

SRIDHAR MAHADEVAN

# INFINITESIMAL CREATIVITY

Learning by Double Involution

VOLUME III OF THE CATEGORICAL AI RESEARCH PROGRAM



# Contents

|                                                                     |    |
|---------------------------------------------------------------------|----|
| <i>Preface</i>                                                      | 13 |
| <i>A Trail Map of Infinitesimal Creativity</i>                      | 19 |
| <i>I Foundations of Infinitesimal Creativity</i>                    | 23 |
| <i>o A Working Language for Creative Theory Change</i>              | 25 |
| <i>o.1 Progressive abstraction as creative reconstruction</i>       | 26 |
| <i>o.2 Categories: typed composition</i>                            | 27 |
| <i>o.3 Diagrams and compositional promises</i>                      | 27 |
| <i>o.4 Functors and natural transformations</i>                     | 28 |
| <i>o.5 Universal constructions: defining by relations</i>           | 29 |
| <i>o.6 Algebraic theories, sketches, and models</i>                 | 30 |
| <i>o.7 Lexical creativity: extending a language</i>                 | 31 |
| <i>o.8 Presheaves, sheaves, and contextual agreement</i>            | 32 |
| <i>o.9 Tangent categories and local possibility</i>                 | 33 |
| <i>o.10 Weil algebras as probe shapes</i>                           | 34 |
| <i>o.11 Classical double Lie algebroids: a concrete realization</i> | 35 |
| <i>o.12 Two axes: Boden and Piaget</i>                              | 37 |
| <i>o.13 The geometry example as a creative trace</i>                | 38 |
| <i>o.14 How to read the rest of the book</i>                        | 39 |

|      |                                                                        |    |
|------|------------------------------------------------------------------------|----|
| 1    | <i>Initial Sketch Acquisition</i>                                      | 41 |
| 1.1  | <i>The initialization problem</i>                                      | 41 |
| 1.2  | <i>Four routes into a domain</i>                                       | 42 |
| 1.3  | <i>Explanation-based learning as a precedent</i>                       | 44 |
| 1.4  | <i>Learning apprentices and continuous knowledge acquisition</i>       | 45 |
| 1.5  | <i>What must be acquired from a demonstration?</i>                     | 45 |
| 1.6  | <i>Coaching as typed counter-witness selection</i>                     | 46 |
| 1.7  | <i>Conservation as developmental sketch acquisition</i>                | 47 |
| 1.8  | <i>From acquisition to double involution</i>                           | 48 |
| 1.9  | <i>An initial acquisition contract</i>                                 | 50 |
| 1.10 | <i>A controlled acquisition benchmark</i>                              | 51 |
| 1.11 | <i>What this chapter does not assume</i>                               | 51 |
| 2    | <i>Creativity as Theory Construction</i>                               | 53 |
| 2.1  | <i>Three different achievements</i>                                    | 54 |
| 2.2  | <i>Assimilation, accommodation, and cognitive development</i>          | 56 |
| 2.3  | <i>Two interacting classes of repair</i>                               | 57 |
| 2.4  | <i>Boden's three modes</i>                                             | 58 |
| 2.5  | <i>Four evaluation media, five admission regimes</i>                   | 59 |
| 2.6  | <i>From LINC'S repair to creative transformation</i>                   | 61 |
| 2.7  | <i>The status-bearing extension dossier</i>                            | 62 |
| 2.8  | <i>What counts as success?</i>                                         | 63 |
| 2.9  | <i>Personal and historical creativity</i>                              | 63 |
| 2.10 | <i>From a target to a mechanism</i>                                    | 64 |
| 3    | <i>The Differential Geometry of Creativity</i>                         | 65 |
| 3.1  | <i>The ambient differential universe</i>                               | 67 |
| 3.2  | <i>Weil algebras as probe shapes</i>                                   | 68 |
| 3.3  | <i>The missing bridge: from tangent semantics to double algebroids</i> | 69 |
| 3.4  | <i>Tensor products of sketches as a semantics for doubling</i>         | 70 |
| 3.5  | <i>A Weil profile of a theory state</i>                                | 76 |

|      |                                                              |     |
|------|--------------------------------------------------------------|-----|
| 3.6  | <i>Double-infinitesimal repair geometry</i>                  | 76  |
| 3.7  | <i>Curve objects and the dynamics of theory revision</i>     | 79  |
| 3.8  | <i>Assimilation and accommodation in differential form</i>   | 82  |
| 3.9  | <i>The three modes in differential form</i>                  | 84  |
| 3.10 | <i>Differential theory extensions</i>                        | 85  |
| 3.11 | <i>A differential creativity process</i>                     | 86  |
| 3.12 | <i>Beyond two directions: the higher-DIAL conjecture</i>     | 87  |
| 3.13 | <i>What the differential account does not claim</i>          | 90  |
| 4    | <i>Skill Optimization for Creative Exploration</i>           | 93  |
| 4.1  | <i>From a script to a creative skill</i>                     | 94  |
| 4.2  | <i>The Lie algebroid of creative moves</i>                   | 95  |
| 4.3  | <i>Brackets expose interaction</i>                           | 96  |
| 4.4  | <i>Assimilation and accommodation in skill geometry</i>      | 97  |
| 4.5  | <i>Latent skill variation</i>                                | 99  |
| 4.6  | <i>Two coupled optimization problems</i>                     | 99  |
| 4.7  | <i>DIAL-SkillOpt: training a discovery skill</i>             | 100 |
| 4.8  | <i>Boden's modes as changes of control geometry</i>          | 105 |
| 4.9  | <i>Scope and experimental consequences</i>                   | 105 |
| 5    | <i>Algebraic Theories as Creative Targets</i>                | 107 |
| 5.1  | <i>Lawvere's functorial semantics</i>                        | 108 |
| 5.2  | <i>Sketches present theories</i>                             | 108 |
| 5.3  | <i>From products to finite limits</i>                        | 109 |
| 5.4  | <i>PROPs and monoidal theories</i>                           | 110 |
| 5.5  | <i>From presentations to Grothendieck toposes</i>            | 111 |
| 5.6  | <i>The creative target package</i>                           | 112 |
| 5.7  | <i>From a double repair theory to a finite theory change</i> | 113 |
| 5.8  | <i>Kan extensions as the transport semantics of DIAL</i>     | 115 |
| 5.9  | <i>Developmental towers of theory</i>                        | 118 |
| 5.10 | <i>Boden's modes revisited</i>                               | 119 |

|      |                                                            |     |
|------|------------------------------------------------------------|-----|
| 6    | <i>Landscapes of Computational Creativity</i>              | 123 |
| 6.1  | <i>A map by creative object</i>                            | 124 |
| 6.2  | <i>Conceptual spaces and transformations</i>               | 124 |
| 6.3  | <i>Creativity as a neurocognitive process</i>              | 125 |
| 6.4  | <i>Assimilation and accommodation across the landscape</i> | 126 |
| 6.5  | <i>Prediction, explanation, and structural possibility</i> | 128 |
| 6.6  | <i>Novelty, information, and surprise</i>                  | 130 |
| 6.7  | <i>Discovery as heuristic search</i>                       | 131 |
| 6.8  | <i>Local novelty, integrability, and persistence</i>       | 133 |
| 6.9  | <i>Artifacts, process, and evaluation</i>                  | 134 |
| 6.10 | <i>Imagination and counterfactual worlds</i>               | 135 |
| 6.11 | <i>Cellular automata and emergent conceptual structure</i> | 135 |
| 6.12 | <i>Evolution as a creative constructor</i>                 | 137 |
| 6.13 | <i>AI Feynman as a controlled testbed</i>                  | 138 |
| 6.14 | <i>A four-level symbolic-discovery challenge</i>           | 139 |
| 6.15 | <i>Large language models in scientific discovery</i>       | 141 |
| 7    | <i>Combinational Creativity</i>                            | 147 |
| 7.1  | <i>From juxtaposition to composition</i>                   | 148 |
| 7.2  | <i>Interface discovery</i>                                 | 149 |
| 7.3  | <i>Assimilation and accommodation</i>                      | 149 |
| 7.4  | <i>Differential tests for interaction</i>                  | 150 |
| 7.5  | <i>Combination inside a tensor-product theory</i>          | 151 |
| 7.6  | <i>Combinations across the AGENTIC testbeds</i>            | 152 |
| 7.7  | <i>A combinational experiment protocol</i>                 | 152 |
| 7.8  | <i>Failure modes and admission</i>                         | 153 |
| 8    | <i>Exploratory Creativity</i>                              | 155 |
| 8.1  | <i>The boundary of a conceptual space</i>                  | 155 |
| 8.2  | <i>Assimilation as exploration</i>                         | 156 |
| 8.3  | <i>Local exploration with Weil probes</i>                  | 157 |

|       |                                                                   |     |
|-------|-------------------------------------------------------------------|-----|
| 8.4   | <i>Exploration as an integrable path</i>                          | 157 |
| 8.5   | <i>Active probe selection</i>                                     | 158 |
| 8.6   | <i>Novelty, surprise, and interestingness</i>                     | 159 |
| 8.7   | <i>From exploration to obstruction</i>                            | 159 |
| 8.8   | <i>Exploration across the testbeds</i>                            | 160 |
| 8.9   | <i>The exploratory audit record</i>                               | 161 |
| 9     | <i>Transformational Creativity</i>                                | 163 |
| 9.1   | <i>What can be transformed?</i>                                   | 163 |
| 9.2   | <i>Declaration interventions</i>                                  | 165 |
| 9.3   | <i>Accommodation and equilibration</i>                            | 165 |
| 9.4   | <i>From infinitesimal evidence to a finite jump</i>               | 166 |
| 9.5   | <i>Transport, forgetting, and conceptual loss</i>                 | 167 |
| 9.6   | <i>An obstruction-to-extension workflow</i>                       | 168 |
| 9.7   | <i>Transformations across the testbeds</i>                        | 169 |
| 9.8   | <i>Evaluating transformational creativity</i>                     | 170 |
| 9.9   | <i>The boundary of the claim</i>                                  | 171 |
| II    | <i>Algorithmic Realizations</i>                                   | 173 |
| 10    | <i>DIAL as an Algorithmic Family</i>                              | 175 |
| 10.1  | <i>The missing algorithmic layer</i>                              | 176 |
| 10.2  | <i>The four-level implementation invariant</i>                    | 177 |
| 10.3  | <i>Where creativity enters</i>                                    | 178 |
| 10.4  | <i>Three targets for every DIAL-X method</i>                      | 181 |
| 10.5  | <i>The DIAL-X contract</i>                                        | 183 |
| 10.6  | <i>The generic DIAL-X algorithm</i>                               | 186 |
| 10.7  | <i>DIAL-GIRL: the target is an admitted decision state</i>        | 188 |
| 10.8  | <i>DIAL-ALLORA: the target is an admitted adapter composition</i> | 190 |
| 10.9  | <i>CLIC: creative causal learning</i>                             | 191 |
| 10.10 | <i>OPTIC: creative skill optimization</i>                         | 191 |

|       |                                                                  |     |
|-------|------------------------------------------------------------------|-----|
| 10.11 | <i>Further members of the DIAL-X family</i>                      | 192 |
| 10.12 | <i>How DIAL-X sharpens an experiment</i>                         | 193 |
| 10.13 | <i>Boundary of the algorithmic claim</i>                         | 194 |
| 11    | <i>Causal Learning with Infinitesimal Creativity</i>             | 197 |
| 11.1  | <i>The CLIC target card</i>                                      | 197 |
| 11.2  | <i>Exact closure and typed repair</i>                            | 199 |
| 11.3  | <i>Finite operators and propagated uncertainty</i>               | 200 |
| 11.4  | <i>Active counter-witness interventions</i>                      | 201 |
| 11.5  | <i>Explicit alternative posteriors</i>                           | 203 |
| 11.6  | <i>From local repair to causal admission</i>                     | 204 |
| 12    | <i>Skill Optimization with Infinitesimal Creativity</i>          | 207 |
| 12.1  | <i>The OPTIC target card</i>                                     | 207 |
| 12.2  | <i>From vector-field brackets to involution algebroids</i>       | 209 |
| 12.3  | <i>The OPTIC contract</i>                                        | 210 |
| 12.4  | <i>An executable OPTIC episode</i>                               | 211 |
| 12.5  | <i>Evidence for skill factorization</i>                          | 213 |
| 12.6  | <i>What has been learned—and what has not</i>                    | 215 |
| 12.7  | <i>The open construction problem</i>                             | 216 |
| 12.8  | <i>CLIC, OPTIC, and RELIC</i>                                    | 216 |
| 13    | <i>Reinforcement Learning with Infinitesimal Creativity</i>      | 219 |
| 13.1  | <i>What is creative about reinforcement learning?</i>            | 219 |
| 13.2  | <i>The RELIC target card</i>                                     | 220 |
| 13.3  | <i>From GIRL to RELIC</i>                                        | 222 |
| 13.4  | <i>Schema learning in an embodied text world</i>                 | 223 |
| 13.5  | <i>The ALFWorld benchmark contract</i>                           | 224 |
| 13.6  | <i>Competence of persistent structural artifacts</i>             | 224 |
| 13.7  | <i>Why novelty and one-step guidance are insufficient</i>        | 225 |
| 13.8  | <i>Admitting a supplied temporal composition (B<sub>3</sub>)</i> | 226 |



|       |                                                                        |     |
|-------|------------------------------------------------------------------------|-----|
| 13.9  | <i>A DIAL reading of the result</i>                                    | 227 |
| 13.10 | <i>Constructing the temporal composition (B4)</i>                      | 228 |
| 13.11 | <i>Residual failure and declaration scope</i>                          | 229 |
| 13.12 | <i>Constructing a missing stage</i>                                    | 229 |
| 14    | <i>Composing Creative Workflows</i>                                    | 235 |
| 14.1  | <i>Three maintained theories</i>                                       | 236 |
| 14.2  | <i>The exchange object</i>                                             | 237 |
| 14.3  | <i>A small algebra of workflow composition</i>                         | 238 |
| 14.4  | <i>Three modes of meta-creativity</i>                                  | 239 |
| 14.5  | <i>Cross-workflow obstructions</i>                                     | 240 |
| 14.6  | <i>Canonical compositions</i>                                          | 240 |
| 14.7  | <i>An integrated episode</i>                                           | 240 |
| 14.8  | <i>An integrated testbed</i>                                           | 242 |
| 14.9  | <i>Ablations and interaction effects</i>                               | 242 |
| 14.10 | <i>A controlled demonstration: active evidence across the workflow</i> | 244 |
| 14.11 | <i>What surprising outcomes would mean</i>                             | 247 |
| 14.12 | <i>From orchestration sketch to workflow theory</i>                    | 247 |
| III   | <i>AGENTIC Testbeds and Open Domains</i>                               | 249 |
| 15    | <i>Mathematical Theory Construction</i>                                | 251 |
| 15.1  | <i>The AGENTIC mathematical contract</i>                               | 252 |
| 15.2  | <i>Historical and generated worlds</i>                                 | 254 |
| 15.3  | <i>A hierarchy of mathematical construction</i>                        | 254 |
| 15.4  | <i>An AGENTIC mathematical episode</i>                                 | 256 |
| 15.5  | <i>Admission records and formal certificates</i>                       | 256 |
| 15.6  | <i>Persistence</i>                                                     | 257 |
| 15.7  | <i>Early calibration: theory extension with Lea</i>                    | 257 |
| 15.8  | <i>A worked episode: inventing an intrinsic property</i>               | 259 |
| 15.9  | <i>From interaction laws to latent presentations</i>                   | 260 |

|       |                                                                         |     |
|-------|-------------------------------------------------------------------------|-----|
| 15.10 | <i>Symbolic laws and active mathematical experiments</i>                | 262 |
| 15.11 | <i>Cellular-automaton theory induction</i>                              | 264 |
| 15.12 | <i>An integrated benchmark</i>                                          | 265 |
| 15.13 | <i>A grand challenge: constructing a coalgebraic theory of learning</i> | 266 |
| 15.14 | <i>Pattern repair before causal explanation</i>                         | 281 |
| 16    | <i>Simulator-Grounded Theory Extension</i>                              | 285 |
| 16.1  | <i>The AGENTIC simulator contract</i>                                   | 286 |
| 16.2  | <i>Executable semantics</i>                                             | 287 |
| 16.3  | <i>Three validity layers</i>                                            | 288 |
| 16.4  | <i>Double involution in an executable world</i>                         | 289 |
| 16.5  | <i>The integrated loop</i>                                              | 289 |
| 16.6  | <i>Three Prometheus executable substrates</i>                           | 290 |
| 16.7  | <i>A worked theory-repair trace</i>                                     | 295 |
| 16.8  | <i>Ontology extension</i>                                               | 295 |
| 16.9  | <i>An evidence hierarchy</i>                                            | 297 |
| 16.10 | <i>AGENTIC-SIM: the integrated benchmark</i>                            | 298 |
| 16.11 | <i>What would count as scientific creativity?</i>                       | 299 |
| 16.12 | <i>Open problems</i>                                                    | 300 |
| 17    | <i>From Scientific Documents to Testable Theories</i>                   | 303 |
| 17.1  | <i>The transition from simulation to literature</i>                     | 304 |
| 17.2  | <i>Two directions between theory and evidence</i>                       | 305 |
| 17.3  | <i>Documents as implicit sketch presentations</i>                       | 306 |
| 17.4  | <i>The AGENTIC corpus contract</i>                                      | 307 |
| 17.5  | <i>From reports to local models</i>                                     | 309 |
| 17.6  | <i>Why a corpus is not a simulator</i>                                  | 310 |
| 17.7  | <i>The Theory Extension Package</i>                                     | 310 |
| 17.8  | <i>Controlled evidence for corpus-to-theory construction</i>            | 311 |
| 17.9  | <i>A corpus-to-theory evidence hierarchy</i>                            | 323 |
| 17.10 | <i>The GLP-1 challenge</i>                                              | 324 |

|       |                                                                         |     |
|-------|-------------------------------------------------------------------------|-----|
| 17.11 | <i>A grant proposal as a capstone artifact</i>                          | 328 |
| 17.12 | <i>Admission beyond the corpus</i>                                      | 346 |
| 17.13 | <i>Toward an integrated open-domain benchmark</i>                       | 346 |
| 18    | <i>Generative Visual Languages</i>                                      | 349 |
| 18.1  | <i>Diffusion is one infinitesimal direction</i>                         | 350 |
| 18.2  | <i>The second direction: proto-accommodation and semantic extension</i> | 351 |
| 18.3  | <i>A candidate double geometry</i>                                      | 351 |
| 18.4  | <i>From a new token to a new language</i>                               | 352 |
| 18.5  | <i>ARTISTIC as an AGENTIC system</i>                                    | 353 |
| 18.6  | <i>The ARTISTIC procedure</i>                                           | 354 |
| 18.7  | <i>Textual inversion as staged calibration</i>                          | 354 |
| 18.8  | <i>A visual-language benchmark</i>                                      | 356 |
| 18.9  | <i>An evidence architecture for visual assurance</i>                    | 357 |
| 18.10 | <i>Lexical calibration and typed realization</i>                        | 358 |
| 18.11 | <i>Learning the declaration</i>                                         | 359 |
| 18.12 | <i>Transfer to natural photographs</i>                                  | 359 |
| 18.13 | <i>Intervals, abstention, and active acquisition</i>                    | 362 |
| 18.14 | <i>From image repair to structural assurance</i>                        | 365 |
| 18.15 | <i>Visual theory invention in Bongard problems</i>                      | 371 |
| 18.16 | <i>What the experiments establish</i>                                   | 381 |
| 18.17 | <i>Measurements and failure modes</i>                                   | 382 |
| 18.18 | <i>Open mathematics</i>                                                 | 382 |
| 19    | <i>Artistic Theory Extension</i>                                        | 385 |
| 19.1  | <i>What CAN makes creative</i>                                          | 387 |
| 19.2  | <i>The state space is a space of artistic theories</i>                  | 387 |
| 19.3  | <i>Two local artistic directions</i>                                    | 388 |
| 19.4  | <i>From noncommutativity to structural obstruction</i>                  | 389 |
| 19.5  | <i>The Weil profile of an obstruction</i>                               | 390 |
| 19.6  | <i>Constructing a candidate technique</i>                               | 391 |

|       |                                                           |     |
|-------|-----------------------------------------------------------|-----|
| 19.7  | <i>Transporting the old artistic theory</i>               | 392 |
| 19.8  | <i>The admission firewall</i>                             | 392 |
| 19.9  | <i>The hierarchy from CAN to DILATE</i>                   | 393 |
| 19.10 | <i>Calibration of the mixed-order diagnostic</i>          | 394 |
| 19.11 | <i>A paintbrush over the space of generative dynamics</i> | 395 |
| 19.12 | <i>Observer-induced metric and discrete transport</i>     | 396 |
| 19.13 | <i>Registered paintbrush evidence</i>                     | 397 |
| 19.14 | <i>Theory routing across natural videos</i>               | 400 |
| 19.15 | <i>The withheld-generator experiment</i>                  | 402 |
| 19.16 | <i>Baselines and ablations</i>                            | 403 |
| 19.17 | <i>Open mathematical obligations</i>                      | 404 |
| 20    | <i>Creativity as Compositional Theory Change</i>          | 407 |
| 20.1  | <i>The argument in one view</i>                           | 407 |
| 20.2  | <i>What the present evidence supports</i>                 | 409 |
| 20.3  | <i>Why two interacting directions matter</i>              | 411 |
| 20.4  | <i>What DIAL does not explain</i>                         | 412 |
| 20.5  | <i>Music as an open compositional domain</i>              | 413 |
| 20.6  | <i>Other domains not explored here</i>                    | 416 |
| 20.7  | <i>Evaluation beyond novelty</i>                          | 417 |
| 20.8  | <i>Open obligations</i>                                   | 418 |
| 20.9  | <i>Conclusion</i>                                         | 418 |
|       | <i>Bibliography</i>                                       | 421 |
|       | <i>Glossary of Notation</i>                               | 435 |
|       | <i>Subject Index</i>                                      | 445 |

# Preface

*Infinitesimal Creativity* studies a question that lies beyond the optimization of a fixed model: how can an artificial reasoner construct, maintain, and extend a conceptual world?

*“All things are made of atoms.”*

—Richard Feynman, *The Feynman Lectures on Physics*

Feynman introduces the atomic hypothesis through a thought experiment. If a cataclysm destroyed scientific knowledge and only one statement could reach a future civilization, which statement would preserve the most information in the fewest words? His answer identifies atoms, their perpetual motion, and their attractive and repulsive interactions [Feynman et al., 1963, Section 1–2].

