, and repair its own organization without there being anything it is like to be that system. The converse also remains open: failure to satisfy this engineering certificate would not prove the absence of consciousness. The framework is designed to prevent functional evidence from silently becoming a metaphysical conclusion.

Moral status is likewise not inferred. Neither coherence, causal efficacy, relational irreducibility, nor self-maintenance is by itself a criterion of moral patienthood. Future arguments connecting these properties to consciousness or welfare would need independent premises and evidence. Until then, the framework can support careful descriptions of organization and agency without treating a collaborative system as a person or dismissing the moral standing of the humans who participate in it.

99. Limitations

The theory is deliberately broad, but its conclusions remain bounded by several limitations.

1. **Admissible-decomposition burden.** Part XVII formalizes faithful coarse-graining and family-robust certification, but the choice of primitive carriers and admissible alternatives remains an empirical and methodological judgment. A narrow family can still overstate robustness.
2. **Admissible-specification burden.** Part XVIII makes the target, loss, horizon, intervention family, null panel, and thresholds explicit. The admissible family $\mathfrak{S}$ still requires scientific judgment. A family chosen after seeing the results can conceal fragility even when every reported analysis is technically correct.

3. **Limited observability and partial identification.** Many practical systems expose messages and outputs rather than complete local states. Transfer maps, predictive states, reconstruction operators, and even the certificate sign may therefore be only partially identified. A point claim is unjustified when both passing and failing realizations remain compatible with the declared experiment.
4. **Linear and finite-dimensional emphasis.** The clearest structural theorems use finite graphs and linear Hilbert-space models. Nonlinear, stochastic, symbolic, and category-valued extensions require additional work.
5. **Gauge and model ambiguity.** Even with abundant data, local coordinates need not be identifiable. Learned transfers are often meaningful only up to a gauge class and relative to a selected model family.
6. **Causal identification burden.** Workspace interventions must be feasible and matched. Observational dependence, activation, or predictive accuracy cannot substitute for interventions when a causal claim is made.
7. **Certification cost.** The complete certificate requires coalition tests, null models, replacement studies, longitudinal interventions, admissible specification sweeps, and identified-set or uniform uncertainty calculations. Exhaustive certification may be impractical in large systems, and adaptive exploration can invalidate nominal error guarantees unless selection and certification are separated or valid selective procedures are used.
8. **Conditional mathematical results.** The theorems establish what follows from stated assumptions. They do not show that naturally occurring collaborations satisfy those assumptions, and several constants may be too conservative for practical estimation.
9. **Open empirical status.** The motivating human–AI collaboration has not been subjected to the complete frozen certificate. It remains a plausible case study and a source of hypotheses, not a demonstrated instance of relational agency.
10. **Open priority boundary.** The literature comparison is broad but cannot prove absolute firstness. Closely related work may exist under other terminology or appear after the review date.

These limitations do not make the framework vacuous. They specify the conditions under which its claims can fail and identify the measurements needed to strengthen them.

100. Open Problems

Several directions are mathematically and empirically immediate.

1. Develop nonlinear and stochastic analogues of the transfer Laplacian, including kernel, manifold, and category-valued realizations.
2. Establish sharp uniform finite-sample estimators for workspace dimension, coalition risk, matched-null gaps, replacement identity, and specification envelopes under adaptive data collection.
3. Characterize the identified set when local states are latent and only messages, artifacts, or actions are observed. Determine the minimum intervention family needed to identify each certificate gate and the complete conjunctive certificate.

4. Extend gluing and persistence to changing graph topology, asynchronous communication, and delayed or lossy transfers.
5. Extend the joint decomposition-and-specification theory to latent shared resources, overlapping component systems, continuous coarse-graining, target families, and data-driven discovery of admissible interface and specification families without invalidating certification.
6. Study the interaction between workspace formation and strategic participants whose local objectives need not agree.
7. Build benchmark suites in which structural, causal, persistence, and repair quantities are known exactly but remain nontrivial to infer.
8. Run preregistered human–AI and multi-agent experiments that estimate the complete certificate rather than isolated performance gains.
9. Compare certificate-directed self-maintenance with established notions of organizational resilience, institutional memory, and collective agency.
10. Clarify which additional evidence, if any, could connect functional workspace predicates to theories of consciousness without assuming the connection in advance.

101. Conclusion

A stratified relational workspace is a temporally persistent, causally efficacious, jointly reconstructible structure represented only partially in several heterogeneous components. Its identity is determined by invariants preserved under admissible transfer, and its operative capacity is relational only when no individual component or matched local-storage null reproduces the relevant function.