The force of the example is not brevity alone. The atomic hypothesis names objects, dynamics, and interaction laws from which an extraordinary range of phenomena can be reconstructed. It is a compact *generative theory*, not merely a compressed archive of observations. This book adopts that image as its governing aspiration: synthetically discover compositional structures whose models explain what is already known and make new deductions, predictions, and experiments possible.

The preceding books, *Categories for AGI* and *Machine Learning from Enforcing Compositionality*, developed the categorical language on which this book builds. They introduced diagrams, sketches, sheaves, tangent structure, typed obstructions, localization, repair, and admission. Chapter 0 provides a working primer sufficient to enter the present argument; the earlier books remain the route to broader motivation and fuller foundations. Our subject here is what happens when the declaration itself is no longer adequate.

No declaration, however, appears from nowhere. Before a reasoner can assimilate experience or accommodate a failing theory, it must acquire a provisional sketch of the domain. It may receive that sketch through explicit instruction, reconstruct it by watching competent behavior, or learn it as an apprentice whose teacher diagnoses and corrects its own attempts. Chapter 1 makes this initialization step explicit. The

acquired sketch is not presumed complete or correct; it is the inspectable starting theory on which the later double-involution process operates.

Frontier models now prove difficult theorems, generate sophisticated programs, and search spaces that were previously inaccessible. Yet solving a problem inside number theory is not the same achievement as constructing number theory. A durable mathematical or scientific field introduces objects, operations, invariants, and laws that remain useful for problems not known when the concepts were proposed.

### Challenge

Can a system recognize that its current conceptual language is inadequate, propose a more productive theory, transport what was previously known into that theory, and make the extension usable by other reasoners?

THE PHRASE DOUBLE INVOLUTION NAMES THE BOOK'S PROPOSED ORGANIZATION OF LOCAL CHANGE IN TWO INTERACTING DIRECTIONS. It does not mean that a creative theory arrives through two mechanical steps, nor that creativity is reduced to an ordinary gradient update. The first direction is *assimilatory*: observations and models vary while the current conceptual schema is retained. The second is *proto-accommodative*: local probes test how that schema itself might have to be reorganized. These terms adapt Piaget's distinction between interpreting experience through an existing schema and modifying the schema when it no longer suffices [Piaget, 1952, 1985].

Mathematically, each direction will be represented by an involution-algebroid structure, an abstract infinitesimal geometry for admissible local variation. The *double* structure retains both directions and, crucially, their mixed interaction: does assimilating and then probing reorganization agree with probing reorganization and then assimilating? A failure of compatibility does not itself invent a theory. It localizes a conceptual frontier. Before a finite theory extension is proposed, the candidate local repair must persist under the registered probes, compose into a coherent finite realization, and survive less-disruptive explanations: parameter change, stronger search inside the current language, measurement or observer error, and other domain-appropriate controls. Only then may a new object, mechanism, invariant, or law be proposed, transported, and independently admitted. Chapter 2 develops this interpretation, and Chapters 3 and 4 supply its formal geometry.

For brevity, we call the proposed *Double Involution ALgebroid* architecture DIAL. In Figure ,  $\mathbb{D}_{\text{rep}}$  denotes the joint local variation,  $\Theta_{\text{int}}$  its proposed internal interchange defect, and  $\Omega_{\text{cre}}$  the corresponding classical cochain observer when a realization is supplied. Chapter 3

states the formal target and its unresolved bridge obligations.

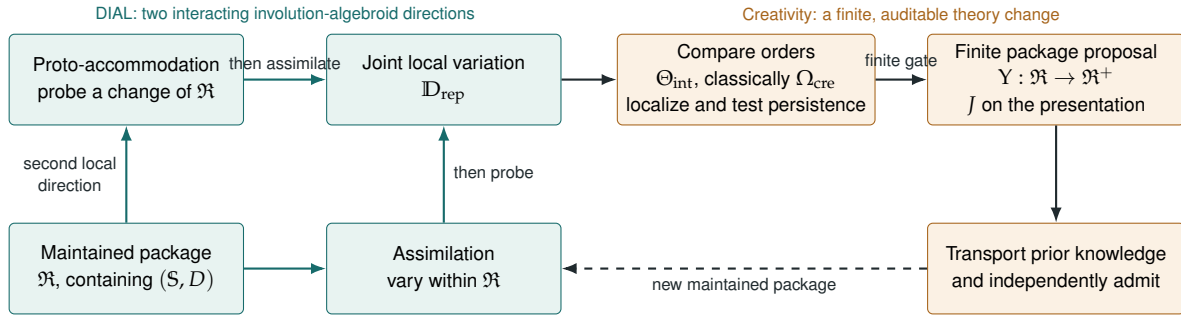


Figure 1: A pedagogical reading of double involution. Assimilation varies a realization inside the maintained representational package; proto-accommodation probes a second local direction. DIAL compares their two orders, localizes any incompatibility, and tests whether the candidate repair persists, has a coherent finite realization, and survives less-disruptive controls. That evidence can motivate, but does not determine, a finite package proposal. When its presentation changes,  $J : S \rightarrow S^+$  is one component of  $Y$ ; transport and independent admission are still required.

THE WORD DOUBLE HAS A USEFUL BIOLOGICAL ANALOGUE. The importance of the DNA double helix lies not merely in having two strands, but in their structured complementarity. The pairing suggested how hereditary information could be copied: each strand could serve as a template for its partner [Watson and Crick, 1953]. Fidelity preserves an organization across generations, while variation and recombination make altered organizations possible. Yet DNA does not create an organism by itself. Replication, expression, development, and selection require a larger cellular and environmental system.

DIAL makes an analogous architectural wager. Assimilation preserves and updates a maintained representational package; proto-accommodation explores how its generative organization might vary. Their double structure records not only the two directions but also their interdependence. But double involution alone is not creativity, just as the double helix alone is not life. Proposal machinery must construct a finite package change, and an admission environment must test whether that change remains coherent, transports prior knowledge, and survives new evidence. Only a change that extends the presentation or generative language supports the stronger claim of theory extension; many valuable repairs leave that language fixed. The central question of this book is whether this doubly connected learning architecture can repeatedly support creative theory construction across otherwise different AI applications.

TWO HISTORICAL EPISODES CLARIFY BOTH THE AMBITION AND ITS LIMITS. The first comes from mathematics. Formulas for equations of degrees two, three, and four encouraged a centuries-long search for a corresponding formula for the general quintic. Ruffini and Abel established the obstruction to a general solution by radicals; Galois supplied a deeper reorganization. Instead of seeking one more expression in the old language, his work related solvability to the structure of permutations of roots—what later developed into Galois theory and

modern group theory [Brunk, n.d.]. The achievement was not merely a more successful search through formulas. It changed the objects and invariants in which the problem could be stated.

This is a paradigmatic finite conceptual jump. A familiar problem generated pressure on an established representational language; a new structural language made the obstruction intelligible and remained useful far beyond the original problem. Nothing in this history, however, establishes that Galois reasoned by double involution. At most it supplies a demanding target for an artificial creativity architecture: can a system recognize that continued search in the current language is inadequate and construct a new algebraic theory with independently checkable consequences?

DARWIN'S DISCOVERY REVEALS A COMPLEMENTARY PATTERN. The diversity, geographical distribution, and fossil history of species demanded an account of change, but a scientifically adequate mechanism was missing. Darwin's notebooks already treated species as mutable. His reading of Malthus in 1838 then supplied a transferable relational pattern: populations can grow faster than limited resources, producing a struggle in which not all individuals survive and reproduce. Applied to heritable biological variation, that pattern became a causal mechanism for natural selection [Darwin Correspondence Project, 2015b,a].

The episode combines two forms of theory construction. Darwin transported a schema from political economy into biology, but the result was not a literal copy of Malthus. Variation, inheritance, differential reproduction, and environmental pressure gave the transported pattern new biological meaning. The creative act was therefore both analogical and transformational: a cross-domain structural correspondence helped produce a new mechanism in the target theory.

Darwin's geological work made a second idea equally important. Small effects need not have small historical consequences. Variation and differential reproduction can be locally slight while their repeated action over deep time produces divergence, adaptation, and extinction. The mechanism in nature is incremental even though the scientific concept that makes it intelligible is a finite reorganization of theory.

The eye made this distinction especially vivid. Paley's earlier watch-maker argument contrasted a stone found on a heath with a watch whose coordinated parts appeared to require a designer [Paley, 1802]. The eye seemed to present the biological version of the watch: what possible use could an incomplete intermediate organ have? Wilberforce's 1860 review pressed related objections concerning missing transitional varieties and the absence of directly observable favorable variations [Wilberforce, 1860]. Darwin addressed the difficulty directly. The relevant standard was not whether "half an eye" performs the work



of a modern eye, but whether a sequence can exist from simple light-sensitive structures toward more complex ones, with every intermediate stage useful to its possessor. Slight inherited variations, accumulated under selection, could then explain apparent design without requiring the finished organ to appear in one leap [Darwin, 1859, Chapter VI].

These examples separate three notions that this book must not collapse. An *infinitesimal process in the world* can accumulate into a macroscopic effect. An investigator can use local variations, failed predictions, and cross-domain comparisons to probe the boundary of a theory. But a *creative theory change* is finite: it introduces a new object, mechanism, invariant, or law and then submits the enlarged theory to independent tests.

#### Boundary

Galois and Darwin are motivating examples, not retrospective validations of DIAL. The book does not claim that either thinker literally used double involution, nor that all creativity is infinitesimal. It asks whether typed local probes and their interactions can help an artificial system decide when to search beyond its current language and guide a finite, auditable theory extension.

With that distinction in view, the technical core uses two interacting infinitesimal geometries. Leung’s Weil theory first supplies tangent-category semantics for local probes, while involution algebroids connect one such tangent direction to classical Lie-algebroid geometry. This book proposes the corresponding double internalization for two interacting classes of repair: assimilatory repair inside a maintained theory and a proto-accommodative probe of how its organization might change. Mackenzie’s classical double Lie algebroids provide the concrete smooth model, including tangent, matched-pair, and cotangent Lie-bialgebroid forms of doubling [Mackenzie, 2000a]. Meinrenken–Pike’s distinct, bigraded cochain Weil algebra then supplies two differentials whose failure to super-commute localizes an incompatibility between the realized repair processes. This classical criterion motivates, but does not by itself establish, the proposed construction in every tangent category or smooth topos. A genuine creative jump remains a finite sketch extension whose transport and consequences must be admitted.

This proposed internal structure is the DIAL architecture introduced above. Its capitalization records the acronym rather than a new spelling of algebroid. DIAL is the concrete mathematical target of the book. Its proposed axioms and functorial semantics must recover the relevant classical double-Lie structure under stated realization hypotheses; its computational value must be established separately by comparisons

showing whether two-direction diagnostics improve theory repair.

## Challenge

Can DIAL turn two interacting modes of conceptual variation into an internal, transportable diagnostic whose realizations improve finite theory extension over single-direction tangent repair and equally resourced search?

We use *synthetic creativity* for this broader research program and *infinitesimal creativity* for its governing architecture. Its working maxim is:

- probe in two directions  $\longrightarrow$  localize their incompatibility
- $\longrightarrow$  test persistence and finite realization
- $\longrightarrow$  extend and admit finitely.

Infinitesimal probes characterize the boundary of the current conceptual space. They do not imply that a creative leap is infinitesimal. The leap is an explicit theory transformation, and its value must be established by proof, simulation, prediction, experiment, or sustained reuse.

The book is organized around three questions. Part I asks what a creative theory change is and develops its categorical and differential language. Part II asks how that language becomes an algorithm whose target, proposal, and admission obligations remain typed. Part III asks what the resulting discipline reveals in mathematics, simulation, scientific literature, and visual generation—distinguishing structural repair from artistic theory extension—then separates the evidence established by those testbeds from the open domains and obligations that remain. Margaret Boden’s combinational, exploratory, and transformational modes cut across all three parts rather than assigning one mode to each application.

The trail map that follows gives conceptual, algorithmic, and testbed routes through this structure. It is deliberately navigational: the chapters retain the responsibility for defining their objects, evidence, and claim boundaries.

The ambition is deliberately larger than the results reported here. The purpose of this book is to turn a philosophical question into a sequence of typed, falsifiable challenge problems.

# *A Trail Map of Infinitesimal Creativity*

Readers new to categories, sketches, tangent categories, or Weil probes should begin with Chapter 0. The book can then be read along three routes.

The routes differ in entry point, not in destination. The conceptual route supplies the meanings of the objects, the algorithmic route turns those meanings into executable contracts, and the testbed route asks what evidence can warrant the resulting constructions. A reader may follow one route first, but Part II assumes the working vocabulary of the early foundational chapters, and Part III assumes the target-card discipline introduced in Chapter 10.

*Conceptual route.* Chapter 1 explains how instruction, imitation, and apprenticeship supply a provisional sketch. Chapter 2 states the theory-construction target; Chapters 3 and 4 develop its local geometry and executable skills; Chapter 5 specifies the finite theory artifact; and Chapter 6 locates that artifact in the wider creativity literature. Chapters 7–9 then distinguish the three creative modes using the common language established by those layers.

*Algorithmic route.* Chapter 10 turns DIAL into the DIAL-X family by declaring domain targets, side operations, mixed witnesses, creative mandates, finite-realization tests, less-disruptive controls, package comparisons, admission contracts, and persistent records. Chapter 11 develops *Causal Learning with Infinitesimal Creativity* (CLIC), Chapter 12 develops *Optimizing Programs Through Infinitesimal Creativity* (OPTIC), and Chapter 13 develops *Reinforcement Learning with Infinitesimal Creativity* (RELIC). Chapter 14 composes the three into *Acquiring Generalizable Novel Theories with Infinitesimal Creativity* (AGENTIC), a proposed meta-creative workflow architecture.

*Testbed route.* Chapters 15–17 use formal certificates, executable simulators, and source-bearing scientific corpora. Chapter 18 develops *Acquiring Representational Theories through Infinitesimal Semantic Transport in Image Creation* (ARTISTIC), a structural-assurance regime for generated images; Chapter 19 develops *Double Involution Learning for Artistic Theory Extension* (DILATE), an open-ended regime for

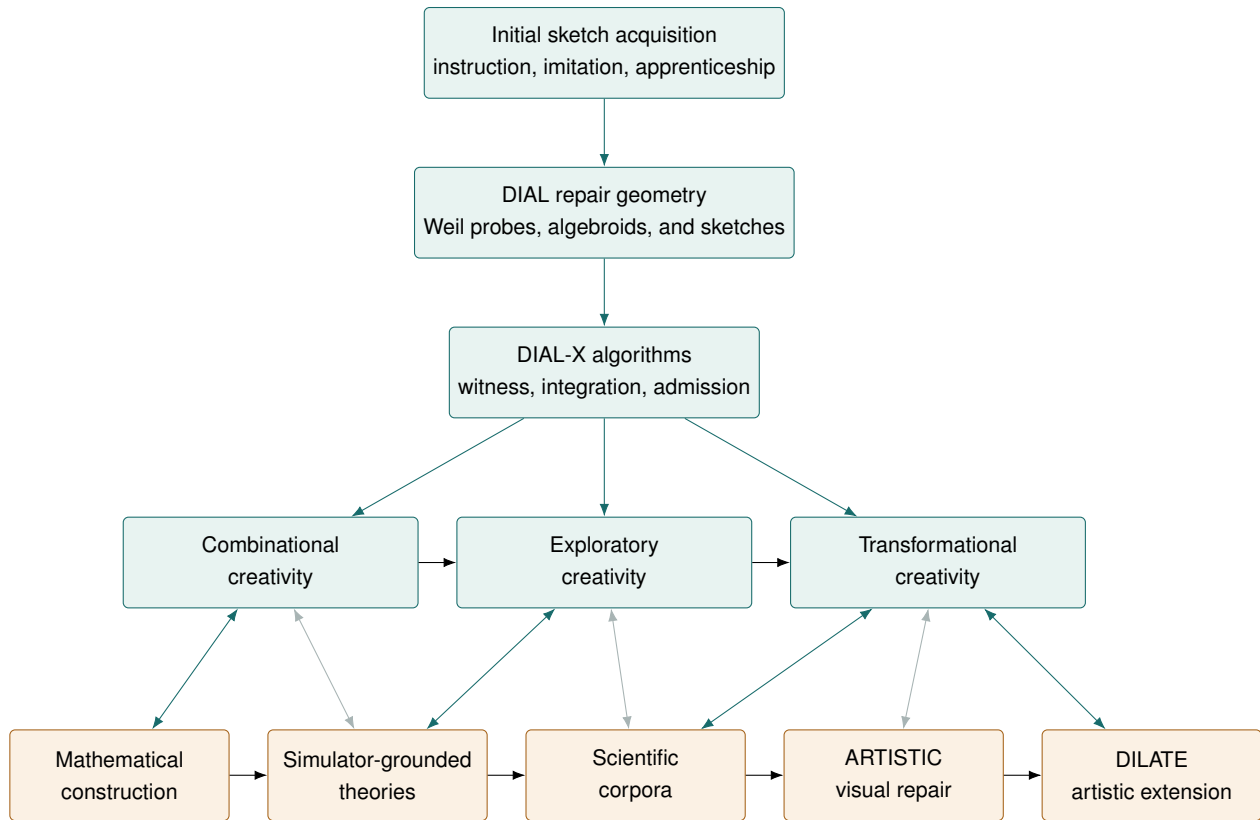
reusable artistic operations. Four media thus yield five admission regimes. They are complementary rather than progressively weaker versions of one ground truth. Chapter 20 separates the evidence established by them from the broader program's open obligations.

The routes are intentionally asymmetric in length. Chapters 7–9 isolate three conceptual distinctions; Chapters 11–14 isolate domain mechanisms and their interfaces; several testbed chapters preserve longer evidence ladders, including negative rungs, because the sequence of failed and repaired contracts is part of the result. Page count should therefore not be read as a ranking of maturity or importance. The evidential status stated inside each chapter—exact, estimated, constructed, admitted, rejected, or conjectural—is the relevant guide.

#### Design principle

Use the matrix as a discipline, not a taxonomy of model personalities. A single creative episode may combine registered structures, explore their consequences, encounter an obstruction, and then transform the theory.

The vertical and diagonal arrows matter. No testbed is assigned to only one mode. Every experiment is located by both a mode and a testbed.





## **Part I**

# **Foundations of Infinitesimal Creativity**





O

## *A Working Language for Creative Theory Change*

Mathematics does not progress only by proving more theorems inside a fixed language. It also invents representations in which previously difficult questions can be composed, varied, and generalized. Euclidean geometry organized points, lines, circles, constructions, and proofs synthetically. Cartesian geometry connected those objects to numbers and polynomial equations. The geometric world did not disappear; a new interface made algebraic operations available to it [Euclid, 1956, Descartes, 1954].

That transition is a useful miniature of infinitesimal creativity. Coordinates combine two conceptual systems. They open an exploratory space of symbolic calculation. At a broader historical resolution, they transform the accepted language of geometry and prepare the way for calculus, differential geometry, and later coordinate-free abstraction. The purpose of this chapter is to introduce the categorical and differential vocabulary needed to describe such changes precisely.

This is a working-language chapter rather than a compressed substitute for the theory that follows. It introduces each construction through the question it lets us ask: categories type composition, sketches declare a theory, functorial models interpret it, sheaves test contextual agreement, tangent structure describes local possibility, and double structure compares two classes of repair. Chapters 2–5 supply the formal development and delimit which parts are established mathematics and which belong to the proposed *Double Involution ALgebroid* (DIAL) research program.

**Design principle**

The mathematical language in this book separates four phases. *Declare* the current theory, its models, and its compositional promises. *Diagnose* failure with typed observers and interacting local probes. *Construct* a coherent finite proposal only after testing less-disruptive explanations. *Admit* the proposal through transport and evidence independent of its construction. Creativity can enter at several phases, but diagnosis, proposal, and admission should not be collapsed into one event.

*0.1 Progressive abstraction as creative reconstruction*

Abstraction is sometimes described as deleting detail. In mathematics it is often a constructive act: identify what is invariant, assign it a new type, and define operations that make the invariant reusable. A coordinate system, a group, a vector space, a manifold, and a category all forget some features while making other relations explicit.

Consider a circle and a line. A synthetic treatment reasons through incidence, congruence, construction, and previously proved propositions. After choosing coordinates, the same configuration can be represented by

$$x^2 + y^2 = r^2, \quad y = mx + b.$$

Intersection becomes substitution and elimination. Parameters  $m, b, r$  can be varied systematically. Degeneracies can be detected algebraically. The coordinate interface changes the available operations without making the original geometric meaning irrelevant.

Three lessons recur throughout this book.

1. A representation is productive when it preserves the relations that matter while enabling new constructions.
2. A translation between domains is not merely a relabeling; it must state how composition is transported.
3. A new language creates new blind spots as well as new powers. Coordinate formulas may hide invariance under a change of coordinates, motivating a later coordinate-free reconstruction.

The history is therefore not a simple ladder from concrete to abstract. It is an alternation of representation, exploration, obstruction, and reconstruction. Category theory is useful because it describes these translations and invariants without requiring every theory to use the same underlying objects.

### 0.2 Categories: typed composition

**Definition 0.1** (Category). A category  $\mathcal{C}$  consists of objects  $X, Y, Z, \dots$ , morphisms  $f : X \rightarrow Y$ , an identity  $1_X : X \rightarrow X$  for each object, and a composite

$$g \circ f : X \rightarrow Z$$

whenever  $f : X \rightarrow Y$  and  $g : Y \rightarrow Z$ . Composition is associative and identities are neutral:

$$h \circ (g \circ f) = (h \circ g) \circ f, \quad f \circ 1_X = f = 1_Y \circ f.$$

The definition is deliberately silent about what objects contain. In *Set*, objects are sets and morphisms are functions. In a category of vector spaces they are vector spaces and linear maps. In a path category, objects are states and morphisms are composable paths. In a category of theories, objects may be presentations and morphisms may be admissible translations or extensions.

THE FIRST READING HABIT IS TYPE FIRST. Before evaluating a formula or a generated proposal, identify the source and target of every map. A composition is meaningful only when adjacent types agree. This catches a basic form of creative nonsense: combining ideas whose verbal similarity hides incompatible roles.

Categories do not assert that everything composes. They record which compositions exist. Nor does associativity say that a composite is true, useful, or executable. It says that a legal chain has a coherent parenthesization. Semantic interpretation and admission supply the remaining obligations.