The theory gives this definition several independent layers. Exact shared content appears as a global compatible section and, in the invertible case, as a holonomy-fixed core. Approximate shared content appears as a low-energy spectral subspace of the transfer Laplacian. Predictive states and feasible interventions determine whether that structure matters for a declared future. Coalition tests and matched nulls determine whether the capacity is genuinely distributed. Reconstruction frames, temporal transport, and identity profiles determine whether it survives loss and replacement. Descent determines whether local workspaces compose globally. Provenance tests distinguish learned organization from scaffolded answers. Goal, value, damage, and repair interventions determine whether the workspace supports domain-relative agency and self-maintenance.

No one layer can stand in for the others. Coherence can describe an archive. Performance can describe a centralized controller. Persistence can be trivial if all content collapses. A task-equivalent replacement can alter the causal organization. Local compatibility can fail to glue globally. Endogenous-looking structure can be supplied through hidden leakage, and successful recovery can be open-loop replay rather than repair. The value of the framework lies in making these alternatives explicit and testable.

Relationality also cannot be detached from decomposition. Faithful coarse-graining preserves the exact structural core, but coalition counts and local sufficiency change when components are merged. Arbitrary splitting can create a formally irreducible code around a centralized encoder. The strongest claims therefore quantify their margins across an admissible family of interfaces rather than relying on one favorable diagram.

A complete scientific claim must also survive scrutiny of the specification and the experiment. The

strongest form of the certificate reports a positive gatewise lower envelope across a predeclared family of decompositions and scientific specifications, shows that the sign of that envelope is identified across all realizations compatible with the intervention data, and retains a positive margin after uniform finite-sample uncertainty. When these conditions fail, the correct conclusion is specification-specific, fragile, partially identified, or unidentified rather than simply positive or negative.

The central identity principle is therefore

workspace identity = persistent invariant causal organization under admissible transport.

This principle is compatible with changing carriers, representations, and participants, but only when structural, operational, causal, reconstructive, and non-collapse defects remain within declared tolerances. It provides a formal language for studying distributed collaboration without turning that collaboration into a mysterious third substance and without inferring consciousness from organization alone.

References

[1] S. Abramsky and A. Brandenburger. The sheaf-theoretic structure of non-locality and contextuality. *New Journal of Physics*, 13:113036, 2011. doi:10.1088/1367-2630/13/11/113036.

[2] F. Adams and K. Aizawa. The bounds of cognition. *Philosophical Psychology*, 14(1):43–64, 2001. doi:10.1080/09515080120033571.

[3] A. D. Ames, X. Xu, J. W. Grizzle, and P. Tabuada. Control barrier function based quadratic programs for safety critical systems. *IEEE Transactions on Automatic Control*, 62(8):3861–3876, 2017. doi:10.1109/TAC.2016.2638961.

[4] C. Amornbunchornvej. Collective counterfactual planning: Coordination, consent, and verification under representational constraints. arXiv:2608.17932, 2026. doi:10.48550/arXiv.2608.17932.

[5] W. R. Ashby. *An Introduction to Cybernetics*. Chapman & Hall, London, 1956.

[6] J.-P. Aubin. *Viability Theory*. Birkhäuser, Boston, 1991. doi:10.1007/978-0-8176-4910-4.

[7] B. J. Baars. *A Cognitive Theory of Consciousness*. Cambridge University Press, Cambridge, 1988.

[8] X. E. Barandiaran, E. Di Paolo, and M. Rohde. Defining agency: Individuality, normativity, asymmetry, and spatio-temporality in action. *Adaptive Behavior*, 17(5):367–386, 2009. doi:10.1177/1059712309343819.

[9] D. Blackwell. Equivalent comparisons of experiments. *Annals of Mathematical Statistics*, 24(2):265–272, 1953. doi:10.1214/aoms/1177729032.

[10] R. Berk, L. Brown, A. Buja, K. Zhang, and L. Zhao. Valid post-selection inference. *The Annals of Statistics*, 41(2):802–837, 2013. doi:10.1214/12-AOS1077.

[11] G. E. Bredon. *Sheaf Theory*. 2nd edition, Springer, New York, 1997.

[12] M. E. Bratman. Shared cooperative activity. *The Philosophical Review*, 101(2):327–341, 1992. doi:10.2307/2185537.