### 0.3 Diagrams and compositional promises

A directed graph records generating objects and arrows. Its path category adds identities and every finite composite. A *diagram* of shape  $J$  in  $\mathcal{C}$  is a functor  $D : J \rightarrow \mathcal{C}$ ; informally,  $J$  is a wiring pattern and  $D$  supplies an interpretation of that pattern.

Two paths are parallel when they have the same source and target. A square

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ u \downarrow & & \downarrow v \\ X' & \xrightarrow{f'} & Y' \end{array}$$

commutes when

$$v \circ f = f' \circ u.$$

The equation can express preservation under a change of coordinates, agreement between two calculations, compatibility of local descriptions, or transport of an experiment between models. The diagram says which routes are promised to agree. A model can fail that promise.

**Definition 0.2** (Compositional obstruction). Relative to a declared diagram, a compositional obstruction is the typed failure of two required parallel routes to agree, of a required factorization to exist, or of a designated universal property to be realized. An observer may turn the obstruction into vector-, relation-, distribution-, language-, or scalar-valued evidence, but that evidence is not the obstruction’s type-independent definition.

This distinction is central. The equation  $\|vf - f'u\|^2$  is meaningful only after the codomain has enough linear or metric structure. The commutative square is meaningful in any category. Synthetic creativity begins with the structural promise and then chooses observers suited to the application.

This also marks the progression across the three books. Diagrammatic Backpropagation in *Categories for AGI* introduced diagrammatic curvature as an optimizable scalar penalty. LINC’s subsequently separated the typed obstruction from any chosen scalarization. The present book inherits that separation and asks the additional question of when persistent evidence warrants a finite extension of the declaration itself [Mahadevan, 2026b,e].

#### 0.4 Functors and natural transformations

**Definition 0.3** (Functor). A functor  $F : \mathcal{C} \rightarrow \mathcal{D}$  assigns objects to objects and morphisms to morphisms while preserving identities and composition:

$$F(1_X) = 1_{F(X)}, \quad F(g \circ f) = F(g) \circ F(f).$$

A functor is therefore a compositional translation. Cartesian coordinates can be viewed schematically in this spirit: geometric constructions receive algebraic representations, and a sequence of constructions must correspond to the associated sequence of algebraic operations. Establishing a literal functor requires specifying the source and target categories; the historical analogy alone does not do that work.

Functors also separate a formal theory from a realization. A schema or sketch specifies roles and legal routes; a functor interprets them as sets, smooth spaces, programs, database tables, or simulator components. The same theory can therefore have many models.

**Definition 0.4** (Natural transformation). Given functors  $F, G : \mathcal{C} \rightarrow \mathcal{D}$ , a natural transformation  $\eta : F \Rightarrow G$  assigns a map  $\eta_X : F(X) \rightarrow G(X)$

to every object so that for each  $f : X \rightarrow Y$ ,

$$\begin{array}{ccc} F(X) & \xrightarrow{F(f)} & F(Y) \\ \eta_X \downarrow & & \downarrow \eta_Y \\ G(X) & \xrightarrow{G(f)} & G(Y) \end{array}$$

commutes.

Naturality says that a family of local changes is coherent with the structure connecting its components. This is one model of a compositional repair. A system whose modules are independently modified may improve local scores while destroying the routes that gave those modules meaning.

#### Boundary

A metaphorical correspondence between two fields is not automatically a functor, and a family of edits is not automatically a natural transformation. The objects, maps, and preservation equations must be stated before those terms carry mathematical force.

### 0.5 Universal constructions: defining by relations

Many important objects are characterized not by an implementation but by how all other objects map to or from them. Such a *universal property* specifies a construction up to canonical isomorphism.

A product  $X \times Y$  supports paired observations. Given maps  $f : Z \rightarrow X$  and  $g : Z \rightarrow Y$ , there is a unique pairing  $\langle f, g \rangle : Z \rightarrow X \times Y$  with the expected projections. A pullback

$$\begin{array}{ccc} X \times_Z Y & \longrightarrow & Y \\ \downarrow & & \downarrow g \\ X & \xrightarrow{f} & Z \end{array}$$

represents pairs that agree over a shared target. A pushout reverses this pattern and glues objects along a common interface.

These constructions already have creative readings:

*Product.* Place several probes or descriptions in a common context.

*Pullback.* Discover which candidates satisfy two compatible views.

*Pushout.* Combine theories or modules through a registered interface.

*Equalizer.* Identify the locus on which parallel accounts agree.

*Coequalizer.* Quotient distinctions declared irrelevant or redundant.

An optimizer may approximate one of these constructions, but low loss does not state its universal property. The declaration and its numerical observer remain different levels.

## 0.6 Algebraic theories, sketches, and models

A signature lists operation symbols. An algebraic theory also records how operations compose and which equations they satisfy. Lawvere expressed a one-sorted algebraic theory as a category  $\mathbb{T}$  with finite products, generated by a distinguished object and its finite powers [Lawvere, 1963]. A model in a category  $\mathcal{C}$  with finite products is a product-preserving functor

$$M : \mathbb{T} \longrightarrow \mathcal{C}.$$

For the theory of monoids, the presentation contains multiplication  $\mu : X \times X \rightarrow X$ , a unit  $e : 1 \rightarrow X$ , and the associative and unit equations. A model in  $\mathbf{Set}$  is an ordinary monoid. A model in another suitable category interprets the same formal theory there. Theory and model must not be conflated: a parameter vector is one realization, not the compositional specification shared across realizations.

A completed theory can be large. A *sketch* presents one compactly by a graph of generating objects and arrows, declared commutative diagrams, and designated cones or cocones [Ehresmann, 1968, Barr and Wells, 1999]. A model maps the presentation to a semantic category and preserves the specified structure.

**Definition 0.5** (Candidate presentation extension). A candidate presentation extension is a typed map of presentations

$$J : \mathbb{S} \longrightarrow \mathbb{S}^+$$

together with an interpretation of the new generators and relations. The map records how the old language is intended to transport into the proposed one. It is the presentation-changing component of a candidate package comparison, not by itself an admitted theory extension.

Restriction along  $J$  sends an expanded model back to its old visible part:

$$J^* : \mathbf{Mod}(\mathbb{S}^+, \mathcal{C}) \longrightarrow \mathbf{Mod}(\mathbb{S}, \mathcal{C}).$$

The fiber over an old model records its possible expansions. A unique expansion may merely name structure that was already determined. Several inequivalent expansions expose a genuinely new choice. No expansion means that the new theory rejects or must reinterpret that old model.

This gives the book a concrete target for transformational creativity: propose not only a sentence, but an augmented presentation whose models, transport, and new consequences can be examined. Chapter 2 adds the finite-realization and independent-admission conditions under which such a proposal earns an unqualified theory-extension status.

### 0.7 *Lexical creativity: extending a language*

Human language makes theory extension unusually concrete. Speakers continually introduce new expressions and new senses for old expressions. Some disappear; others acquire stable grammatical roles, meanings, and relations to the rest of a language. The Oxford English Dictionary records this continuing process in quarterly updates that add words and senses and revise existing entries from historical evidence [Oxford University Press, 2026]. Lexicographical inclusion is recognition of an innovation that has acquired sufficient evidence of use, not the act that originally created it.

This exposes an important difference between producing an unfamiliar string and changing a language. A conventional language model normally has a fixed tokenizer and a fixed set of learned parameters. Subword tokenization allows it to assemble known pieces into rare or previously unseen surface forms [Sennrich et al., 2016]. It can therefore spell a plausible word that did not occur in its training corpus. But the generation alone does not add a new token to the model, assign the expression a persistent grammatical and semantic type, or revise a shared lexicon for later speakers.

It is useful to distinguish three achievements.

*Novel string generation.* Produce a sequence not previously observed.

This may be combinational novelty over a fixed token vocabulary.

*Conceptual innovation.* Use an expression to mark a useful distinction, operation, object, or social practice that the prior vocabulary did not mark adequately.

*Lexical accommodation.* Modify a persistent language so that the new word or sense becomes a reusable compositional resource for future speakers.

The final achievement can be represented as a declaration extension

$$J_{\text{lex}} : S_{\text{lex}} \longrightarrow S_{\text{lex}}^+.$$

Here the extended sketch does not merely contain another character string. It records a new lexical generator—or a new sense of an existing one—together with relevant syntactic behavior, semantic relations, and

conventions of use. Restriction along  $J_{\text{lex}}$  forgets the innovation and recovers the earlier lexicon. Transport in the other direction asks how old utterances and distinctions survive after the extension.

Human linguistic communities perform proposal, exploration, and admission at different scales. An individual or group coins an expression. Speakers test it across contexts, alter its meaning, and sometimes derive related forms. Sustained public use supplies evidence that the expression has become part of a shared language. In Piagetian terms, a novel string may be assimilated by existing grammar, whereas a genuinely new conceptual distinction may require accommodation of the lexicon. In Boden's terms, the episode can begin as combination, become exploratory as uses proliferate, and become transformational when it changes what speakers can concisely express and reason about.

#### Design principle

Do not identify linguistic creativity with out-of-vocabulary generation. The synthetic target is a persistent extension of the language: a new expression or sense with a type, compositional role, relations to prior expressions, and an admission history.

This example sharpens the book's contrast with a compression-centered account. Predictive compression may induce a useful latent representation of an emerging word. Synthetic creativity asks for the additional transition by which that distinction becomes an explicit, transportable component of the language in which later thought is conducted.

### 0.8 *Presheaves, sheaves, and contextual agreement*

Some theories are observed only through overlapping contexts. A presheaf assigns information  $F(U)$  to each context  $U$  and restriction maps

$$\rho_V^U : F(U) \longrightarrow F(V), \quad V \subseteq U,$$

that compose coherently. A sheaf adds a local-to-global principle: compatible local sections on a cover glue to a unique global section.

Crossword puzzles give a simple intuition. Each clue determines a set of candidate words. Intersections of across and down entries are overlaps. Candidates are locally compatible when their shared letters agree. A complete solution is a global section. Local agreement need not guarantee that a global solution exists unless the relevant gluing conditions are met.

In scientific creativity, contexts may be experiments, papers, simulators, scales, or coordinate charts. A failure to glue can reveal an



inconsistent claim, a missing latent variable, or an inadequate global theory. Sheaf language keeps localization visible: the system can say where agreement holds without pretending that it holds everywhere [Mac Lane and Moerdijk, 1992].

### 0.9 Tangent categories and local possibility

Calculus describes local variation on numerical spaces. A tangent category abstracts the compositional structure required to speak of tangent objects and maps even when ordinary coordinates are not primary [Cockett and Cruttwell, 2014]. At minimum, the reader should retain the following data.

The intended picture is simpler than the axioms. An object  $X$  represents a space of current possibilities;  $TX$  represents a possibility together with an admissible first-order way of varying it. A map  $f : X \rightarrow Y$  transports both the current possibility and its local variation through  $Tf : TX \rightarrow TY$ . The additional structure below makes zero variation, addition, lifting, and the interchange of two successive variations compositional. Chapter 3 states the formal hypotheses; here the purpose is to make local experiments available outside ordinary coordinate spaces.

There is an endofunctor

$$T : \mathcal{M} \longrightarrow \mathcal{M}$$

sending  $X$  to a tangent object  $TX$ , with a projection  $p_X : TX \rightarrow X$ , a zero section, addition of tangent vectors over the same base point, a vertical lift, and a canonical flip on  $T^2X$ . These maps obey coherence and universality axioms. The chain rule appears as functoriality:

$$T(g \circ f) = Tg \circ Tf.$$

The tangent category is not one search space. It is an ambient universe in which selected objects possess a declared notion of local variation. A tangent direction might vary a coefficient, a simulator mechanism, an observer, a skill, or a local causal field. The application must supply that meaning.

For the Cartesian circle

$$q(x, y) = x^2 + y^2 - r^2 = 0,$$

a first-order displacement  $(\dot{x}, \dot{y})$  tangent to the circle satisfies

$$Dq_{(x,y)}(\dot{x}, \dot{y}) = 2x\dot{x} + 2y\dot{y} = 0.$$

The equation does not replace the circle. It probes which local variations preserve its defining relation. This pattern generalizes: declare a structure, differentiate its obligations, and use the resulting failure to localize where the current model is unstable.

### 0.10 Weil algebras as probe shapes

The elementary first-order infinitesimal algebra is the dual-number algebra

$$W = \mathbb{N}[\varepsilon]/(\varepsilon^2).$$

Here  $\mathbb{N}$  is the initial commutative rig used in Leung’s algebraic classification. This object should not be confused with the familiar smooth dual numbers  $\mathbb{R}[\varepsilon]/(\varepsilon^2)$ , or with the Kock–Lawvere infinitesimal object  $D = \{d \in R \mid d^2 = 0\}$  internal to a smooth topos. The constructions are related through their first-order role, but they live at different semantic levels: one presents a universal probe shape, while the others realize infinitesimals in a chosen smooth setting [Cockett and Cruttwell, 2014]. The nilpotent generator retains first-order displacement while suppressing higher powers. More elaborate Weil algebras encode several directions and relations among them. In Leung’s formulation, an appropriate monoidal category  $\text{Weil}_1^{\text{tan}}$  acts through a strong monoidal functor

$$F : \text{Weil}_1^{\text{tan}} \longrightarrow \text{End}(\mathcal{M}),$$

with the generating algebra realized as the tangent functor [Leung, 2017, 2018].

For a first reading, one may treat  $\varepsilon$  as a tagged perturbation: evaluate what changes linearly when the tag is present, while deliberately ignoring effects that require two or more copies of the same tag. The categorical construction makes that probing discipline compositional.

The important intuition is experimental. A Weil algebra is not the theory of the domain; it is a *probe shape*. First-order probes test sensitive directions. Paired probes test interaction. Iterated probes can expose order-effects or curvature-like behavior. A finite implementation selects a probe family rather than observing the entire infinitesimal structure.

**Boundary: Local diagnosis and the two Weil constructions**

Local diagnosis does not uniquely construct a global theory. A tangent direction may fail to integrate, and many finite theory extensions can share the same local profile. Infinitesimal probes constrain a creative proposal; they do not eliminate the abductive jump.

Two different constructions below carry the name *Weil algebra*. Leung's  $\text{Weil}_1^{\text{tan}}$  is a monoidal theory classifying tangent structure. Meinrenken–Pike's  $\mathcal{W}_{\text{Lie}}^{\bullet,\bullet}(\mathbb{D})$  is a bigraded cochain algebra attached to a classical double vector bundle. They must not be identified. Involution algebroids provide an established first-order bridge: every ordinary Lie algebroid induces an involution algebroid, while an involution algebroid induces a Lie bracket on sections (and, given a suitable unit object, the Leibniz law) [Burke and MacAdam, 2019, Sections 4.2.1–4.2.4]; the corresponding double bridge is a construction proposed in Chapter 3. That chapter formulates the candidate DIAL object as a research target and states the classical realizations and coherence conditions that any completed construction must recover.

The internal involution is more than a change of notation for a bracket. Burke and MacAdam impose a Yang–Baxter-style flip law on the prolongation of an anchored bundle; a Lie bracket satisfying Jacobi is then induced on its sections. This is distinct from Drinfel'd's classical Yang–Baxter equation for  $r$ -matrices and Poisson–Lie geometry, although both express a three-way coherence of local composition [Drinfel'd, 1983, Burke and MacAdam, 2019]. In the proposed double theory, each DIAL side must satisfy this internal law before the compatibility between the two sides is tested.

### 0.11 Classical double Lie algebroids: a concrete realization

A single tangent direction describes one kind of local variation. Creativity, however, often couples two qualitatively different processes: repair inside a maintained organization and a probe of how that organization itself might have to change. A double vector bundle provides the elementary square in which two such directions can coexist:

$$\begin{array}{ccc} \mathbb{D}_{\text{rep}} & \longrightarrow & B_{\text{acc}} \\ \downarrow & & \downarrow \\ A_{\text{asm}} & \longrightarrow & M. \end{array}$$

Read the square from the bottom upward. A point of  $M$  is a maintained theory state. The two side bundles list two different kinds of locally available move at that state. The object at the upper left records a compatible pair of moves and any mixed information visible only when both are considered. The first question is therefore not “what does the algebroid optimize?” but “does the result depend on the order in which the two kinds of move are applied?” The bracket and interchange constructions below type that order effect.

Here  $M$  is a space of maintained theory states. We interpret  $A_{\text{asm}}$  as locally available *assimilatory* repairs and  $B_{\text{acc}}$  as *proto-accommodative* directions that test a possible reorganization. The labels are an application-level semantics, not part of the abstract definition. The total space  $\mathbb{D}_{\text{rep}}$  records their joint variation. Its core  $K$  contains mixed directions that project trivially to both sides and can therefore encode interaction data invisible in either repair class alone.

In the classical category of smooth manifolds, Mackenzie’s definition equips a double vector bundle with compatible Lie-algebroid structures whose compatibility is expressed through double-vector-bundle duality [Mackenzie, 2000a]. This is substantially stronger than placing two unrelated brackets on a square. When these conditions hold, the square is a *double Lie algebroid*. Meinrenken and Pike associate to it a bigraded cochain Weil algebra

$$\mathcal{W}_{\text{Lie}}^{\bullet,\bullet}(\mathbb{D}_{\text{rep}})$$

with horizontal and vertical differentials of bidegrees  $(1,0)$  and  $(0,1)$  [Meinrenken and Pike, 2021, Pike, 2020]. Their classical Weil-algebra criterion states that the two VB-algebroid structures form a double Lie algebroid exactly when these differentials supercommute [Meinrenken and Pike, 2021, Theorem II]. For the creativity interpretation, write these differentials as  $d_{\text{asm}}$  and  $d_{\text{acc}}$ .

Their graded commutator defines the *creative interchange obstruction*

$$\Omega_{\text{cre}} := [d_{\text{asm}}, d_{\text{acc}}]_{\text{gr}} = d_{\text{asm}}d_{\text{acc}} + d_{\text{acc}}d_{\text{asm}}.$$

A nonzero value says that “assimilate, then probe accommodation” disagrees with “probe accommodation, then assimilate.” Given a representation and a declared localization observer, it can localize an order effect between repair processes. It is evidence that their current coupling is not a compatible double structure; it is not yet the content of a replacement theory. Internally, the proposed tangent-categorical double construction has an interchange defect  $\Theta_{\text{int}}$ ;  $\Omega_{\text{cre}}$  is its classical cochain realization when the required bridge exists.

Chapter 3 develops the canonical tangent- prolongation example, the vacant matched-pair case, and the splitting-dependent description by 2-term representations up to homotopy. For the primer, the essential

point is only that double compatibility contains strictly more data than two unrelated local update rules.

#### Boundary

A proto-accommodative direction is not an admitted accommodation. The classical double Lie algebroid realizes local interaction and its obstruction; the proposed internal object records the corresponding tangent-categorical comparison. A presentation-level theory change is recorded separately by a finite sketch map

$$J : \mathbb{S} \longrightarrow \mathbb{S}^+,$$

together with transport of old models and an independent admission record. More general accommodation is recorded by an admitted package comparison; when the presentation changes, the sketch map states that language-level component of the jump. The infinitesimal geometry diagnoses pressure for either change.

For later use, call

$$\mathfrak{R}_{\text{dbl}} = (\mathfrak{R}_{\text{diff}}, \mathbb{D}_{\text{rep}}, \Theta_{\text{int}}, \text{Real}, d_{\text{asm}}, d_{\text{acc}})$$

a *double repair package*. It adds interacting repair geometry to the differential representational package without replacing its observers, generators, or admission procedures. The entries following the semicolon belong to the classical cochain realization and are present only when that realization has been supplied.

#### 0.12 Two axes: Boden and Piaget

The book combines two classifications that answer different questions. Boden distinguishes how a conceptual space is used:

*Combinational creativity* connects registered objects or domains.

*Exploratory creativity* searches consequences generated by fixed rules.

*Transformational creativity* changes the rules generating the space.

Piaget distinguishes how experience is related to a maintained schema. *Assimilation* interprets or repairs within the schema; *accommodation* revises the schema itself. These axes correlate but are not identical. A combination may be assimilatory when its interface already exists, or proto-accommodative when a new mediating type must be proposed. A long exploration can remain assimilatory. A small declaration edit can be transformational.

To keep the levels explicit, later chapters use a representational package

$$\mathfrak{R} = (\mathbb{S}, \mathfrak{D}, \Lambda, \mathcal{G}, \mathcal{A}),$$

where  $\mathbb{S}$  is the presentation,  $\mathfrak{D}$  the preservation doctrine,  $\Lambda$  the observers,  $\mathcal{G}$  the generative operations, and  $\mathcal{A}$  the admission procedures. Its differential refinement is

$$\mathfrak{R}_{\text{diff}} = (\mathbb{S}, \mathfrak{D}, \mathcal{M}, F, \Lambda, \mathcal{G}, \mathcal{A}).$$

Assimilation keeps the relevant package fixed. Proto-accommodation probes a second local repair direction inside  $\mathfrak{R}_{\text{dbl}}$ . Accommodation changes at least one representational component through an admitted finite map and records transport to the new package.

### 0.13 *The geometry example as a creative trace*

The Euclidean-to-Cartesian example can now be read as a structured trace rather than a slogan about abstraction.

*Prior presentation.* Geometric objects, incidence relations, constructions, and synthetic propositions.

*Combination.* Coordinates provide an interface between geometric objects and algebraic quantities.

*Exploration.* Substitution, elimination, parameter variation, and later differential operations traverse consequences enabled by the interface.

*Accommodation.* The accepted representational package expands to include coordinate models and algebraic transformations.

*Transport.* Synthetic constructions receive coordinate descriptions; geometric interpretation transports selected algebraic results back.

*Obstruction.* Coordinate dependence obscures invariant content and motivates later affine, projective, manifold, and coordinate-free formulations.

*Tangent probe.* Differentiating the equation of a curve exposes its local admissible directions without reconstructing the entire curve.

This is a retrospective schematic analysis, not a claim that the historical development followed a formal algorithm or a single theory map. It shows the kind of record a synthetic-creativity system should attempt to build: prior language, bridge, new operations, transported results, limitations, and later repairs.

**Admission contract**

Do not infer creativity from abstraction alone. An abstraction earns admission when it preserves relevant prior structure, enables independently valuable constructions or predictions, and remains interpretable across more than the example that suggested it.

*0.14 How to read the rest of the book*

Each subsequent chapter can be approached through four phases. The detailed questions belong inside these phases rather than forming a long checklist of independent machinery.

*Declare.* What is the current sketch, what category contains its models, which operations generate its conceptual space, and which diagrams or universal properties are promised?

*Probe and localize.* Which observers make failure detectable, which Weil probes describe local alternatives, and what are the two repair classes? If the internal interchange law fails, where is  $\Theta_{\text{int}}$  supported, and—when a classical cochain realization exists—what does  $\Omega_{\text{cre}}$  reveal?

*Control and construct.* Does a candidate local repair persist and compose into a coherent finite realization? Do no-change, parameter, in-language, measurement, observer, and domain-specific alternatives explain it? Is the surviving response assimilatory, proto-accommodative, or a proposed package comparison, and is its creative mode combinational, exploratory, or transformational?

*Transport and admit.* What is preserved, forgotten, or newly predicted, and which evidence independent of proposal construction admits, rejects, or defers the change?

This four-phase lens is a reading aid, not a replacement for the more precise DIAL-X algorithmic contract in Chapter 10.

Chapter 1 first explains how a learner can obtain the provisional package on which these questions operate. Chapter 2 then turns them into a theory- extension package. Chapters 3 and 4 develop local geometry and executable skills. Chapter 5 makes algebraic theories the target artifact. Chapter 6 situates that target among the major traditions in computational creativity, after which Chapters 7–9 develop Boden’s three modes. Chapter 10 turns the double repair geometry into a family of target-specific algorithms before the testbeds and experimental architecture put them under controlled evaluation.

### *Further Reading*

The foundational material is treated more broadly in *Categories for AGI* and more deeply for structural diagnosis and repair in the LINC book [Mahadevan, 2026b,e]. For categories, sketches, and functorial models, see Barr and Wells [1999], Lawvere [1963], and Ehresmann [1968]. Mac Lane and Moerdijk develop sheaf and topos semantics [Mac Lane and Moerdijk, 1992]. Cockett and Cruttwell introduce tangent categories as an abstract differential setting, and Leung relates ordinary tangent structure to monoidal theories of Weil algebras [Cockett and Cruttwell, 2014, Leung, 2017, 2018]. Burke and MacAdam introduce involution algebroids as a tangent-categorical generalization of Lie algebroids, and MacAdam develops their functorial semantics [Burke and MacAdam, 2019, MacAdam, 2023]. Mackenzie develops classical double Lie algebroids and distinguishes their tangent, matched-pair, and cotangent Lie-bialgebroid realizations [Mackenzie, 2000a, 2005]. Meinrenken and Pike characterize double compatibility through the commuting differentials of a distinct bigraded cochain Weil algebra [Meinrenken and Pike, 2021, Pike, 2020]. Gracia-Saz, Jotz Lean, Mackenzie, and Mehta give the equivalent splitting-dependent presentation by matched pairs of 2-term representations up to homotopy [Gracia-Saz et al., 2014].



# 1

## *Initial Sketch Acquisition*

### IMITATION, EXPLANATION, AND LEARNING APPRENTICES

Infinitesimal creativity cannot begin from nowhere. Assimilation presupposes a schema within which experience can be interpreted; accommodation presupposes a schema that can fail and be revised. A creator must therefore begin with some provisional theory of the domain. That theory may be supplied explicitly, reconstructed from demonstrations, acquired through coached practice, inherited from an earlier domain, or assembled from all four.

This chapter makes that initialization process explicit. Its target is not a perfect theory. It is an inspectable first sketch whose objects, arrows, equations, observers, and uncertainties are explicit enough for subsequent learning to begin. The sketch remains corrigible: acquisition establishes a starting declaration, not an immutable ontology.

#### Design principle

The DIAL cycle begins only after a learner possesses a provisional sketch. Initial sketch acquisition constructs that starting point from instruction, demonstration, explanation, interaction, and prior theories. Assimilation and accommodation then test and revise what has been acquired.

#### 1.1 *The initialization problem*

Let a maintained representational package be

$$\mathfrak{R} = (\mathcal{S}, \mathcal{D}, \Lambda, \mathcal{G}, \mathcal{A}),$$

where  $\mathcal{S}$  is a sketch presentation,  $\mathcal{D}$  states which structure its models must preserve,  $\Lambda$  contains observers,  $\mathcal{G}$  contains admissible generative operations, and  $\mathcal{A}$  contains admission procedures. The later DIAL machinery assumes that some version of this package is already available.

Initial sketch acquisition asks where it came from.

Write the acquisition evidence as

$$\mathcal{E} = (\mathcal{I}, \mathcal{D}, \mathcal{F}, \mathcal{X}),$$

with explicit instructions  $\mathcal{I}$ , expert demonstrations  $\mathcal{D}$ , corrective feedback  $\mathcal{F}$ , and the learner's own interventions  $\mathcal{X}$ . An acquisition procedure produces

$$\text{Acquire}(\mathcal{E}, \mathfrak{R}_{\text{prior}}) \rightsquigarrow (\hat{\mathfrak{R}}_0, \hat{D}_0, \Gamma_0, U_0).$$

Here  $\hat{\mathfrak{R}}_0$  is the provisional package,  $\Gamma_0$  records how its declarations are supported by the evidence, and  $U_0$  records unresolved alternatives and unsupported components. The term  $\hat{D}_0$  is a current realization or task state interpreted under the provisional package. It may be assembled from an observed example, an expert trace, or the learner's present environment. The squiggled arrow is deliberate. Acquisition need not be a unique functor determined by the observations; several sketches, and several realizations of each sketch, may explain the same behavior.

Within this framework, the formulation rules out two unhelpful extremes. A theory-free learner has no declared invariants against which to recognize a structural failure. A learner supplied with a complete and correct theory has no acquisition problem at all. The scientifically interesting regime lies between them: partial instruction, finite demonstrations, incomplete observers, and a theory that is useful precisely because it remains revisable.

A useful initial sketch is not necessarily true or complete. It must be explicit enough that later evidence can say where and how it fails.

## 1.2 Four routes into a domain

Four acquisition routes should be distinguished. The first three supply new evidence directly; the fourth reuses a theory acquired elsewhere and therefore enters the acquisition map through  $\mathfrak{R}_{\text{prior}}$ .

*Explicit instruction.* A teacher directly names objects, relations, rules, goals, or invariants. The learner receives part of the declaration in symbolic or linguistic form: a racket strikes a ball; a legal serve begins behind the baseline; quantity is preserved when liquid is poured without loss.

*Imitation.* The learner observes one or more competent performances and reconstructs a task theory from their regularities. It must infer which entities play stable roles, which actions compose, and which apparent details are incidental.

*Apprenticeship.* The learner acts in the task while a teacher selects examples, redirects attention, corrects actions, supplies explanations, and introduces counterexamples. The evidence is interactive and depends on the learner's current theory.

*Theory transport.* The learner imports a typed construction from an earlier domain through an analogy, ontology alignment, or explicit theory map. The imported structure is provisional: its source types, preserved relations, and unsupported target-side assumptions must remain visible.

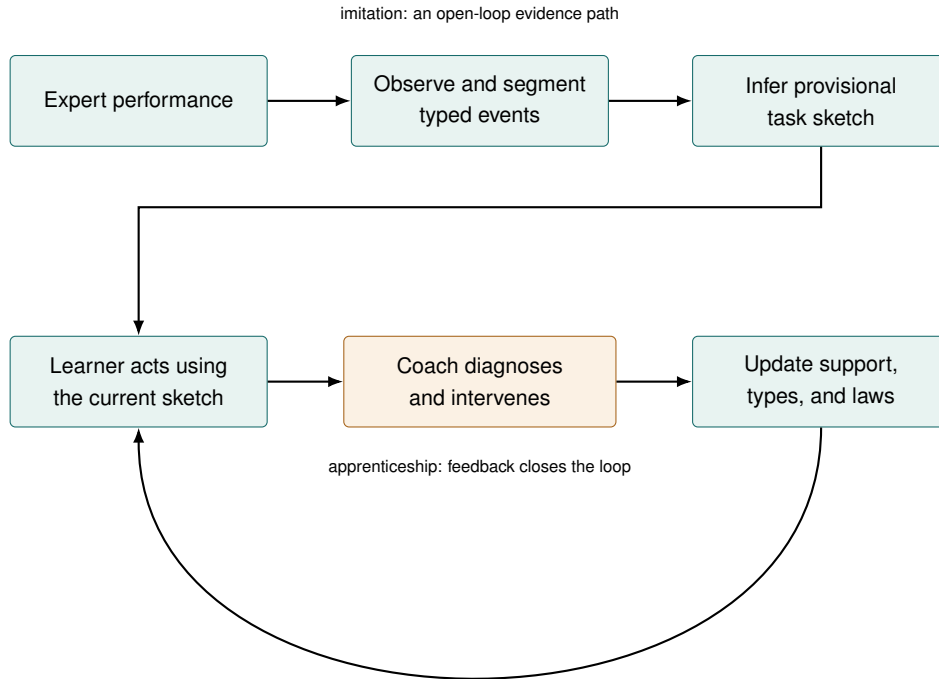
The distinction between imitation and apprenticeship is structural, not merely terminological. Under imitation, the evidence can be represented as a fixed family of expert traces

$$\mathcal{D} = \{\tau_1^E, \dots, \tau_n^E\}.$$

Under apprenticeship, the teacher observes a learner trace  $\tau_t^L$  and returns a context-sensitive intervention

$$f_t = \text{Coach}(\tau_t^L, \hat{\mathcal{R}}_t),$$

which may correct an action, identify a missing precondition, contrast two cases, or amend the learner's explanation. The next evidence therefore depends on the current learner. Apprenticeship is a closed-loop acquisition process.



Modern literature sometimes uses *apprenticeship learning* for learning from demonstrations, including inverse reinforcement learning [Abbeel

Figure 1.1: Imitation reconstructs a task sketch from observed expert traces. Apprenticeship adds an adaptive teacher who diagnoses the learner's own attempts and supplies corrections, contrasts, and explanations. The resulting sketch is an auditable starting theory for later assimilation and accommodation.

and Ng, 2004]. This book uses the narrower distinction above because teacher feedback changes the acquisition problem. Passive traces reveal what an expert did; coaching can reveal why an action was chosen, which latent error the learner made, and what it should attend to next.

### 1.3 *Explanation-based learning as a precedent*

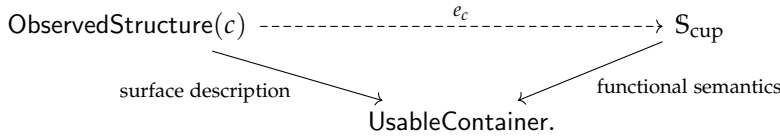
Explanation-based learning (EBL) provides a direct historical precedent for theory-guided sketch acquisition. Rather than estimate a concept from surface correlations across many examples, EBL constructs an explanation of why a particular example satisfies a goal concept under a domain theory. It then generalizes the proof while retaining the dependencies required by the explanation [Mitchell et al., 1986].

The classical cup example begins with a functional specification. A cup is an open vessel that is stable and liftable. Domain knowledge connects these functions to observable structure: a suitable cavity can hold liquid, a flat base can support stability, and a graspable part together with manageable weight can support lifting. From one explained example, the learner can derive a recognition rule that ignores features unused by the proof [Mitchell et al., 1986, Mooney and Bennett, 1986].

In sketch language, let  $S_{\text{cup}}$  contain types such as

Container, Liquid, Opening, Support, Grasp,

and arrows representing containment, filling, lifting, and support. A particular object  $c$  is admitted as a cup when its observed structure factors through the functional declaration:



The explanatory map  $e_c$  records why the example realizes the concept. Color and logo need not appear in the factorization. They can therefore vary without invalidating the explanation.

#### Boundary

One-example generalization is not induction from a sample of size one. Its force comes from a supplied domain theory and an explanation that identifies which properties are proof-relevant. If the theory is wrong, the generalized concept may be wrong in a systematic and highly confident way.

This is both the strength of EBL and the point at which infinitesimal creativity extends it. Classical EBL primarily operationalizes consequences of a supplied theory. DIAL also permits the theory to become an object of diagnosis and revision. A declaration that every cup must possess a handle, for example, fails on handleless drinking vessels. The repair may replace the specific morphological condition with a more general grasping, holding, or drinking affordance. The counterexample changes not merely a parameter but the explanatory vocabulary.

#### 1.4 *Learning apprentices and continuous knowledge acquisition*

The learning-apprentice tradition makes the feedback loop operational. LEAP, developed for VLSI design, was conceived as an interactive knowledge-based assistant that acquired rules by observing the ordinary problem-solving steps of expert users. When a user overrode its advice and manually refined a circuit, the system could treat that intervention as evidence of a rule it should have possessed [Mitchell et al., 1985].

The later account of apprentice-based knowledge acquisition separated at least two learning obligations. LEAP learned feasible problem reductions and also learned control knowledge for choosing among competing reductions. Its explanation-based procedures operated with incomplete domain theories rather than presuming that every useful dependency had already been encoded [Mahadevan et al., 1993]. This distinction anticipates the present separation between acquiring arrows that make an operation possible and acquiring observers or preferences that determine when the operation should be used.

The historical architecture also contains a lesson for current frontier models. Knowledge acquisition need not be isolated in a separate training phase. A useful assistant can learn while participating in the task, provided that it records the teacher's contributions, reconstructs their explanatory role, and preserves uncertainty about what has not yet been established. Language models greatly expand the available interface for explanations and corrections, but fluent dialogue does not by itself produce a coherent task theory. The acquired claims must still be typed, related, and tested.

#### 1.5 *What must be acquired from a demonstration?*

A demonstration is not merely a sequence of states and actions. For sketch acquisition, it supplies evidence about several interacting layers:

1. **Ontology:** persistent objects, agents, tools, locations, and materials.

2. **Typing:** which actions accept which objects and which outcomes they produce.
3. **Affordances:** what an object makes possible in the current task.
4. **Preconditions and effects:** when an action is applicable and what changes after it is executed.
5. **Composition:** which actions form meaningful subroutines and in what order.
6. **Invariants:** what the expert preserves while other quantities vary.
7. **Objectives:** which terminal states or maintained conditions count as success.
8. **Recovery structure:** how the expert recognizes and repairs a failed attempt.

Consider learning tennis. Watching matches can reveal that players alternate striking a ball across a net, that serves begin points, that court regions matter, and that certain action sequences recur. It may not reveal the subtle mechanics of a grip, weight transfer, toss location, or recovery step. A coach can intervene directly: rotate the grip, move behind the baseline, keep the eye on the ball, recover toward an appropriate court position. Such feedback identifies latent variables and localizes errors that are weakly observable from expert performance alone.

The learner should therefore infer not only a policy  $\pi(a \mid s)$ , but a compact inventory for a typed task sketch,

$$S_{\text{task}} = (O, A, E, R, Q),$$

where  $O$  contains object types,  $A$  admissible actions,  $E$  their effects,  $R$  compositional relations and invariants, and  $Q$  observers for quality, legality, and failure. This tuple is a mnemonic for the generators and constraints of a presentation, not an alternative formal definition of a sketch. Policies are models or controllers defined relative to this theory. They are not substitutes for it.

### 1.6 Coaching as typed counter-witness selection

An effective coach does more than label an action incorrect. The coach selects evidence that distinguishes the learner's current theory from a better alternative. This gives coaching a natural counter-witness semantics.

Suppose the learner considers candidate sketches  $S_1, \dots, S_k$ . A pedagogical intervention  $q$  has value when the possible teacher responses

separate these candidates:

$$q^* \in \arg \max_q \text{Discrim}(p(f \mid q, S_1), \dots, p(f \mid q, S_k)).$$

A tennis coach may exaggerate a motion, physically reposition the learner, contrast two grips, or ask for the same serve from a different stance. Each intervention changes the evidence so that a hidden structural distinction becomes observable.

Teacher corrections can target different components of the sketch:

| Correction       | Example                                                   | Sketch effect                                       |
|------------------|-----------------------------------------------------------|-----------------------------------------------------|
| Type correction  | “Hold the racket by the grip, not the frame.”             | Refines an action’s admissible argument type.       |
| Precondition cue | “Begin the serve behind the baseline.”                    | Adds a condition to a legal transition.             |
| Invariant cue    | “Keep the racket face stable through contact.”            | Declares a relation to preserve during a motion.    |
| Segmentation cue | “Recover before preparing the next stroke.”               | Introduces a missing subaction or boundary.         |
| Explanatory cue  | “This position leaves both sides of the court reachable.” | Connects an instruction to its functional role.     |
| Counterexample   | Demonstrate when the usual stroke is inappropriate.       | Restricts a rule or introduces a contextual branch. |

The same correction may update execution and theory simultaneously. The learner can first assimilate it as a local motor adjustment. Repeated failure to transport that adjustment may then reveal a missing type, affordance, or contextual distinction and trigger accommodation.

Table 1.1: A coach can intervene at several distinct locations in an acquired task sketch. Feedback should preserve this type information rather than being collapsed into an undifferentiated scalar reward.

### 1.7 Conservation as developmental sketch acquisition

Piaget and Inhelder’s conservation task illustrates why initial theory construction cannot be reduced to passive perception [Piaget and Inhelder, 1974]. A young child may use the salient rule

$$\text{greater visible height} \implies \text{greater quantity}.$$

Pouring the same water into a taller, narrower vessel then appears to increase the amount. A more adequate sketch distinguishes the changing geometry of the container from a quantity invariant under an admissible pour:

$$Q(x) = Q(\text{pour}_{c_1 \rightarrow c_2}(x)),$$

provided nothing is added, removed, spilled, or transformed.

The acquired theory contains more than the word *volume*. It contains a class of transformations, an observer  $Q$ , and boundary conditions under which  $Q$  is preserved. Reversing the pour supplies a particularly useful intervention because it exposes the transformation as reversible while restoring the earlier appearance. Measurement, social correction, and repeated manipulation can provide further evidence. The conceptual achievement is the construction of an invariant across appearances.

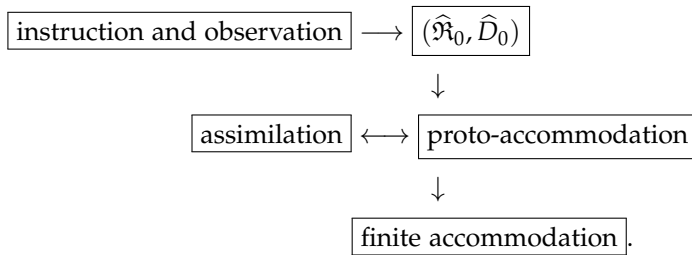
#### Design principle

Instruction can name an invariant; imitation can reveal that competent agents act as if it were preserved; apprenticeship can select transformations that make the invariant observable. The acquired sketch records all three forms of support separately.

This example also exposes an important boundary. The framework does not claim that a child's developmental transition is literally a categorical algorithm or a DIAL computation. It extracts a computational problem: how can a learner move from a perceptually attractive but structurally inadequate rule to a theory organized by transformations and invariants?

### 1.8 From acquisition to double involution

Initial acquisition, assimilation, and accommodation form three different stages, although an extended learning episode may cycle among them:



Acquisition supplies both the declaration and a realized base over which the two DIAL directions are defined. The assimilatory side asks how  $\hat{D}_0$ , its models, and its skills vary while  $\hat{\mathfrak{R}}_0$  is retained. The proto-accommodative side probes local changes to the organization of that package. Their mixed interaction can localize an interchange defect. A package comparison then states the proposed finite change; a sketch map records its presentation-changing component when one exists.

This sequencing prevents a circular explanation. DIAL does not create its first objects and observers merely by differentiating them;



differentiation requires a space already organized well enough to admit tangent directions. Conversely, sketch acquisition does not have to solve creativity before DIAL begins. It need only construct a provisional and revisable space.

The support record  $\Gamma_0$  is essential at this interface. It should distinguish:

- declarations stipulated by a teacher;
- relations inferred from repeated demonstrations;
- functional dependencies supported by explanations;
- corrections observed only in a narrow context;
- hypotheses introduced by the learner;
- unresolved alternatives retained in  $U_0$ .

When a later obstruction appears, this provenance helps decide whether to repair perception, execution, an inferred rule, or a teacher-supplied axiom. Without it, accommodation risks rewriting well-supported structure to cover a local modeling failure.

### 1.8.1 *The readiness boundary*

Acquisition may itself involve many tentative revisions: a learner segments a demonstration differently, retracts an inferred precondition, or asks a teacher for a contrast. These are changes to a candidate representation during bootstrapping. This book reserves *accommodation* for a later event: a registered maintained declaration is found inadequate, a finite change is constructed, and that change is independently admitted. The distinction is procedural rather than metaphysical, but it keeps the evidence ledger honest.

A provisional package is ready to initialize DIAL when it supplies three interfaces:

*A realized base.* At least one current realization  $\hat{D}_0$  is typed by the declaration, so local variation is variation of something explicit rather than movement in an unspecified latent space.

*Observable obligations.* At least one declared relation, invariant, or universal property has an observer with stated semantics. Otherwise the system cannot distinguish structural failure from observer silence.

*A revision boundary.* The maintained components, unresolved alternatives, admissible proposal language, and admission test are recorded. Otherwise a change cannot be classified as assimilation, proposal, or admitted accommodation.

This is a minimum interface, not a completeness claim. A weak initial sketch may omit precisely the concept that later learning must invent. It is usable when its omissions can become diagnosable, not when they have already been eliminated.

### 1.9 *An initial acquisition contract*

An acquired sketch should be evaluated before it is treated as the maintained theory. We use the following contract.

#### **Admission contract**

An initial sketch acquisition claim must declare:

1. the instruction, demonstration, feedback, intervention, and prior-theory channels made available to the learner;
2. the typed objects, arrows, equations, observers, and generative operations acquired from those channels;
3. the provenance and confidence of each acquired component;
4. at least one held-out composition or transfer task not copied directly from a demonstration;
5. counterexamples that test whether salient surface features were mistaken for functional invariants;
6. unsupported alternatives on which the learner must abstain; and
7. the interface through which later assimilation and accommodation may update the sketch.

This contract separates sketch acquisition from behavioral cloning. A learner can reproduce familiar trajectories while possessing no reusable task theory. Conversely, it can acquire a useful structural theory while its motor execution remains inferior to the expert's. The two achievements should be measured separately.

It also separates acquisition from creativity. Faithfully importing a theory supplied by a teacher may solve the initialization problem without constructing anything novel. An acquisition episode supports a creative claim only when it proposes structure absent from the available instruction, demonstrations, feedback, and prior theory, and when the provenance record and held-out tests distinguish that construction from copying or paraphrase. The same episode may therefore contain an acquisitional step and a separately admitted creative step; the audit record should not collapse them.

### 1.10 *A controlled acquisition benchmark*

A controlled benchmark can use a deterministic hand–eye world or an instruction-rich environment such as ALFWorld. An expert solves tasks while the learner receives progressively richer acquisition channels. The sequence is an evidence hierarchy: each condition tests what an additional channel contributes to the acquired sketch.