[13] P. G. Casazza and J. Kovačević. Equal-norm tight frames with erasures. *Advances in Computational Mathematics*, 18:387–430, 2003. doi:10.1023/A:1021349819855.

[14] O. Christensen. *An Introduction to Frames and Riesz Bases*. 2nd edition, Birkhäuser, Cham, 2016.

[15] A. Clark and D. Chalmers. The extended mind. *Analysis*, 58(1):7–19, 1998. doi:10.1093/analys/58.1.7.

[16] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. 2nd edition, Wiley, Hoboken, 2006.

[17] C. Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. III. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970. doi:10.1137/0707001.

[18] S. Dehaene and J.-P. Changeux. Experimental and theoretical approaches to conscious processing. *Neuron*, 70(2):200–227, 2011. doi:10.1016/j.neuron.2011.03.018.

[19] D. Dellermann, P. Ebel, M. Söllner, and J. M. Leimeister. Hybrid intelligence. *Business & Information Systems Engineering*, 61:637–643, 2019. doi:10.1007/s12599-019-00595-2.

[20] E. A. Di Paolo. Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences*, 4:429–452, 2005. doi:10.1007/s11097-005-9002-y.

[21] R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasoning About Knowledge*. MIT Press, Cambridge, Massachusetts, 1995.

[22] P. Gupta, T. Nguyen, C. Gonzalez, and A. W. Woolley. Fostering collective intelligence in human–AI collaboration: Laying the groundwork for COHUMAIN. *Topics in Cognitive Science*, 17(2):189–216, 2025. doi:10.1111/tops.12679.

[23] J. Y. Halpern and Y. Moses. Knowledge and common knowledge in a distributed environment. *Journal of the ACM*, 37(3):549–587, 1990. doi:10.1145/79147.79161.

[24] J. Hansen and R. Ghrist. Toward a spectral theory of cellular sheaves. *Journal of Applied and Computational Topology*, 3:315–358, 2019. doi:10.1007/s41468-019-00038-7.

[25] A. Hatcher. *Algebraic Topology*. Cambridge University Press, Cambridge, 2002.

[26] B. Hayes-Roth. A blackboard architecture for control. *Artificial Intelligence*, 26(3):251–321, 1985. doi:10.1016/0004-3702(85)90063-3.

[27] J. Hollan, E. Hutchins, and D. Kirsh. Distributed cognition: Toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction*, 7(2):174–196, 2000. doi:10.1145/353485.353487.

[28] E. Hutchins. *Cognition in the Wild*. MIT Press, Cambridge, Massachusetts, 1995.

[29] T. Kato. *Perturbation Theory for Linear Operators*. Springer, Berlin, 2nd edition, 1995.

[30] A. S. Klyubin, D. Polani, and C. L. Nehaniv. Empowerment: A universal agent-centric measure of control. In *2005 IEEE Congress on Evolutionary Computation*, pages 128–135, 2005. doi:10.1109/CEC.2005.1554676.

[31] G. W. Imbens and C. F. Manski. Confidence intervals for partially identified parameters. *Econometrica*, 72(6):1845–1857, 2004. doi:10.1111/j.1468-0262.2004.00555.x.

[32] E. E. Leamer. Let's take the con out of econometrics. *The American Economic Review*, 73(1):31–43, 1983.

[33] K. G. Larsen and A. Skou. Bisimulation through probabilistic testing. *Information and Computation*, 94(1):1–28, 1991. doi:10.1016/0890-5401(91)90030-6.

[34] K. Lewis. Measuring transactive memory systems in the field: Scale development and validation. *Journal of Applied Psychology*, 88(4):587–604, 2003. doi:10.1037/0021-9010.88.4.587.

[35] C. List and P. Pettit. *Group Agency: The Possibility, Design, and Status of Corporate Agents*. Oxford University Press, Oxford, 2011.

[36] L. Ljung. *System Identification: Theory for the User*. 2nd edition, Prentice Hall, Upper Saddle River, New Jersey, 1999.

[37] S. Mac Lane. *Categories for the Working Mathematician*. 2nd edition, Springer, New York, 1998.

[38] H. R. Maturana and F. J. Varela. *Autopoiesis and Cognition: The Realization of the Living*. D. Reidel, Dordrecht, 1980.

[39] R. Olfati-Saber, J. A. Fax, and R. M. Murray. Consensus and cooperation in networked multi-agent systems. *Proceedings of the IEEE*, 95(1):215–233, 2007. doi:10.1109/JPROC.2006.887293.