1. **Trace imitation:** expert state–action sequences only.
2. **Explained imitation:** traces plus subgoals, object roles, and action rationales.
3. **Corrective apprenticeship:** the learner acts and receives localized corrections from a teacher.
4. **Active apprenticeship:** the learner requests demonstrations, contrasts, or explanations that distinguish its remaining candidate sketches.
5. **Acquisition plus DIAL:** the acquired sketch initializes RELIC or another DIAL-X system, which must recognize and repair a withheld structural change.

Evaluation crosses task performance with structural recovery. Relevant measures include type accuracy, precondition and effect recovery, equation and invariant recovery, calibration of unsupported alternatives, teacher interventions required, compositional transfer, and preservation of earlier knowledge after accommodation. A held-out recombination is particularly important: it tests whether the learner acquired a theory that composes rather than a library of trajectories.

The tennis analogy suggests the expected outcome. Watching can supply a broad theory of the game. Coaching should be most valuable where the decisive variables are difficult to infer from expert behavior: grip, body position, timing, attention, and recovery. DIAL becomes relevant when the learner encounters a task variant for which the acquired ontology or skill grammar is insufficient, rather than merely executing an established stroke poorly.

### 1.11 *What this chapter does not assume*

Initial sketch acquisition does not eliminate the observability problem. A teacher may omit crucial knowledge, demonstrations may underdetermine the theory, and the learner’s perceptual vocabulary may be unable to express the relevant distinction. Nor does feedback guarantee truth: teachers can be mistaken, inconsistent, or strategically unhelpful.

The chapter therefore does not posit a privileged channel through which the correct ontology simply enters the machine. It makes the channels explicit, records their provenance, and preserves uncertainty. The result is a provisional theory that can be criticized by later experience.

This makes the initialization problem explicit without assuming it away. The next chapter begins from that provisional, source-bearing package and its current realization. It asks what it would mean for a learner to improve the realization, revise another component of the package, or extend its theory presentation—and how to distinguish each achievement from mere output novelty.

### Challenge

Can an apprentice acquire a compact task sketch from a small number of demonstrations and corrections, use it to explain and recombine expert behavior, and then recognize when a new situation requires accommodation rather than further imitation?

### *Further Reading*

Mitchell, Keller, and Kedar-Cabelli unify explanation-based generalization and illustrate it with the functional cup example [Mitchell et al., 1986]; Mooney and Bennett develop a domain-independent generalizer for proof and action explanations [Mooney and Bennett, 1986]. The LEAP work introduced learning-apprentice systems for continuous acquisition from expert problem solving [Mitchell et al., 1985], and the later synthesis develops explanation-based learning under incomplete domain theories [Mahadevan et al., 1993]. Abbeel and Ng provide the influential inverse-reinforcement-learning formulation of apprenticeship from expert demonstrations [Abbeel and Ng, 2004]. Piaget and Inhelder's conservation studies motivate acquisition of invariants across changing appearances [Piaget and Inhelder, 1974], while Drescher supplies a constructivist computational treatment of schema acquisition and revision [Drescher, 1985, 1991].

## *Creativity as Theory Construction*

A machine may solve a difficult problem without creating the conceptual world in which future problems become easy to state. The distinction is visible in mathematics: proving an isolated theorem is an achievement, but introducing objects such as prime numbers, Fourier series, or Lie groups creates a durable language of objects, operations, laws, examples, and questions. The latter is the phenomenon this book calls *theory construction*.

Chapter 1 supplied the starting point: a provisional sketch, a current realization typed by it, and the provenance and uncertainty of its declarations. The present chapter asks what can happen after that acquisition. Improving a model of the sketch is one achievement; changing the sketch so that new kinds of model, explanation, or experiment become expressible is another. This distinction turns initialization into the first state of a creative process rather than an unexplained precondition.

The source-bearing sketch is an input to that process, not automatically a creative artifact. Creativity begins when the learner constructs and admits a novel composite, consequence, or theory change relative to what its acquisition channels already supplied. Provenance is what makes that relative claim auditable.

Feynman's atomic-hypothesis thought experiment gives the scientific analogue. Its achievement is not that a sentence has a short code. The hypothesis declares a small ontology of particles, motion, and interaction from which a large family of observable consequences follows [Feynman et al., 1963, Section 1–2]. The desired economy is therefore *generative compression*: a compact theory whose compositional structure supports reconstruction, explanation, intervention, and extension.

### **Boundary**

The shortest description of existing data need not be the most creative theory. A synthetic theory earns its economy by exposing reusable objects and laws that generate consequences beyond the observations from which it was constructed.

Chapter 0 supplied the working categorical language for representing such structure. The preceding books develop that language and learning from failures of compositionality in greater depth [Mahadevan, 2026b,e]. We do not rederive those foundations here. Our concern is what happens when repair of the current realization is no longer sufficient. The surviving proposal may revise another component of the maintained package, or it may make the stronger move of changing the objects, arrows, equations, or generative language of the theory presentation.

#### Design principle

Synthetic creativity is the broader program of constructing novel, persistent, and admissible conceptual extensions. Infinitesimal creativity is its governing architecture: use interacting local repairs to diagnose where a maintained package is inadequate, then state the creative jump as a finite, typed package comparison. A change of presentation or generative language is the special case of theory extension, in which new answers, predictions, and experiments become expressible.

A theory is persistent when it can be communicated, reused, criticized, and extended by reasoners other than its originator. Novel output alone does not meet this standard.

### 2.1 Three different achievements

Let a reduced view of a conceptual space be represented by

$$\mathcal{K} = (\mathcal{S}, D, \Lambda, \mathcal{U}),$$

where  $\mathcal{S}$  declares the relevant types and compositional laws,  $D$  is the current realization,  $\Lambda$  is a family of observers, and  $\mathcal{U}$  is the language of admissible moves. This notation is not intended as a complete theory of concepts. It supplies a boundary among three levels of achievement. It is also a deliberately reduced view of the fuller representational package  $\mathfrak{R}$  introduced in Chapter 0:  $\mathcal{K}$  suppresses the preservation doctrine, proposal machinery, and admission procedure when they are not needed for the problem-solving/theory-extension distinction.

**Definition 2.1** (Problem solving). A problem-solving transition changes the realization or its state while preserving the maintained package:

$$(\mathfrak{R}, D) \longrightarrow (\mathfrak{R}, D').$$

The result may be extremely difficult or surprising, but its objects and criteria of success remain available in the inherited theory.

**Definition 2.2** (Candidate package change). A candidate package change proposes a registered comparison

$$(\mathfrak{R}, D) \xrightarrow{Y} (\mathfrak{R}^+, D^+),$$

identifying which presentation, doctrine, semantics, probes, observers, generative operations, or admission procedures change, and how established content is intended to transport. At this stage coherence, transport, and usefulness remain to be tested.

**Definition 2.3** (Candidate theory extension). A candidate theory extension is a candidate package change whose presentation or generative language changes through a registered map, span, or zigzag containing

$$J : \mathbb{S} \longrightarrow \mathbb{S}^+.$$

The presentation component may add, split, identify, or reorganize objects, arrows, equations, or primitive operations. A change only to an observer, estimator, probe family, or admission threshold is a package change but not, by itself, a theory extension.

**Definition 2.4** (Admitted package change). An admitted package change is a candidate package change whose componentwise coherence, claimed transport, and novel consequences have passed a registered test independent of proposal construction.

**Definition 2.5** (Admitted theory extension). An admitted theory extension is an admitted package change whose package comparison contains a non-equivalent presentation or generative-language change  $J$ . An unqualified claim that a system *constructed a theory extension* refers to this status; earlier stages are proposals or candidate extensions.

This refines the primer's presentation-level datum by including a realization and its transport. The displayed map is still the simplest case. An inclusion-like map naturally records added generators or equations, but a revision that forgets, splits, or identifies earlier structure may require a span or zigzag of presentations, together with an explicit account of information lost in each direction. Subsequent uses of  $J$  should therefore be read as registered theory maps, not as a claim that every conceptual change is literally an inclusion.

A mere relabeling or invertible change of coordinates is not a theory extension: it may improve tractability or reveal structure, but it does not by itself enlarge what the theory can express. Nor is an arbitrary incompatible replacement an extension, because it supplies no accountable relation to the earlier theory. Between these cases lies the creative target: a candidate must expose new consequences, state what survives, and earn admission. This requirement turns creativity from unconstrained novelty into a problem of *typed theory surgery*.

Six statuses therefore recur: a model update keeps  $\mathfrak{R}$  fixed; an equivalent presentation changes notation without changing expressivity; a package proposal registers  $\mathbb{Y}$ ; a theory-extension proposal records its presentation-changing component  $J$ ; an admitted package change adds

verified transport and independent evidence; and an admitted theory extension also establishes the claimed non-equivalent language change. An equivalent presentation may be mathematically or computationally valuable, but it is not evidence of theory construction merely because it is unfamiliar.

## 2.2 *Assimilation, accommodation, and cognitive development*

Jean Piaget's developmental epistemology provides a striking psychological precursor to this distinction. Piaget described adaptation through two coupled processes. In *assimilation*, experience is interpreted through an existing cognitive schema. In *accommodation*, the schema itself is modified because the existing organization cannot adequately incorporate the experience. Cognitive development proceeds through repeated attempts to restore an always provisional equilibrium between them [Piaget, 1952, 1985].

In the notation above, assimilation is analogous to changing a realization while retaining its package,

$$(\mathfrak{R}, D) \longrightarrow (\mathfrak{R}, D'),$$

whereas proto-accommodation may propose a package transition

$$(\mathfrak{R}, D) \xrightarrow{Y} (\mathfrak{R}^+, D^+).$$

Independent admission turns that proposal into accommodation. Only when the presentation or generative language changes does the admitted transition also contain a theory map  $J : \mathbb{S} \rightarrow \mathbb{S}^+$  and qualify as a theory extension. The analogy is not an identification of a child's cognition with a categorical machine. Its value is architectural: learning cannot always be understood as adjusting content inside a fixed representational scheme. Sometimes the scheme that determines what counts as an object, an invariant, or an explanation must itself change.

Chapter 1 used conservation to show how an invariant can be acquired across changing appearances. Object permanence gives the complementary example: a theoretical object organizes appearances even while occluded [Piaget, 1954, Piaget and Inhelder, 1974]. Here we retain only their common architectural question: when should a reasoner vary a model of existing objects, and when does persistent failure warrant a new object, invariant, or law?



**Design principle**

Assimilation repairs a model within a maintained conceptual declaration. Proto-accommodation probes how that organization might have to change. Accommodation installs an admitted package change; theory extension is the case that changes its presentation or generative language. A creative system needs all three distinctions, together with evidence separating a genuine need for theory change from an ordinary failure of fit.

This is a conceptual lineage, not a categorical model of child development or a commitment to rigid stages. The computational problem is to make the boundary between the two operations explicit and auditable.

*A computational precedent.* Gary Drescher made this connection to artificial intelligence explicit at the MIT Artificial Intelligence Laboratory. His early thesis introduced the *Schema Mechanism*; the subsequent paper, also issued as an AI Memo, was aptly titled “Genetic AI—Translating Piaget into Lisp” [Drescher, 1985, 1986]. The mechanism begins with sensorimotor primitives and learns predictive schemas organized as context, action, and result. It can refine schemas in response to experience, compose actions, and construct new representational items when its existing vocabulary is inadequate.

Drescher’s 1989 doctoral thesis was later published as *Made-Up Minds: A Constructivist Approach to Artificial Intelligence* [Drescher, 1989, 1991]. Its simulated infant supplied an executable test of whether a schema-building mechanism could reproduce parts of Piagetian sensorimotor development, including steps toward a persistent object concept. In the language of this book, the Schema Mechanism is an early computational architecture in which assimilation updates predictive structure inside an available vocabulary, while accommodation can appear as the construction of a new item or coordinated schema.

The resemblance to the synthetic-creativity program is substantive but incomplete. Drescher supplied an embodied microworld, an explicit action loop, and a mechanism for concept invention—all features absent from purely textual generation. The present program asks how this constructivist approach can be generalized from sensorimotor schemas to typed algebraic theories, active scientific experiments, explicit theory transport, and independent admission.

### 2.3 *Two interacting classes of repair*

Piaget’s distinction suggests more than two labels for updates. It suggests two processes whose interaction is itself informative. The double

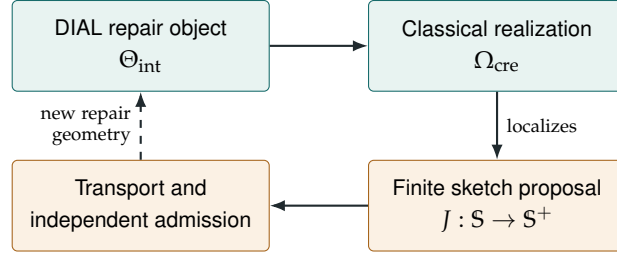


Figure 2.1: The technical spine of infinitesimal creativity. A candidate DIAL object has an interchange defect; a classical double-Lie realization makes that defect computable as a graded commutator. A finite sketch extension states the theory change, and admission tests it.

repair package introduced in Chapter 0 assigns repairs expressible inside the current declaration to the assimilatory side  $A_{\text{asm}}$ , and local probes of a possible reorganization to the proto-accommodative side  $B_{\text{acc}}$ . Those probes test how a schema might have to change; they are not yet the finite change.

The interaction is represented internally by the proposed interchange defect  $\Theta_{\text{int}}$  and, when a classical cochain realization is supplied, by  $\Omega_{\text{cre}}$ . Either object localizes an incompatibility only relative to a declared observer. It does not determine a replacement theory or certify a creative frontier. Chapter 3 constructs the formal target, distinguishes the established classical criterion from the proposed tangent-categorical bridge, and states the realization requirement precisely.

This architecture deliberately leaves an abductive gap. The obstruction  $\Theta_{\text{int}}$ , or its realized observer  $\Omega_{\text{cre}}$ , does not determine a unique  $S^+$ . A frontier model, program synthesizer, scientist, or other proposal engine must construct candidate extensions. Finite sketches then make those candidates explicit enough to compare, transport, and test. Thus DIAL supplies the typed diagnosis, its classical double-Lie realization supplies a computable observer, and sketch surgery supplies the global proposal.

## 2.4 Boden's three modes

Margaret Boden distinguishes combinational, exploratory, and transformational creativity [Boden, 1998, 2004]. We use these as operational modes rather than mutually exclusive psychological essences.

*Combinational creativity (C).* Previously registered objects or constructions are brought into a new typed relation. Analogy, conceptual blending, and reuse belong here, provided the combination is more than juxtaposition.

*Exploratory creativity (E).* A system searches the possibilities generated by a fixed conceptual space. The rules remain stable while their consequences are explored. Theorem proving, program search, optimization, and many evolutionary procedures can be creative in this

sense.

*Transformational creativity (T).* The conceptual space's generating rules change. New primitives, observers, constraints, equivalences, or forms of explanation become available. This is the mode most closely associated with a conceptual jump.

The modes often form a cycle rather than a taxonomy of disjoint episodes. A new combination opens an exploratory region; exploration reveals a stable obstruction; transformation changes the space; and the enlarged space admits new combinations.

#### Creative transition

An infinitesimal probe can reveal that the current theory has reached a boundary. The jump across that boundary is a finite proposal  $S \rightarrow S^+$ , not an infinitesimal update. Local diagnosis and global transformation therefore play different roles.

### 2.5 Four evaluation media, five admission regimes

The creativity mode says *what kind of change* occurred. A testbed says *where the claim can be evaluated*. We use four testbed media whose admission mechanisms are deliberately different. The visual medium supports two regimes because assurance relative to a declared scene and invention of a reusable artistic operation warrant different claims.

*Mathematical theory construction.* Formal systems offer proof, countermodel, and conservative-extension tests. Historical reconstructions provide recognizable controls, while generated miniature worlds reduce contamination by memorized mathematical prose.

*Simulator-grounded theory extension.* An executable simulator supplies interventions and consequences that are not available from text alone. The Prometheus packages—including singing mice, the Indus Valley collapse, and microparticle pollution—offer initial scientific environments in which proposed abstractions can be tested under held-out regimes.

*Scientific documents to testable theories.* A closed, source-bearing corpus is converted into local models, disagreements, candidate mechanisms, and an extensible algebraic theory. This stage can admit provenance and corpus-relative derivations, but causal admission requires a simulator, laboratory, clinical study, or independent structured evidence. The GLP-1 challenge is an initial demanding case: the aim is not a fluent review,

but a theory that organizes mechanisms and supports discriminating predictions.

*Generative visual-language construction.* A text-to-image system supplies a continuous generative process whose outputs are immediately inspectable, while its conditioning vocabulary exposes a representational boundary. ARTISTIC first asks whether a supplied or estimated scene declaration can diagnose and repair violations while preserving unaffected content. DILATE asks the stronger question: whether persistent failure warrants a reusable visual operation unavailable in the frozen language, and whether that operation survives transport and independent admission. A successful repair is therefore infrastructure for artistic theory extension, not evidence of that extension by itself.

|                          | Combinational (C)                                | Exploratory (E)                      | Transformational (T)                               |
|--------------------------|--------------------------------------------------|--------------------------------------|----------------------------------------------------|
| <b>Mathematics</b>       | compose known constructions                      | search consequences of fixed axioms  | introduce a new primitive, law, or equivalence     |
| <b>Simulation</b>        | connect mechanisms or modules                    | explore parameters and interventions | extend the ontology or simulator interface         |
| <b>Scientific corpus</b> | relate source-backed claims                      | compare contextual models            | posit a latent mechanism, context, or law          |
| <b>Visual language</b>   | combine known objects, attributes, and relations | explore conditioned generative flows | introduce a grounded concept, relation, or grammar |

Table 2.1: The creativity matrix. Rows are evaluation environments; columns are modes of conceptual change. A single episode may traverse several cells.

This matrix is the organizing canvas of the book. It prevents two common collapses: identifying creativity with unconstrained generation, and identifying a creativity mode with a particular domain. Each experiment is tagged by a pair such as Mathematics–E, Simulation–T, or Visual–T.

## 2.6 From LINCS repair to creative transformation

LINCS begins with declared compositional obligations, probes their failures, localizes an obstruction, proposes a typed repair, and independently admits or rejects the proposal. Infinitesimal creativity inherits that discipline but widens the repair language.

For fixed-theory learning, the repair acts inside  $S$ :

$$D \rightsquigarrow D'.$$

For transformational creativity, a persistent obstruction can instead motivate a declaration intervention:

$$\begin{aligned} S &\xrightarrow{\text{diagnose}} (\omega, \gamma) \xrightarrow{\text{propose}} S^+, \\ S^+ &\xrightarrow{\text{transport and test}} \{\text{admit, revise, reject, abstain}\}. \end{aligned}$$

The obstruction may include the internal interchange defect  $\Theta_{\text{int}}$ , or its classical cochain observer  $\Omega_{\text{cre}}$ , but it does not logically determine the extension. Many incompatible theories can repair the same failure. Frontier models are therefore useful as proposal engines: they can generate analogies, mechanisms, new types, and candidate laws. They do not thereby become their own judges.

### Boundary

Non-compositionality is a source of diagnostic evidence, not a complete theory of creativity. It can identify where an inherited structure fails without uniquely specifying the concept that should replace or extend it.

## 2.7 The status-bearing extension dossier

To make a creative claim auditable, every proposed jump should produce a *status-bearing extension dossier*

$$\Xi = ((\mathfrak{R}, D), \omega, \gamma, Y, (\mathfrak{R}^+, D^+), \tau, \Pi, \sigma),$$

with the following components:

*Prior state*  $(\mathfrak{R}, D)$ . The maintained representational package and current realization before the proposal.

*Obstruction*  $\omega$ . The localized failure motivating the change, possibly including  $\Theta_{\text{int}}$  or its realized observer  $\Omega_{\text{cre}}$ . For an opportunity-driven proposal with no diagnosed failure, the record should state  $\omega = \emptyset$  rather than retrofit an obstruction after the fact.

*Finite-realization record*  $\gamma$ . The persistence or integration record showing whether a candidate local repair composes coherently beyond the diagnostic neighborhood. A failed realization may motivate accommodation but does not select the extension.

*Package comparison*  $Y$ . The explicit, componentwise change of presentation, doctrine, semantics, probes, observers, operations, or admission. When the presentation or generative language changes,  $Y$  includes the theory map  $J$ .

*Extended state*  $(\mathfrak{R}^+, D^+)$ . The proposed package and a realization interpreted in it.

*Transport*  $\tau$ . An account of how established content is preserved, reinterpreted, or deliberately abandoned.

*Predictions*  $\Pi$ . New theorems, simulations, interventions, or empirical consequences made possible by the proposed change.

*Status*  $\sigma$ . Proposed, admitted, revised, rejected, or presently unidentifiable.

This record separates invention from validation and makes failed creative attempts scientifically useful. Its status field also prevents a candidate from being reported as an admitted change merely because it was generated fluently or reduced the original defect. A rejected proposal can still reveal a missing observer, an inadequate simulator, or a false analogy.

## 2.8 *What counts as success?*

Novelty and validity are necessary but too weak. A random new axiom can be novel, and an isolated theorem can be valid. A persistent theory should be assessed along several dimensions:

1. *coherence*: the extension has a well-defined internal semantics;
2. *conservativity*: established results are preserved where claimed;
3. *explanatory reach*: several prior anomalies receive a common account;
4. *generativity*: the theory produces new questions, constructions, or predictions;
5. *transfer*: its concepts remain useful beyond the originating example;
6. *persistence*: other agents can reuse and extend it; and
7. *nontriviality*: the gain is not explained by paraphrase, memorization, or a disguised restatement.

Formal proof can certify only some of these properties. Simulation yields simulator-relative evidence, not truth about the physical world. Corpus-based evaluation supports empirical claims only to the extent licensed by its sources and identification assumptions. Visual assurance can demonstrate structural control and counterfactual grounding without establishing that a learned representation corresponds to a human concept; open-ended artistic extension has no unique target image and requires reuse, irreducibility, persistence, and independent critical value. The five admission regimes are valuable precisely because their weaknesses differ.

## 2.9 *Personal and historical creativity*

Boden also distinguishes psychological creativity—an idea new to the reasoner—from historical creativity—an idea new to a community. The first can be tested under controlled information conditions. The second requires search, priority evidence, and expert judgment. Our experiments report these statuses separately. Rediscovering a known theorem inside a generated world may demonstrate a mechanism of personal creativity without constituting a historical discovery.

The AM program is an important precedent: beginning from elementary concepts, it used heuristics to construct mathematically interesting concepts and conjectures [Lenat, 1976]. Its successes and limitations suggest that a creative system needs both a generative language and explicit criteria for interestingness. Contemporary frontier models greatly enlarge the proposal engine; they do not remove the need for persistent representations and independent admission.

### 2.10 *From a target to a mechanism*

The status-bearing dossier specifies what a creative episode must leave behind, but not how a system discovers such an extension. That is the next problem. Chapter 3 supplies local probes of a theory and its models; Chapter 4 makes the operations conducting those probes inspectable and trainable; and Chapter 5 states precisely which theory and model semantics a finite proposal changes. Together these chapters separate the local evidence for inadequacy from the global act of theory construction.

That separation remains in force throughout the book. A proposal engine may suggest a finite change, but neither local diagnosis nor fluent generation admits it. The creative claim belongs to the complete record: obstruction, finite realization, less-disruptive controls, package comparison, its presentation component when one exists, transport, novel consequence, and evidence independent of the proposal.

Nor does a silent observer certify completeness: the inherited probe language may be blind to the missing distinction. Chapter 3 therefore treats observability as declared experimental data, not as a property supplied by tangent structure.

#### Challenge

Can a system construct a theory that competent independent reasoners choose to retain because it organizes old knowledge, survives tests, and creates a fruitful space of new problems? This is a stronger target than producing a novel answer and a more operational target than declaring a model creative.

### *Further Reading*

Boden supplies the taxonomy and personal–historical distinction [Boden, 1998, 2004]; Lenat and Lakatos treat concept discovery and theory transformation [Lenat, 1976, Lakatos, 1976]. McCarthy motivates theoretical objects beyond appearance, Zahavy the abductive jump beyond compression, and Feynman compact generative theories [McCarthy, 2007, Zahavy, 2026, Feynman et al., 1963]. Piaget supplies the developmental precedent for assimilation and accommodation [Piaget, 1952, 1954, Piaget and Inhelder, 1974, Piaget, 1985]; Drescher gives its computational continuation [Drescher, 1985, 1986, 1989, 1991]. For the geometric bridge, see involution algebroids [Burke and MacAdam, 2019, MacAdam, 2023], classical double Lie algebroids, and the Meinrenken–Pike bigraded construction [Mackenzie, 2005, Meinrenken and Pike, 2021, Pike, 2020].



### 3

## *The Differential Geometry of Creativity*

### TANGENT SEMANTICS AND DOUBLE REPAIR

Chapter 2 described creativity as a transition between conceptual spaces. We now give that picture a differential geometry. The starting point is not a numerical parameter manifold, but a tangent category whose objects carry admissible notions of infinitesimal variation. Poon Leung’s classification of tangent structures by Weil algebras then supplies a family of algebraic probe shapes for exploring that variation [Leung, 2017, 2018].

In plain language, the chapter asks two questions. What small changes can be made while the current theory is held fixed? And what becomes visible when a second class of small changes probes the theory’s organization itself? DIAL begins from the order effects between those two experiments; it does not identify every local variation with a creative act.

The six mathematical levels below refine the four reading phases introduced in Chapter 0. They describe what must be represented inside the workflow; they are not six additional stages that a reader must execute in sequence.

This viewpoint suggests a precise six-level account of infinitesimal creativity:

1. an ambient category describes the current world of theories or models;
2. a monoidal action of tangent-probe algebras describes how that world can be probed infinitesimally;
3. an involution algebroid internalizes one Lie-algebroid direction in the tangent category;
4. a proposed double internalization organizes two interacting repair directions, with classical double Lie algebroids providing its concrete smooth realization;

*Chapter roadmap.* The chapter first introduces established tangent and Weil semantics, then separates classical double models from the proposed and conjectural bridges. It next formulates the DIAL geometry and process before stating the construction’s limits. Leung’s classification connects the axiomatic tangent-category and synthetic differential-geometric views.

5. a finite-realization observer—called an integration observer when it tests a local-to-finite passage—asks whether a local repair composes into a coherent finite realization; and
6. after less-disruptive controls have been tested, a finite package comparison proposes a change to the presentation, semantics, probes, observers, operations, or admission interface; a sketch map records its presentation component when one exists, and transport plus independent admission determine whether the proposal becomes an accommodation.

The third level is essential. Differential structure can reveal the edge of a conceptual space, but a transformational creative act need not be a tangent vector inside that space.

#### Design principle

A tangent category specifies a geometry of admissible local variation. Leung's Weil theory specifies composable shapes of infinitesimal probe. Involution algebroids supply the established bridge to one Lie-algebroid direction. This book proposes the corresponding double bridge for assimilation and proto-accommodation, using classical double Lie algebroids as its concrete model. Infinitesimal creativity uses their compatibility defect to diagnose a conceptual space, tests whether the resulting local repair persists and integrates, compares it with less-disruptive controls, and records any surviving finite package change before transport and independent admission. A finite sketch map appears when that change extends the presentation.

#### Boundary: Scale convention

The words *infinitesimal* and *finite* name representational levels, not psychological speed or numerical magnitude. An infinitesimal probe is expressed in the maintained tangent semantics; a finite theory change is recorded at the level of whole presentations. Many local updates may remain assimilatory, while adding a single generator or law is already a finite extension. Local evidence may motivate such an extension, but it neither decomposes nor determines the creative jump.

Likewise, a *finite realization* is the application-level record that a candidate local repair persists as a path, rollout, model, or other declared whole. It is broader than the specific Lie-theoretic question of whether a Lie algebroid integrates to a Lie groupoid. The latter is called *algebroid integration* when the distinction matters.

**Boundary: Formal status convention**

Five labels are used deliberately. A *cited theorem* imports a published mathematical result with its stated hypotheses and source; it is not presented as a new proof of this book. An unqualified *definition* stipulates local terminology or data and does not assert that nontrivial examples exist. A *proposed definition* specifies a research target whose existence, classification, or realization remains open. A *hypothesis* or *conjecture* is not used as a proved premise unless an application explicitly assumes it. A *design principle* states an architectural or empirical discipline, not a theorem.

Under this convention, tangent categories, Leung’s classification, involution algebroids, and classical double Lie algebroids are cited mathematics. DIAL objects and their realization maps are proposed constructions. Interpreting their directions as *assimilation* and *proto-accommodation* is an empirical design hypothesis. Later computational experiments can test a finite proxy without proving the proposed internal DIAL theory.

### 3.1 *The ambient differential universe*

Let  $\mathcal{M}$  be a category whose objects are structured theory states, models, or realizations and whose arrows are admissible structure-preserving transformations. A tangent structure equips  $\mathcal{M}$  with an endofunctor

$$T : \mathcal{M} \longrightarrow \mathcal{M}$$

and natural transformations expressing projection, zero, fiberwise addition, vertical lift, and interchange [Cockett and Cruttwell, 2014]. For an object  $X$ , the projection

$$p_X : TX \longrightarrow X$$

organizes first-order variations over their base state.

It is tempting to call  $\mathcal{M}$  “the search space.” That phrase is useful only with care. A tangent category is a categorical universe containing many spaces and transformations, not one set of candidate answers. The tangent structure says which local variations are meaningful and how they compose. The application supplies the semantics: a tangent direction may vary an axiom weight, an intervention field, a simulator mechanism, an observer, or a local model.

**Definition 3.1** (Conceptual differential universe). A conceptual differential universe is a tangent category  $(\mathcal{M}, T)$  together with an interpretation assigning theory states and admissible conceptual variations to selected objects and tangent fibers.

The interpretation clause prevents a purely formal tangent vector from being mistaken for a meaningful conceptual change. It plays the same role as the observability caveat in *Learning in Infinitesimal Non-Compositional Sketches* (LINCS): structure must be connected to an observer and an experimental semantics before it can guide learning.

### 3.2 Weil algebras as probe shapes

The basic first-order infinitesimal algebra is

$$W = \mathbb{N}[\varepsilon] / (\varepsilon^2).$$

The coefficient rig  $\mathbb{N}$  belongs to Leung’s universal algebraic presentation. It is not a claim that the semantic tangent fibers are natural-number spaces. A realization functor may send this probe shape into a smooth, synthetic, or other tangent setting; familiar real dual numbers and the Kock–Lawvere object arise only after such semantic structure has been chosen. Its nilpotent generator records a first-order displacement while suppressing higher powers. More elaborate Weil algebras describe several infinitesimal directions and relations among them. Write  $\text{Weil}_1^{\text{tan}}$  for Leung’s category  $\mathbb{N}\text{-Weil}_1$ : it is generated from the basic algebra  $W$  using the monoidal coproduct and the designated products and foundational pullbacks. Its product powers include algebras of the form

$$W_n = \mathbb{N}[\varepsilon_1, \dots, \varepsilon_n] / (\varepsilon_i \varepsilon_j \mid 1 \leq i \leq j \leq n),$$

together with their structure-preserving maps.

The crucial point is that a Weil algebra is not a model of the whole conceptual space. It is a *probe shape*. A monoidal functor realizes that shape as an operation on every object of the ambient category.

**Theorem 3.2** (Leung’s classification, cited). *Giving a tangent structure on a category  $\mathcal{M}$  is equivalent, up to isomorphism, to giving a strong monoidal functor*

$$F : \text{Weil}_1^{\text{tan}} \longrightarrow \text{End}(\mathcal{M})$$

*that preserves Leung’s designated foundational pullbacks and the equalizer encoding the universality axiom for the vertical lift. The generating algebra  $W$  is sent to the tangent functor  $T$ , while maps among Weil algebras induce the structural natural transformations of the tangent category [Leung, 2017, Theorem 14.1].*

Thus  $F(A)X$  is the realization at  $X$  of the infinitesimal experiment encoded by  $A$ . The monoidal product organizes composition of probe shapes: up to the coherence supplied by the strong monoidal structure,

$$F(A \otimes B) \cong F(A) \circ F(B).$$

Projection, zero, addition, lift, and canonical flip arise from distinguished maps between the relevant Weil algebras. What appears axiomatically as a list of tangent-category operations is thereby organized by a monoidal algebraic theory.

### 3.3 *The missing bridge: from tangent semantics to double algebroids*

There is a terminological and mathematical seam that must not be hidden. Leung’s  $\text{Weil}_1^{\text{tan}}$  and the Weil algebra later used for a double Lie algebroid are *not the same construction*. The former is a monoidal algebraic theory whose actions classify tangent structure. The latter is a bigraded cochain algebra

$$\mathcal{W}_{\text{Lie}}^{\bullet,\bullet}(\mathbb{D})$$

attached by Meinrenken and Pike to a classical double vector bundle. Its two differentials encode candidate horizontal and vertical VB-algebroid structures. Sharing the name Weil reflects a common infinitesimal lineage, not an identity between the two objects.

An established first-order bridge is available, but it is not an equivalence of theories. Under the tangent-category limits and negatives assumed by Burke and MacAdam, they define *involution algebroids* internally, and MacAdam develops their functorial semantics. In the ordinary smooth model, every classical Lie algebroid determines an involution algebroid. Conversely, an involution algebroid carries a Lie bracket on the sections of its underlying bundle; a suitable unit object supplies the Leibniz law [Burke and MacAdam, 2019, Sections 4.2.1–4.2.4].

The smooth comparison is only an injection on objects, not an equivalence of the full categories [Burke and MacAdam, 2019, Sections 4.2.1–4.2.4]. The warranted relationships are therefore

$$\text{Weil}_1^{\text{tan}}\text{-action} \longrightarrow \text{tangent structure},$$

which supplies ambient semantics for involution algebroids, while

$$\text{Lie algebroid} \longrightarrow \text{involution algebroid} \longrightarrow \text{section bracket}.$$

The word *involution* hides an important coherence law. For an anchored bundle  $\pi : A \rightarrow M$ , Burke and MacAdam replace a bracket on sections by an arrow  $\alpha$  on the prolongation  $A \times_{\rho, T\pi} T(A)$ . Writing  $\sigma = (p\pi_1, \alpha)$ , their involution axiom gives  $\sigma^2 = 1$ , while their *flip* axiom can be written in the Yang–Baxter-style form

$$\begin{aligned} & (\sigma \times c)(1 \times T\sigma)(\sigma \times c) \\ &= (1 \times T\sigma)(\sigma \times c)(1 \times T\sigma), \end{aligned}$$

where  $c : T^2 \Rightarrow T^2$  is the canonical tangent flip. Thus three infinitesimal rearrangements agree independently of which adjacent pair is resolved first. After passage to sections, this coherence induces a Lie bracket satisfying Jacobi. Jacobi is therefore not discarded; it is recovered from a section-free internal law that makes sense in a tangent category [Burke and MacAdam, 2019].

This has a distinguished precedent but requires a careful distinction. Drinfel'd related the *classical* Yang–Baxter equation for  $r$ -matrices to Hamiltonian Poisson–Lie structures and Lie bialgebras [Drinfel'd, 1983]. Burke and MacAdam's equation is instead a Yang–Baxter-style *braid relation* for an involution on a prolongation; it is not Drinfel'd's classical  $r$ -matrix equation. The common geometric lesson is that a local algebraic law can encode order-independent composition before its familiar bracket-level identity is extracted.

For DIAL this yields a sharper architecture. Each of the horizontal and vertical directions must first satisfy its own Yang–Baxter-style flip law. Only then is it meaningful to impose the mixed interchange, matched-pair, or super-commutation condition between them. A failure internal to one side is a malformed repair process; a failure between two individually coherent sides is a genuinely double obstruction.

Mackenzie's *Notions of Double for Lie Algebroids* is the closest classical precursor to the required second step [Mackenzie, 2000a]. It unifies several meanings of double: Drinfel'd and Manin-triple doubling, matched pairs, the cotangent double associated with a Lie bialgebroid, and categorical or iterated doubles such as  $T(TM)$ . A double Lie algebroid is not merely a square carrying two unrelated Lie-algebroid structures. Mackenzie's compatibility uses the duality theory of double vector bundles; it recovers the infinitesimal invariants of double Lie groupoids and specializes to matched pairs and Lie bialgebroids.

This precedent both guides and limits DIAL. Mackenzie observes that the compatibility equation should admit formulation in fairly general categories, but his construction depends crucially on finite-rank vector-bundle duality. An internal tangent-categorical theory must therefore reconstruct that duality, replace it with an appropriate internal comparison, or state explicitly the dualizability doctrine under which DIAL exists. No result cited here automatically performs that internalization. We therefore make the missing step explicit as a research hypothesis.

### 3.4 *Tensor products of sketches as a semantics for doubling*

MacAdam points to a second route across the missing bridge, one that connects the double geometry directly to the theory-building language of this book [MacAdam, 2023]. In a closed monoidal doctrine of limit

sketches in which the required tensor and internal model categories exist, tensor-product semantics gives equivalences

$$\mathrm{Mod}(\mathbb{A}, \mathrm{Mod}(\mathbb{B}, \mathbf{Set})) \simeq \mathrm{Mod}(\mathbb{A} \otimes \mathbb{B}, \mathbf{Set}) \simeq \mathrm{Mod}(\mathbb{B}, \mathrm{Mod}(\mathbb{A}, \mathbf{Set})).$$

The middle sketch contains both presentations. Its models preserve the distinguished limits in either coordinate, and its commuting squares encode interchange between the two structures. In this sense an  $\mathbb{A}$ -structure internal to  $\mathbb{B}$ -models can be uncurried into one model of a tensor-product theory and curried again in the opposite order.

Let  $\mathbb{I}_{\mathrm{Lie}}$  denote the proposed enriched limit sketch, or classifying tangent-categorical theory, for an involution algebroid. The candidate presentation

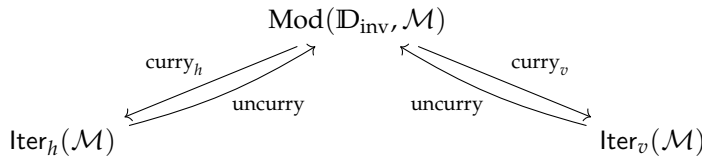
$$\mathbb{D}_{\mathrm{inv}} := \mathbb{I}_{\mathrm{Lie}} \otimes \mathbb{I}_{\mathrm{Lie}}$$

is therefore a natural syntactic target for DIAL. A model would be an involution-algebroid structure internal to involution-algebroid models, with the tensor-product squares recording horizontal–vertical interchange. More generally,  $\mathbb{I}_{\mathrm{Lie}} \otimes \mathbb{V}$  is the corresponding candidate presentation for a vector-bundle object in involution algebroids, the tangent-categorical analogue suggested by VB–Lie algebroids. These formulas are a research construction, not a theorem that ordinary unenriched finite-limit sketches already encode every tangent, smooth, duality, and bracket requirement. The relevant tensor product must live in the enriched sketch doctrine appropriate to tangent categories.

This viewpoint explains why the two DIAL directions should not be assembled as an arbitrary pair of learners. They are two models of declared theories, and their interaction is itself presented. It also provides a formal version of order-independence. Under the tensor-product theorem for the selected limit-sketch doctrine, exchanging the two factors gives the semantic symmetry

$$\mathrm{Mod}(\mathbb{A}, \mathrm{Mod}(\mathbb{B}, -)) \simeq \mathrm{Mod}(\mathbb{B}, \mathrm{Mod}(\mathbb{A}, -)).$$

MacAdam observes that Mackenzie’s symmetry of partial Lie derivatives should become immediate from this perspective: for a double Lie groupoid, applying the Lie functor in the two orders produces canonically corresponding double infinitesimal structures [Mackenzie, 2000b, MacAdam, 2023].



The construction suggests a precise separation of three questions:

Figure 3.1: Sketch semantics for double involution. The two outer expressions denote the horizontal-first and vertical-first iterated model categories. They read one tensor-product model in opposite orders. Establishing the appropriate enriched doctrine and its classical realization is part of the DIAL program.

1. *presentation*: does  $\mathbb{I}_{\text{Lie}} \otimes \mathbb{I}_{\text{Lie}}$  carry the required side and interchange axioms in an enriched tangent doctrine?
2. *realization*: are its finite-rank smooth models equivalent to Mackenzie double Lie algebroids, including the required duality rather than merely two internal algebroid structures?
3. *integration*: when does such a smooth infinitesimal model arise from a double Lie groupoid?

Only the first question is formal sketch semantics. Even an ordinary Lie algebroid need not integrate to a Lie groupoid: integrability is controlled by global monodromy obstructions [Crainic and Fernandes, 2003]. In the double setting, sidewise integrability is necessary but is not by itself a proof that the integrations assemble into a smooth double groupoid; compatible integration is a further global obligation [Stefanini, 2009].

#### Boundary

Tensor-product symmetry is a semantics of *order independence*; it is not an integrability theorem. It can explain why two already defined partial Lie constructions agree, and it can present the compatibility law to be tested. It does not remove monodromy, smoothness, source-simply-connectedness, or double-integration obstructions. DIAL must report these as separate admission conditions whenever local repair fields are interpreted as finite creative trajectories.

For synthetic creativity, sketches now have two roles at different scales. The tensor product presents the local two-direction repair calculus. The finite map  $J : S \rightarrow S^+$  presents a proposed change in the domain theory.

The passage between those scales has three gates. First, failure of a typed interchange equation localizes mixed generators that may require repair. Second, a finite-realization test distinguishes a formal infinitesimal direction from a coherent transition. Third, an admitted transition may support construction or selection of  $S^+$ . This sequence joins the geometry of this chapter to the algebraic-theory targets of Chapter 5 without identifying local compatibility with global creative success.

*Homotopy coordinates.* Gracia-Saz, Jotz Lean, Mackenzie, and Mehta give a third, computationally useful characterization [Gracia-Saz et al., 2014]. Choose a linear splitting  $\Sigma$  of a classical double vector bundle with side bundles  $A, B$  and core  $C$ . The two VB-algebroid structures



induce:

- an  $A$ -2-representation on  $\partial_B : C \rightarrow B$ ,
- a  $B$ -2-representation on  $\partial_A : C \rightarrow A$ .

For the chosen splitting, their Theorem 3.4 states that the original structure is a double Lie algebroid if and only if these two 2-term representations up to homotopy form a matched pair [Gracia-Saz et al., 2014, Theorem 3.4]. The matched-pair condition expands into nine compatibility equations involving anchors, boundary maps, connections, and curvature.

For a chosen splitting, write

$$r_{\Sigma}^{\text{DIAL}} = (r_{\Sigma}^{(1)}, \dots, r_{\Sigma}^{(9)})$$

for the left-minus-right residuals of those equations. Its vanishing is equivalent to classical double-Lie compatibility. The coordinates depend on  $\Sigma$ , but the vanishing claim does not. A change of splitting therefore acts like a change of gauge: an empirical DIAL diagnostic must either be invariant under that change or report its dependence explicitly.

#### Boundary

Representation *up to homotopy* does not mean that an arbitrary interchange failure is acceptable. Curvature is retained coherently inside the two 2-term representations, but the pair must still satisfy all matched-pair conditions. Homotopy enriches the internal coordinates of compatibility; it does not replace admission by an unconstrained tolerance.

#### 3.4.1 The proposed internal DIAL target

The preceding ingredients do not yet define an internal double object. The classical Weil criterion states what smooth compatibility must recover; the tensor-sketch construction suggests how two presentations may interact; and the homotopy coordinates describe one classical realization. The following definition packages these demands as a research target.

**Definition 3.3** (Candidate DIAL object, proposed). Let  $(\mathcal{M}, T)$  be a tangent category with the differential bundles and pullbacks needed below. A candidate *Double Involution ALgebroid* (DIAL) is a square of differential bundles in  $\mathcal{M}$ , equipped with horizontal and vertical involution-algebroid structures, a declared dualizability doctrine, and an interchange comparison whose smooth finite-rank realization is Mackenzie's double-Lie compatibility datum. We reserve the name *DIAL object* for a candidate whose comparison satisfies the required

coherence equations. Until the missing duality and coherence theory is completed, this is a target definition rather than a claim of an established category of such objects.

Write  $\Theta_{\text{int}}$  for the typed failure of this internal interchange comparison. This notation is deliberately schematic: a full theory must specify the precise 2-cells, pullback hypotheses, and coherence laws. The intended adequacy conditions are:

1. in finite-rank **Smooth**, the construction recovers Mackenzie’s classical double Lie algebroids, not merely two VB-algebroid structures on a common square;
2. tangent prolongation is represented internally;
3. the vacant case recovers matched pairs, while the cotangent case recovers the double associated with a Lie bialgebroid;
4. a suitable cochain realization  $\text{Real}$  sends the internal defect to the graded commutator

$$\text{Real}(\Theta_{\text{int}}) = [d_{\text{asm}}, d_{\text{acc}}]_{\text{gr}} =: \Omega_{\text{cre}};$$

5. the construction is stable under the transports used for finite sketch extension;
6. its dualizability assumptions, or its replacement for double-vector-bundle duality, are explicit; and
7. every admissible splitting yields a matched pair of 2-term representations up to homotopy, and compatibility is independent of that auxiliary splitting.