[40] C. F. Manski. *Partial Identification of Probability Distributions*. Springer, New York, 2003. doi:10.1007/b97478.

[41] J. Pearl. *Causality: Models, Reasoning, and Inference*. 2nd edition, Cambridge University Press, Cambridge, 2009.

[42] C. Riedl, S. Savage, and J. Zvelebilova. Cognitive spillover in human–AI teams. *ACM Transactions on Computer-Human Interaction*, 2026. doi:10.1145/3805039; arXiv:2407.17489v3.

[43] M. Robinson. Sheaves are the canonical data structure for sensor integration. *Information Fusion*, 36:208–224, 2017. doi:10.1016/j.inffus.2016.12.002.

[44] R. D. Rupert. Challenges to the hypothesis of extended cognition. *The Journal of Philosophy*, 101(8):389–428, 2004. doi:10.5840/jphil2004101826.

[45] I. Seeber, E. Bittner, R. O. Briggs, T. de Vreede, G.-J. de Vreede, A. Elkins, R. Maier, A. B. Merz, S. Oeste-Reiß, N. Randrup, G. Schwabe, and M. Söllner. Machines as teammates: A research agenda on AI in team collaboration. *Information & Management*, 57(2):103174, 2020. doi:10.1016/j.im.2019.103174.

[46] C. R. Shalizi and J. P. Crutchfield. Computational mechanics: Pattern and prediction, structure and simplicity. *Journal of Statistical Physics*, 104:817–879, 2001. doi:10.1023/A:1010388907793.

[47] M. Sidji, W. Smith, and M. J. Rogerson. Adopting the theory of distributed cognition for human–AI cooperation. In *Proceedings of the 36th Australasian Conference on Human-Computer Interaction (OzCHI '24)*, pages 774–779, 2024. doi:10.1145/3726986.3727024.

[48] U. Simonsohn, J. P. Simmons, and L. D. Nelson. Specification curve analysis. *Nature Human Behaviour*, 4(11):1208–1214, 2020. doi:10.1038/s41562-020-0912-z.

[49] A. Singer and H.-T. Wu. Vector diffusion maps and the connection Laplacian. *Communications on Pure and Applied Mathematics*, 65(8):1067–1144, 2012. doi:10.1002/cpa.21395.

[50] S. Steegen, F. Tuerlinckx, A. Gelman, and W. Vanpaemel. Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, 11(5):702–712, 2016. doi:10.1177/1745691616658637.

[51] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. 2nd edition, MIT Press, Cambridge, Massachusetts, 2018.

[52] D. M. Wegner. Transactive memory: A contemporary analysis of the group mind. In B. Mullen and G. R. Goethals, editors, *Theories of Group Behavior*, pages 185–208. Springer, New York, 1987. doi:10.1007/978-1-4612-4634-3_9.

[53] P. L. Williams and R. D. Beer. Nonnegative decomposition of multivariate information. arXiv:1004.2515, 2010.

[54] A. W. Woolley, C. F. Chabris, A. Pentland, N. Hashmi, and T. W. Malone. Evidence for a collective intelligence factor in the performance of human groups. *Science*, 330(6004):686–688, 2010. doi:10.1126/science.1193147.

[55] H. Zhao and Z. Han. Understanding the mechanism of human–AI interaction: A distributed cognition perspective. *Journal of Documentation*, 82(4):1042–1063, 2026. doi:10.1108/JD-01-2026-0003.

[56] S. Choudhury et al. Networked intelligence: Active shared context graphs for human–AI team science. arXiv:2607.13220v3, 2026. doi:10.48550/arXiv.2607.13220.

[57] A. Goyal, A. Didolkar, A. Lamb, K. Badola, N. R. Ke, N. Rahaman, J. Binas, C. Blundell, M. Mozer, and Y. Bengio. Coordination among neural modules through a shared global workspace. In *International Conference on Learning Representations*, 2022. arXiv:2103.01197v2.

[58] N. Kotalwar, R. Das, and C. Rosé. Searching for synergy in shared workspace human–AI collaboration. arXiv:2606.18413v3, 2026. doi:10.48550/arXiv.2606.18413.

[59] R. Li and B. Jiang. Thinking in graphs with CoMAP: A shared visual workspace for designing project-based learning. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*, 2026. doi:10.1145/3772318.3790273.

[60] M. J. Pickering and S. Garrod. *Understanding Dialogue: Language Use and Social Interaction*. Cambridge University Press, Cambridge, 2021.