#### Boundary

The first-order tangent-category–Lie-algebroid bridge is established. The double internalization and the realization  $\text{Real}(\Theta_{\text{int}}) = \Omega_{\text{cre}}$  are proposed bridge requirements for this research program. Meinrenken–Pike prove the super-commutation criterion in classical double Lie geometry; they do not prove its internalization in an arbitrary tangent category or smooth topos. Mackenzie supplies the classical compatibility target and its major special cases, but not the proposed tangent-categorical DIAL semantics.

#### 3.4.2 Formal status and empirical burden

DIAL turns the missing bridge into two linked obligations. The formal obligation is not a single missing lemma. It comprises five tasks:

1. define a category  $\mathbf{DIAL}(\mathcal{M})$ , including its objects, morphisms, side-wise involution laws, and mixed interchange axiom;
2. prove stability under tangent lift and under the transports required by finite theory extension;
3. establish a realization theorem relating  $\mathbf{DIAL}(\mathbf{Smooth})$  to classical double Lie algebroids, with the required dualizability doctrine stated explicitly;
4. determine whether the resulting objects admit products, pullbacks, representations, cohomology, and integration to an appropriate double groupoid semantics; and
5. internalize the matched-pair description by 2-term representations up to homotopy, prove independence from the chosen splitting, and explain how the two Yang–Baxter-style side laws interact with mixed double coherence.

The last task matters because two valid involution algebroids do not automatically form a DIAL object.

The empirical obligation is to show whether this additional structure earns its keep. A computational DIAL realization must expose two typed repair directions, estimate an interchange witness, localize its support, and test whether the candidate local repair persists as a coherent finite realization. Only then may that information support selection or construction of a finite sketch extension. The complete system must be compared with single-direction tangent repair, untyped two-operator search, fixed-sketch LINCOS, and an equally resourced frontier-model proposer. Success means better admitted extensions, not merely a larger obstruction score.

#### Design principle

DIAL is falsifiable at two levels. It can fail mathematically if no coherent internal double construction has the claimed classical realization, and it can fail computationally if its typed interchange information does not improve proposal quality, sample efficiency, localization, or conservative abstention.

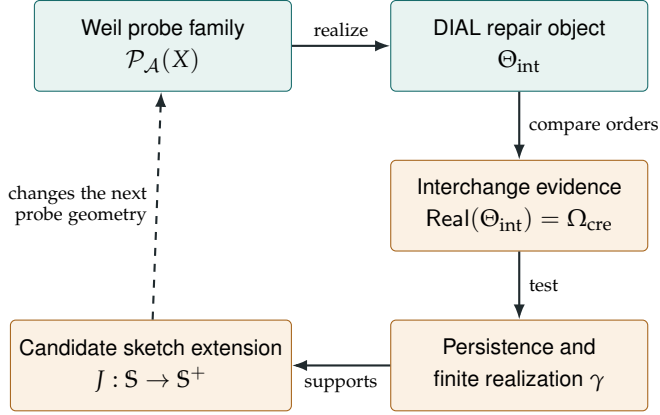


Figure 3.2: The differential scaffolding for infinitesimal creativity. Leung’s Weil theory indexes tangent probes; a candidate DIAL object compares two classes of variation; a classical cochain realization exposes their interchange defect. Persistence and finite realization must still be tested before that evidence can support, but never determine, a finite sketch extension.

### 3.5 A Weil profile of a theory state

For a selected family  $\mathcal{A} \subseteq \text{Weil}_1^{\text{tan}}$ , define the *Weil profile* of a theory state  $X$  schematically by

$$\mathcal{P}_{\mathcal{A}}(X) = \{F(A)X, F(u)_X \mid A, B \in \mathcal{A}, u : A \rightarrow B\}.$$

The profile contains not just individual perturbations but their relations under maps and monoidal composition. It can therefore record questions that a single gradient cannot ask:

- Which first-order directions are admissible at  $X$ ?
- Which pairs of directions can be compared or combined?
- Does changing their order alter the resulting observation?
- Do locally legal variations close inside the current declaration?
- Which variations are null under the chosen observer?

An empirical implementation never observes this entire family. It chooses a finite probe basis, realizes it computationally, and applies observers. Probe selection is therefore part of the creativity architecture rather than a neutral preprocessing step.

### 3.6 Double-infinitesimal repair geometry

Weil profiles describe families of probes, but they do not yet distinguish two repair processes or explain how those processes interact. For that purpose, let

$$\begin{array}{ccc} \mathbb{D}_{\text{rep}} & \xrightarrow{q_h} & B_{\text{acc}} \\ q_v \downarrow & & \downarrow \\ A_{\text{asm}} & \longrightarrow & M \end{array}$$

be a double vector bundle. Its horizontal and vertical vector-bundle structures commute. The side bundle  $A_{\text{asm}} \rightarrow M$  represents assimilatory directions;  $B_{\text{acc}} \rightarrow M$  represents proto-accommodative directions. The application, not the double-vector-bundle axioms, supplies these meanings.

The canonical projection

$$\mathbb{D}_{\text{rep}} \longrightarrow A_{\text{asm}} \times_M B_{\text{acc}}$$

forgets joint information. Its kernel over the zero sections is the core  $K \rightarrow M$ . Core elements project trivially to both side bundles. The core is therefore a candidate carrier of mixed corrections: under an application semantics, an assimilatory repair may alter the meaning of an accommodative probe even when neither side records that effect alone. This interpretation is additional structure, not a consequence of the double-vector-bundle axioms.

In a classical smooth model, equip each direction with a VB-algebroid structure.

This gives horizontal and vertical differentials on Meinrenken–Pike’s bigraded cochain Weil algebra

$$\mathcal{W}_{\text{Lie}}^{\bullet,\bullet}(\mathbb{D}_{\text{rep}}), \quad d_{\text{asm}} : (p, q) \mapsto (p + 1, q), \quad d_{\text{acc}} : (p, q) \mapsto (p, q + 1).$$

Because both differentials have odd total degree, their graded commutator is

$$\Omega_{\text{cre}} := [d_{\text{asm}}, d_{\text{acc}}]_{\text{gr}} = d_{\text{asm}}d_{\text{acc}} + d_{\text{acc}}d_{\text{asm}}.$$

**Theorem 3.4** (Meinrenken–Pike Weil-algebra criterion, cited). *Let  $\mathbb{D}_{\text{rep}}$  be a double vector bundle equipped with VB-algebroid structures over both side bundles. The two structures make  $\mathbb{D}_{\text{rep}}$  a double Lie algebroid if and only if their horizontal and vertical differentials on the bigraded Weil algebra supercommute, equivalently  $\Omega_{\text{cre}} = 0$  [Meinrenken and Pike, 2021, Theorems II and 8.2].*

The theorem supplies a clean diagnostic interpretation in the classical realization. If  $\Omega_{\text{cre}} \neq 0$ , the candidate pair is *not yet* a double Lie algebroid. The two local repair semantics are incompatible. Once a representation and localization rule have been fixed, the value and support of  $\Omega_{\text{cre}}$  can identify an algebraic locus where the two processes fail to supercommute. Treating that locus as a repair frontier requires a declared observer connecting the residual to task semantics. A model can observe this typed defect through a scalar norm or statistical test, but the obstruction itself is the bidegree- $(1, 1)$  operator. At the internal tangent-categorical level, the corresponding diagnostic is  $\Theta_{\text{int}}$ ; identifying it with  $\Omega_{\text{cre}}$  requires the cochain realization described above.

*A coordinate model of a missing mixed direction.* The smallest useful picture lives on  $M = \mathbb{R}^3$ . Let two locally available motions be

$$X = \frac{\partial}{\partial x}, \quad Y = \frac{\partial}{\partial y} + x \frac{\partial}{\partial z}.$$

At the origin, the registered side directions span only the  $x$ - and  $y$ -axes. Their bracket is

$$[X, Y] = \frac{\partial}{\partial z},$$

so an infinitesimal loop that follows  $X$ , then  $Y$ , then reverses the two motions produces a second-order displacement in a direction absent from the original side span. The  $z$ -direction is the elementary coordinate shadow of mixed information that neither side reports alone.

This example is not itself a DIAL object, and a nonzero bracket is not yet a creative event. It shows only the geometric question that the double repair construction types more carefully: does an order effect close inside the declared representation, disappear under the registered observer, or expose a persistent direction for which the current theory has no expression?

**Definition 3.5** (Double repair geometry, proposed). A double repair geometry consists of a candidate DIAL object and, when available, a classical double-vector-bundle realization  $\mathbb{D}_{\text{rep}}$  with candidate horizontal and vertical VB-algebroid structures. It also includes a chosen finite family of tangent-probe observables and an admission semantics for deciding whether an observed interchange defect is persistent.

This definition intentionally allows incompatible candidates. Internally, a compatible DIAL object is the zero- $\Theta_{\text{int}}$  case. In the classical cochain realization, a compatible double Lie algebroid is the zero- $\Omega_{\text{cre}}$  case. Either defect may motivate repair of the coupling, enrichment of the core, or a larger theory change.

The canonical example is the tangent prolongation of a Lie algebroid  $A \rightarrow M$ :

$$\begin{array}{ccc} TA & \xrightarrow{Tq} & TM \\ \downarrow & & \downarrow \\ A & \xrightarrow{q} & M. \end{array}$$

Its double Weil algebra recovers the Weil algebra of  $A$  [Meinrenken and Pike, 2021]. This makes the construction more than a formal analogy: a Lie algebroid of creative skills can be tangent-lifted, so one direction represents available skills and the other their variation across theory states. Chapter 4 develops that connection.

**Design principle**

Use the double structure to compare local repair orders, not to smuggle a finite theory change into tangent notation. The creative interchange obstruction localizes what the current coupled geometry cannot reconcile. A candidate repair must first persist and integrate into a coherent finite realization. If the old theory cannot support such a realization, a proposal engine may express a replacement as a finite sketch extension  $J : S \rightarrow S^+$ ; admission must then test its transport and new consequences.

### 3.7 Curve objects and the dynamics of theory revision

The preceding geometry identifies two infinitesimal repair directions, but it does not yet say how either direction evolves. Cockett, Cruttwell, and Lemay develop differential equations directly in a tangent category by introducing a *curve object*: a structural analogue of the time line that can parameterize solutions and flows [Cockett et al., 2021]. This supplies a disciplined integration semantics for DIAL.

Let  $(\mathcal{M}, T)$  be a tangent category with the products and tangent limits required for this construction. A vector field on an object  $M$  is a section

$$V : M \longrightarrow TM, \quad p_M V = 1_M.$$

A parameterized dynamical system consists schematically of an object  $M$ , a vector field  $V$ , and an initial-state map  $g : X \rightarrow M$ . Given a curve object  $(C, c_0, c_1)$ , a solution is a map

$$\gamma : C \times X \longrightarrow M$$

that assumes the initial value  $g$  at  $c_0$  and whose tangent in the  $C$ -direction agrees with  $V$ . In ordinary smooth geometry one may take  $C = \mathbb{R}$ , with  $c_0 = 0$  and  $c_1$  the constant unit vector field.

The categorical definition contains an important qualification. A curve object is preinitial among dynamical systems in every parameter context: solutions are unique when they exist, but existence is not automatic. A vector field is *complete* when solutions exist for every initial condition. Under the hypotheses of the theory, complete vector fields correspond to flows

$$\Phi : C \times M \longrightarrow M.$$

Thus a curve object is not a trajectory and DIAL is not itself a curve object. The curve object supplies abstract time;  $M$  is the maintained theory-state object; and  $\gamma$  is a possible history of theory repair.

*The involution-algebroid variant.* The definition above is the later, general formulation of curve-object dynamics. Burke and MacAdam use a related but differently packaged notion in their development of involution algebroids [Burke and MacAdam, 2019, Section 2.5]. Their *complete curve object* is a dynamical system

$$R_{\text{curv}} = (0_R : 1 \rightarrow R, \partial : R \rightarrow TR)$$

with a second distinguished point  $1_R : 1 \rightarrow R$ , uniqueness of morphisms from it to any dynamical system, and a lifting condition for linear vector fields. More precisely, for a differential bundle locally trivial with respect to  $R$ , a linear system over a base system with a complete solution must itself have a complete solution. The additional point is not used in their paper, but is retained in anticipation of defining the target map of a groupoid integrating an involution algebroid.

Cockett, Cruttwell, and Lemay subsequently separate these ingredients. Their curve object is preinitial in every parameter context, its distinguished vector field commutes with itself, and that field is complete. They then state *linear completeness* as an additional axiom: a linear vector field on a differential bundle is complete whenever its base vector field is complete. The two presentations should therefore not be conflated. The earlier variant foregrounds the linear lifting needed for algebroid transport; the later framework supplies the contextual theory of solutions, flows, and commuting dynamics used here.

For DIAL, both layers matter. The ordinary curve-object axioms parameterize the two local theory-state flows. Transport of elements in the side or core bundles additionally requires the appropriate linear-completeness doctrine. Indeed, Burke and MacAdam use their complete curve object to extend infinitesimal transport along variations of admissible algebroid paths to full transport. A future integration of DIAL should consequently declare which differential bundles are covered, whether completeness is global or partial, and whether the curve object also carries the unit and function-space structure required for path and homotopy objects.

#### Boundary

In this book, *curve object* refers by default to the contextual notion used for differential equations and flows. When algebroid transport is required, we additionally assume and name the relevant linear-completeness doctrine. This convention prevents uniqueness of trajectories from being mistaken for existence of the linear transports needed to integrate DIAL.



### 3.7.1 Two flows rather than one

The double structure suggests two vector fields on a chosen realization of the theory-state object,

$$V_{\text{asm}}, V_{\text{acc}} : M \longrightarrow TM,$$

representing assimilation and proto-accommodation. This notation presupposes that the relevant sections of the two side algebroids have been realized as vector fields on a common state object. A double Lie algebroid does not supply that realization automatically; anchors, sections, and the required integrability conditions are part of the application contract.

When both fields are complete, they induce flows  $\Phi_{\text{asm}}$  and  $\Phi_{\text{acc}}$ . Their interaction compares the two orders

$$\begin{aligned} &\Phi_{\text{acc}}(t, \Phi_{\text{asm}}(s, m)), \\ &\Phi_{\text{asm}}(s, \Phi_{\text{acc}}(t, m)). \end{aligned}$$

If the flows commute, the two directions assemble into a coherent local two-parameter evolution

$$\Gamma : C \times C \times M \longrightarrow M.$$

Classically, commuting flows correspond to vanishing Lie bracket under the usual completeness hypotheses; Cockett, Cruttwell, and Lemay establish the tangent-categorical relation between commuting vector fields and commuting flows. For DIAL, this gives geometric meaning to the mixed comparison: compatibility says that the infinitesimal rectangle closes. A nonzero interchange defect says that assimilating and probing accommodation in the opposite order lead to distinguishable states.

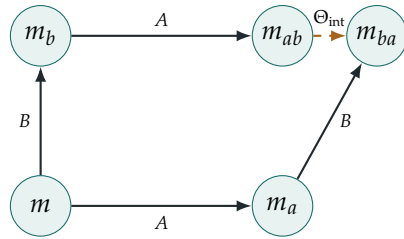


Figure 3.3: DIAL as two-parameter local dynamics, with  $A$  denoting assimilation and  $B$  proto-accommodation. When the flows commute, the infinitesimal rectangle closes and  $m_{ab} = m_{ba}$ . A persistent failure to close localizes order dependence; it is not yet a finite change of theory.

This interpretation connects the formalism to the approximate holonomy experiments described later. A finite empirical realization estimates the discrepancy around small rectangles rather than observing an abstract bracket directly. Shrinking-loop behavior tests whether the discrepancy is consistent with a local interchange obstruction, while uncertainty estimates guard against declaring a structural defect from sampling noise.

### 3.7.2 From flow to a change of theory

A single autonomous flow evolves on a fixed object  $M$ . Full accommodation may instead replace the sketch  $S$  by  $S^+$ , changing the very object on which subsequent dynamics occurs. DIAL therefore requires a hybrid or fibred dynamics. Write  $M_S$  for the state object over a theory presentation  $S$ . Its evolution has the schematic form

$$\begin{aligned} M_S &\xrightarrow{\text{local DIAL flows}} M_S, \\ M_S &\xrightarrow[\text{semantic transport}]{J:S \rightarrow S^+} M_{S^+}, \\ M_{S^+} &\xrightarrow{\text{new local flows}} M_{S^+}. \end{aligned}$$

The middle transition is indexed by a finite sketch map  $J : S \rightarrow S^+$ . It requires the semantic transport developed below—often expressed by Kan extension—rather than another infinitesimal time step. The transported state initializes a new differential system in the new fiber.

Completeness now acquires a useful epistemic interpretation. Assimilatory fields may be complete over a declared regime, expressing the ability to keep repairing within the current theory. Proto-accommodative fields may be only partial: a proposed direction can fail to integrate, encounter a singularity, or leave the admitted state space. Such failure does not uniquely determine a new theory, but it is evidence that the present differential semantics is insufficient.

#### Design principle

Model assimilation and proto-accommodation as partial flows within a theory fiber. Treat their failure to commute or integrate as typed diagnostic evidence. Represent full accommodation separately as a finite sketch change, transport the admitted state into the new fiber, and only then resume the infinitesimal dynamics.

This yields a precise research hypothesis rather than an established theorem: under suitable anchor, realization, and completeness assumptions, the two DIAL directions integrate to partial curve-object flows. Vanishing internal interchange defect should yield a coherent local  $C \times C$ -action, while a persistent nonvanishing defect should constrain a finite theory-extension proposal. Establishing that correspondence is part of the DIAL mathematical program.

### 3.8 Assimilation and accommodation in differential form

Piaget's distinction between assimilation and accommodation can now be made more precise. Let  $X \in \mathcal{M}$  denote the maintained schema or

theory state. An infinitesimal variation at a generalized point  $x : S \rightarrow X$  is a map

$$v : S \longrightarrow TX, \quad p_X v = x.$$

When the application supplies an integration semantics, such a variation may generate a path or finite update  $X \rightsquigarrow X'$ . A bare tangent category does not guarantee that every tangent vector integrates, so the update must remain part of the application-level admission contract.

**Definition 3.6** (Proto-accommodative probe). A proto-accommodative probe is a section of the interpreted side bundle  $B_{\text{acc}} \rightarrow M$  that tests a local change in the organization, observer language, or coupling of a maintained theory. It remains a probe inside the current double repair package; it is not yet a theory extension.

**Definition 3.7** (Differential assimilation). An assimilatory update is an admitted change  $X \rightsquigarrow X'$  whose variation, observations, and repair are expressible using the maintained ambient category, Weil action  $F$ , interpretation, and observer family. It changes the state described by the theory without changing what counts as an admissible local variation.

Assimilation is therefore not synonymous with a small numerical step. A long trajectory, a composite of many probes, or a large parameter update may still be assimilatory if it remains inside the geometry already declared. What matters is whether the current conceptual coordinates and their transition rules suffice to describe the change.

Pressure for finite theory extension begins when they do not. A persistent defect may show that a direction is absent, that locally admissible variations fail to close, that an observer conflates distinct mechanisms, or that the present coordinates do not support the invariant required by the evidence. The response can occur at several levels:

1. move to a new base representation or stratum still available in  $\mathcal{M}$ ;
2. enlarge the observer family or the admitted probe language;
3. replace the tangent action  $F$  by a richer action  $F^+$ ; or
4. propose a change of ambient theory through candidate differential extension data  $(J, \bar{J}, \chi) : (\mathcal{M}, F) \rightarrow (\mathcal{M}^+, F^+)$ .

**Definition 3.8** (Admitted differential accommodation). An admitted differential accommodation is a finite, typed change to the maintained representation or its differential semantics whose coherent finite realization, semantic transport, and new consequences have passed a registered test independent of proposal construction. Such a change may have been proposed because no admitted assimilatory repair accounted for a persistent obstruction. In its strongest form it consists of

a differential theory extension  $(J, \bar{J}, \chi)$ , together with a repaired state  $r : \bar{J}X \rightarrow X^+$  in the extended universe and its admission record.

This definition distinguishes a local proto-accommodative probe from accommodation of the declaration. It also makes accommodation relative to a declared resolution. A change that requires a new object at one level may be an ordinary motion inside a richer theory at another. The audit record must therefore say which of  $X$ , the observers, the probe doctrine,  $F$ , or  $\mathcal{M}$  changed.

Piaget's conservation task illustrates the distinction geometrically [Piaget and Inhelder, 1974]. Pouring water changes visible height and container shape. An appearance-based schema follows those directions but does not yet represent conserved quantity. Accommodation introduces an invariant observer  $q$  for which the pouring direction  $v$  is null:

$$Tq \circ v = 0_Q \circ q \circ x,$$

even though other observers vary. The child has not merely fitted a better coefficient; the representation now quotients a family of appearances by a quantity-preserving transformation.

Object permanence gives the complementary case. Occlusion removes the object from the immediate visual observer without removing it from the maintained world model [Piaget, 1954]. Accommodation introduces a persistent state component whose identity is transported across visible and occluded contexts. Drescher's Schema Mechanism provides an early computational analogue: updates of predictive schemas use existing items, while the construction of a synthetic item enlarges the available representational vocabulary [Drescher, 1986, 1991].

#### Design principle

Attempt assimilation and proto-accommodation before a finite theory change. Use  $\Theta_{\text{int}}$ , or the classical observer  $\Omega_{\text{cre}}$ , to test whether their local orders are compatible. Accommodate the declaration only when a persistent, independently checked obstruction shows that the maintained geometry is inadequate; then record exactly which part of that geometry changed.

### 3.9 The three modes in differential form

Leung's scaffolding does not replace Boden's modes. It clarifies the geometry within which they operate.

| Mode             | Differential reading                                                                                                       | Characteristic risk                                                          |
|------------------|----------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------|
| Combinational    | Compose registered objects, operations, or probe shapes using the available monoidal structure.                            | A formally legal tensor or composition may have no semantic value.           |
| Exploratory      | Traverse and compare variations realized by $F(A)$ while the ambient category and probe semantics remain fixed.            | Local reachability can be mistaken for completeness of the conceptual space. |
| Transformational | Use persistent defects in local profiles to motivate a finite change of ambient theory, interpretation, or tangent action. | The local evidence underdetermines the global extension.                     |

Table 3.1: Boden’s three creativity modes interpreted through tangent structure. The correspondence is diagnostic, not an identification of creativity with differentiation.

Combinational creativity is especially compatible with the monoidal picture: probe shapes and admissible constructions can be composed. Exploratory creativity studies the neighborhood structure generated by those probes. Transformational creativity begins when the neighborhood is persistently inadequate. It may alter the base declaration, enlarge the observer family, or replace the available differential structure.

### 3.10 Differential theory extensions

Let the current and proposed theories be presented by finite sketches  $\mathbb{S}$  and  $\mathbb{S}^+$ , with a typed extension

$$J : \mathbb{S} \longrightarrow \mathbb{S}^+.$$

Suppose their semantic worlds are described by

$$F : \text{Weil}_1^{\text{tan}} \longrightarrow \text{End}(\mathcal{M})$$

and a proposed extension by

$$F^+ : \text{Weil}_1^{\text{tan}} \longrightarrow \text{End}(\mathcal{M}^+).$$

The sketch extension induces or is supplied with an appropriate semantic transport  $\bar{J} : \mathcal{M} \rightarrow \mathcal{M}^+$ . To compare the old and new local geometries, we also need coherent comparison maps

$$\chi_A : \bar{J}F(A) \Longrightarrow F^+(A)\bar{J}, \quad A \in \text{Weil}_1^{\text{tan}}.$$

**Definition 3.9** (Candidate differential theory extension). A candidate differential theory extension consists of a finite sketch map  $J : \mathbb{S} \rightarrow \mathbb{S}^+$ , a semantic transport  $\bar{J}$ , and a monoidally coherent family of comparison maps  $\chi_A$  relating the old and new Weil actions. It becomes an admitted differential theory extension only after the transport claim and the extension’s new consequences pass the registered independent tests of Chapter 2.

For each probe  $A$  and theory state  $X$ , the comparison has the form

$$\begin{array}{ccc} \mathcal{M} & \xrightarrow{F(A)} & \mathcal{M} \\ \bar{J} \downarrow & & \downarrow \bar{J} \\ \mathcal{M}^+ & \xrightarrow{F^+(A)} & \mathcal{M}^+ \end{array}$$

The square is equipped with the 2-cell

$$\chi_{A,X} : \bar{J}F(A)X \longrightarrow F^+(A)\bar{J}X,$$

which measures whether “probe then transport” agrees with “transport then probe.”

If every  $\chi_A$  is invertible, the extension preserves the selected infinitesimal geometry strongly. If not, the discrepancy can reveal genuinely new directions, collapsed distinctions, or a failure to transport old evidence. Such a discrepancy is not automatically an error. Transformational creativity may intentionally change the geometry; the admission record must say which changes are licensed.

#### Creative transition

The creative jump is not  $F(A)X$ ,  $\Theta_{\text{int}}$ , or  $\Omega_{\text{cre}}$ . Its candidate is a finite package comparison  $Y$ . When the presentation changes, that comparison includes the finite sketch map  $J$ , semantic transport  $\bar{J}$ , and an account  $\chi$  of how infinitesimal meaning would change across the jump. The double repair geometry supplies local evidence;  $Y$  records the global proposal, with a differential theory extension as its presentation-changing case. Only independent admission promotes that proposal to an achieved change.

### 3.11 A differential creativity process

The resulting process refines the status-bearing extension dossier of Chapter 2:

1. *Declare* a conceptual differential universe  $(\mathcal{M}, F)$ , its interpretation, observers, and repair double vector bundle.
2. *Select* a finite family of Weil probes appropriate to the claim.
3. *Realize and compose* the probes at the current theory state; when integration semantics are available, evolve their assimilatory and proto-accommodative directions as partial curve-object flows.
4. *Detect* instability, nonclosure, or observer-relative null directions in the resulting Weil profile.

5. *Compare repair orders* through  $\Theta_{\text{int}}$ , or by estimating its classical realization  $\Omega_{\text{cre}}$ , and localize any mixed effect in the core.
6. *Assimilate if possible* by testing repairs expressible in the maintained differential universe, including compatible proto-accommodative probes.
7. *Integrate the candidate repair* by testing whether it persists and composes into a coherent finite path, rollout, or other declared realization inside the maintained theory.
8. *Control and propose if necessary* by testing registered no-change, parameter, in-language, measurement, observer, and domain-specific alternatives. If none explains the admitted finite realization, propose a package comparison  $Y : \mathfrak{R} \rightarrow \mathfrak{R}^+$ . A finite sketch map  $J : S \rightarrow S^+$  records its presentation component when one exists; probe or observer changes are named separately.
9. *Transport* old models and local meaning through the semantic component of  $Y$ , including  $\bar{J}$  and the comparison family  $\chi$  when the presentation changes.
10. *Admit, reject, revise, or defer* the package proposal using proof, simulation, or other independent evidence.

This process is cyclic. An admitted package change determines a new ambient world and hence a new interpretation of subsequent probes. Creativity changes not only what the system believes, but what it can vary, compare, and ask next.

### 3.12 Beyond two directions: the higher-DIAL conjecture

The construction so far separates two local repair directions and then records a finite theory change outside their infinitesimal geometry. Once that base case is explicit, two different forms of iteration become visible. The iterated tangent endofunctor in LINCOS produces a tower

$$X, \quad TX, \quad T^2X, \quad \dots$$

whose successive levels expose higher-order behavior of one maintained tangent structure [Mahadevan, 2026e]. DIAL suggests a related but different iteration. Its two sides are not merely two applications of  $T$ : they are two *typed classes of repair*, interpreted here as assimilation and proto-accommodation, together with a law governing their interchange. Adding another tangent level and adding another repair direction must therefore be distinguished.

This is a bounded conjectural extension of the two-direction architecture developed above. No later algorithm or empirical claim depends on it.

*Tangent depth.* For an existing DIAL object  $\mathbb{D}$ , the sequence

$$\mathbb{D}, \quad T\mathbb{D}, \quad T^2\mathbb{D}, \quad \dots$$

probes higher jets of the same two repair processes. At  $T^2\mathbb{D}$ , for example, one may ask how the assimilatory and proto-accommodative fields accelerate, curve, or change one another. This is the direct analogue of the LINC iteration. It does not create a third semantic mode of repair.

*Repair arity.* A genuinely higher DIAL object would instead contain  $n$  independently typed infinitesimal directions. The natural syntactic candidate is

$$\mathbb{D}_{\text{inv}}^{(n)} := \underbrace{\mathbb{I}_{\text{Lie}} \otimes \cdots \otimes \mathbb{I}_{\text{Lie}}}_{n \text{ factors}}, \quad \mathbb{D}_{\text{inv}}^{(1)} = \mathbb{I}_{\text{Lie}}, \quad \mathbb{D}_{\text{inv}}^{(2)} = \mathbb{D}_{\text{inv}}.$$

Associativity of the appropriate enriched tensor product would give the iterated-model reading

$$\text{Mod}\left(\mathbb{D}_{\text{inv}}^{(n)}, \mathcal{M}\right) \simeq \text{Mod}\left(\mathbb{I}_{\text{Lie}}, \text{Mod}\left(\cdots \text{Mod}\left(\mathbb{I}_{\text{Lie}}, \mathcal{M}\right) \cdots\right)\right).$$

This formula is presently a proposed presentation. It presupposes the enriched sketch doctrine, tangent stability, pullbacks, and dualizability conditions already missing in the double case.

**Definition 3.10** (Candidate  $n$ -fold DIAL object, proposed). For  $n \geq 1$ , a candidate  $\text{DIAL}^{(n)}$  object is a model of  $\mathbb{D}_{\text{inv}}^{(n)}$  with one involution-algebroid structure in each coordinate, a DIAL interchange comparison on every two-dimensional face, and coherent comparisons among different orders of permuting the coordinates. Its smooth finite-rank realization should be an appropriate  $n$ -fold Lie-algebroid object. This is a research target, not an established internal construction.

The first new case,  $n = 3$ , would be cubical rather than square. Its three pairwise DIAL faces would have to satisfy braid-like coherence, and a weak or noisy realization could contain a highest-core defect invisible on every individual face. Pairwise bracket residuals would therefore not by themselves certify a triple object; a higher comparison witness would be required. The book records this possibility but does not use it in any later algorithm or experiment.



**Boundary: Higher-DIAL conjecture**

There should be a coherent theory of  $n$ -fold involution algebroids internal to a suitable tangent category such that: its  $n = 1$  models recover involution algebroids; its  $n = 2$  models recover DIAL; its strict smooth finite-rank realization agrees with the appropriate  $n$ -fold classical Lie-algebroid geometry; and its pairwise and higher comparison cells admit typed obstruction observers. Neither existence of this internal theory nor the claimed realization equivalence is established here.

The distinction between the two towers suggests a two-parameter hierarchy

$$T^k(\text{DIAL}^{(n)}), \quad k \geq 0, \quad n \geq 1,$$

where  $k$  measures differential depth and  $n$  measures repair arity. A future coalgebraic construction could ask whether repeated increases in depth or arity stabilize, but no higher-DIAL endofunctor or final-coalgebra theorem is asserted in this book.

The possible semantic value of a third direction is nevertheless clear. Besides repairing a realization and revising its theory, a discovery system may need to revise the observer or admission doctrine that determines what counts as evidence. A triple object could type those three processes separately and test whether their orders agree. That interpretation is only a motivation: the third coordinate must be declared by an application rather than inferred from the integer 3.

**Design principle**

No empirical force should be attributed to higher DIAL until controlled realizations of ordinary DIAL show that two typed directions and their interchange witness add value beyond appropriate controls. A claim for  $\text{DIAL}^{(3)}$  has the additional burden of exhibiting a reproducible cubical obstruction that no pairwise DIAL model can represent.

### 3.13 *What the differential account does not claim*

#### Boundary

Leung's Weil theory provides algebraic semantics for infinitesimal probes. Involution algebroids provide the established one-direction bridge, while classical double Lie algebroids organize two compatible classes of smooth variation. Involution algebroids replace Jacobi at the internal level by a Yang–Baxter-style flip law, from which Jacobi is recovered on sections. The DIAL bridge remains to be constructed. None of these structures supplies domain meaning, a novelty criterion, or a unique finite extension. A tangent or double-infinitesimal geometry can formalize local possibility without proving that every historical creative jump decomposes into infinitesimal steps.

Four limitations remain visible. First, tangent structure is relative to a chosen category and interpretation. A missing object or observer may be invisible to every current probe. Second, the proposed split into assimilatory and proto-accommodative directions is semantic data, not a canonical decomposition supplied by a double vector bundle. Third, local differential evidence usually underdetermines the global package change. Fourth, preservation of tangent structure is only one admission criterion: a scientifically creative theory must also survive substantive formal, simulator-relative, or empirical tests.

These limitations are productive. They identify the division of labor used in the rest of the book: Weil probes expose local alternatives, DIAL repair geometry diagnoses their interaction, registered controls test competing explanations, frontier models propose, a finite package comparison records the candidate global change, and independent procedures admit or reject it. When the presentation changes, sketch surgery records that component. Chapter 4 next supplies an anchored geometry for the skills that conduct this exploration.

#### *Further Reading*

Leung's classification gives the precise monoidal relationship between tangent structures and the category of Weil algebras [Leung, 2017, 2018]. Cockett and Cruttwell develop tangent categories as an axiomatic setting encompassing classical and synthetic differential geometry [Cockett and Cruttwell, 2014]. Leung's account also suggests viewing  $\text{Weil}_1^{\text{tan}}$  as a monoidal theory for tangent structure, with richer categories of Weil algebras offering probes of higher finite order. Burke and MacAdam's involution algebroids and MacAdam's functorial seman-

tics supply the first-order bridge from tangent categories to Lie theory [Burke and MacAdam, 2019, MacAdam, 2023]. Drinfel'd's Hamiltonian interpretation of the classical Yang–Baxter equation provides the distinct Poisson–Lie and Lie-bialgebraic lineage [Drinfel'd, 1983]. Piaget's assimilation–accommodation distinction gives the developmental lineage for separating change inside a maintained schema from change of the schema itself [Piaget, 1952, 1985]. Drescher supplies the direct computational precedent through predictive schema refinement and synthetic item construction [Drescher, 1986, 1991]. Mackenzie develops classical double Lie algebroids; Meinrenken and Pike construct their distinct bigraded cochain Weil algebras and characterize double compatibility by commuting horizontal and vertical differentials [Mackenzie, 2005, Meinrenken and Pike, 2021, Pike, 2020]. Cockett, Cruttwell, and Lemay develop differential equations, complete vector fields, flows, and commuting dynamics in tangent categories, with curve objects supplying the abstract parameter of evolution [Cockett et al., 2021].



## 4

# *Skill Optimization for Creative Exploration*

## MARKDOWN PROGRAMS AND LIE ALGEBROID GEOMETRY

Chapter 3 supplied probe shapes for local variation. It did not yet say how an artificial agent should represent and select the operations that move a theory through those variations. This chapter adds that operational layer. In the proposed realization, a creative workflow is written as a typed Markdown program; its executable interpretation is a skill; and a Lie algebroid describes how locally available skills act on the current theory state.

The phrase “Markdown-parameterized search space” must be read carefully. Raw token strings do not form a smooth manifold merely because they describe an experiment. Markdown supplies a human-readable control language. The geometry belongs to the semantic theory states and to the anchored bundle of skills compiled from that language.

The proposal adapts LASKO—skill optimization over Lie algebroids—from artifact editing to conceptual exploration [Mahadevan, 2026d].

### Design principle

Markdown declares typed control artifacts. A semantic compiler turns a local skill artifact into an available section, while a distinct policy compiler turns a workflow artifact into choices among typed actions. The anchor map reveals the change that executing a selected skill can produce in the maintained theory. The scripts parameterize exploration without being identified with their differential geometry.

The first sections deliberately study one Lie algebroid  $A \rightarrow M$ . This is the operational unit inherited from LASKO, not yet the full double DIAL object. In DIAL-SkillOpt the construction is applied sidewise to  $A_{\text{asm}}$  and  $B_{\text{acc}}$ ; their double repair geometry supplies the mixed interchange and core information that no single algebroid carries.

#### 4.1 From a script to a creative skill

For one fixed language and interface version, let  $\text{Md}_{\text{adm}}$  denote a category of admissible, typed Markdown artifacts. Its objects include front matter, ordered workflow steps, tool declarations, invariant checks, probe requests, and admission rules. Its arrows are controlled revisions that preserve that version's required interface. A revision of the grammar, type system, or interface instead requires a registered map between versioned artifact categories; it is not silently treated as another arrow of the old category. A typical contract might read:

We distinguish two typed subcollections. A local skill artifact belongs to  $\text{Md}_{\text{sec}}$  and compiles to a section. A discovery-controller artifact belongs to  $\text{Md}_{\text{pol}}$  and compiles to a policy over admissible actions. Both sit inside  $\text{Md}_{\text{adm}}$  and may share syntax, but their semantic codomains differ.

```

---
skill: analogical_transfer
reads: [theory.objects, obstruction.local]
writes: [candidate.interfaces]
invariants: [source_provenance, admitted_results]
probes: [first_order, paired_order]
admission: held_out_counterexample
---
1. Select a source construction with a typed structural match.
2. Propose the smallest target interface needed by the analogy.
3. Record transported laws and laws that are not preserved.
4. Generate a discriminating consequence and send it to admission.

```

The header declares what the skill may read and change. The body describes a proposal policy. Neither is executable until interpreted relative to an environment  $E$ , which includes the parser, model, tools, theory-store version, and authorization policy. We write the resulting partial semantic map as

$$\text{compile}_E^{\text{sec}} : \text{Ob}(\text{Md}_{\text{sec}}) \rightarrow \Gamma_{\text{loc}}(A), \quad m \mapsto s_m.$$

Here  $A \rightarrow M$  is a Lie algebroid over a smooth realization  $M$  of theory states, and  $s_m$  is the local skill section compiled from script  $m$ . The map is partial because a script may be ill typed, request an unavailable tool, or fail its authorization contract.

A local section is a state-dependent choice of one skill direction in each fiber over the region where the skill is defined. Thus  $s_m$  is not merely the Markdown text or a fixed parameter vector: it says which typed operation the compiled program makes available at each compatible theory state. The anchor then converts that internal choice into an observable motion of the base state.

This notation does not assert that every tangent category is already a smooth manifold or possesses a Lie algebroid. It selects a smooth or suitably synthetic realization in which the LASKO construction is

meaningful. Other realizations may require an internal or generalized analogue of the same anchored structure.

#### 4.2 The Lie algebroid of creative moves

A Lie algebroid consists of a vector bundle

$$\pi : A \longrightarrow M,$$

a Lie bracket  $[\cdot, \cdot]_A$  on sections, and an anchor

$$\rho : A \longrightarrow TM$$

satisfying, for sections  $s, t$  and  $f \in C^\infty(M)$ ,

$$[s, ft]_A = f[s, t]_A + \rho(s)(f)t, \quad \rho([s, t]_A) = [\rho(s), \rho(t)],$$

the Leibniz rule and anchor-bracket compatibility [Mackenzie, 2005]. At a theory state  $x \in M$ , the fiber  $A_x$  contains locally available, typed skill directions. For a compiled section  $s_m$ , execution produces the infinitesimal theory motion

$$\dot{x} = \rho(s_m(x)).$$

When this vector field integrates over an interval, it produces a finite rollout  $x \mapsto \Phi_m^t(x)$ . The rollout may edit a conjecture set, add a candidate interface, select an experiment, compare two mechanisms, or update a status-bearing extension dossier.

The three pieces have different semantics:

*Base M.* The maintained semantic states of the current theory.

*Fiber  $A_x$ .* Skills available locally at state  $x$ , including their type, permissions, and provenance requirements.

*Anchor  $\rho$ .* The observable change in the theory state induced by executing a skill direction.

The vector-space structure in each fiber is a modeling commitment, not a property of skill names. It is justified only when local coefficients and linear combinations have an operational semantics—for example as mixtures, rates, or a declared continuous relaxation. Otherwise the Lie algebroid is a local surrogate for the compiled effects, and the compiler must record which discrete executions realize the sections being compared.

This is richer than placing independent coordinates on a prompt. Availability can vary with the state: a proof skill may become legal only after the necessary definitions exist; a simulator intervention may be available only where its mechanism is exposed; and a proposed analogy may be inadmissible until its source and target types have been registered.

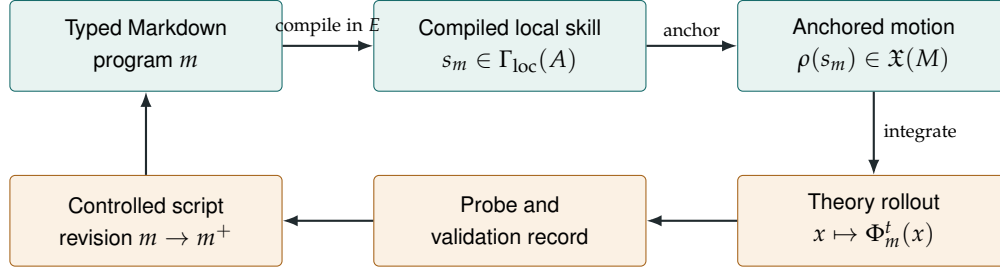


Figure 4.1: The two-level optimization loop. The object-level loop executes an anchored skill on a theory state. The meta-level loop uses probe and validation records to revise the typed Markdown controller; locked admission remains outside this revision loop.

### 4.3 Brackets expose interaction

Creative skills are rarely independent coordinates. Applying analogy before counterexample search can yield a different candidate theory from reversing the order. The Lie bracket records the infinitesimal interaction between two skill sections. If  $m$  and  $n$  compile to  $s_m$  and  $s_n$ , then

$$[\rho(s_m), \rho(s_n)]$$

is the leading infinitesimal order effect obtained from a small commutator loop; the finite loop discrepancy is second order in its side lengths. In an exact Lie algebroid,

$$\rho([s_m, s_n]_A) = [\rho(s_m), \rho(s_n)].$$

This bracket is not a textual difference between two scripts. It compares the state-dependent vector fields produced by their compiled semantics. Two very different scripts can therefore have a negligible bracket when their realized actions commute. Conversely, two small edits can have a large bracket when their order changes what later operations become available. The coordinate construction in Section 3.6 gives the minimal picture: two registered motions generate a third direction only through their order of composition.

The exact identity above belongs to the declared Lie-algebroid model. In a learned implementation, it must itself be checked. The residual

$$r_A(m, n) = [\rho(s_m), \rho(s_n)] - \rho([s_m, s_n]_A)$$

measures failure of the estimated bracket and anchor to satisfy this exact law. It is a model-diagnostic residual, not an additional Lie-algebroid axiom.

Suppose a declared distribution  $\mathcal{D} \subseteq TM$  contains the visible theory changes currently regarded as admissible. On a constant-rank region, choose a local splitting or metric and hence a projection  $P_{\mathcal{D}}$ . The closure residual

$$r_{\text{vis}}(m, n) = (I - P_{\mathcal{D}})[\rho(s_m), \rho(s_n)]$$

tests whether interacting legal skills generate a direction outside the present declaration. Its numerical representative depends on the chosen



projection, whereas vanishing of the quotient class modulo  $\mathcal{D}$  does not. A persistent residual may indicate a missing composite skill, an omitted state variable, an inadequate observer, or a genuine need to extend the theory. It does not by itself choose among these explanations.

Brackets also provide a practical screening rule. Many inexpensive script pairs can be tested with short rollouts; pairs with large order effects or closure defects receive the expensive proof, simulation, or empirical validation budget. The bracket therefore shapes exploration without reducing admission to a scalar objective.

#### 4.4 *Assimilation and accommodation in skill geometry*

Chapter 3 described assimilation as change inside a maintained differential universe and accommodation as a typed change of that universe. The Lie algebroid makes this distinction operational. Fix the creative skill geometry

$$A \xrightarrow{\pi} M, \quad \rho : A \longrightarrow TM, \quad [\cdot, \cdot]_A,$$

together with its Markdown language and compiler. A possibly time-dependent family of compiled local sections  $s_t$  generates an assimilatory trajectory when

$$\dot{x}_t = \rho(s_t(x_t))$$

and the trajectory, its skill compositions, and its admission tests remain expressible in this fixed package.

**Definition 4.1** (Lie algebroidal assimilation). An assimilatory skill update is an admitted rollout generated by sections of the maintained Lie algebroid, using the current anchor, bracket, base-state representation, and semantic compiler. It changes the theory state  $x \in M$  without changing the geometry that specifies which skills are locally available or how their interactions act on  $M$ .

This definition is geometric rather than metric. A long rollout or a large finite displacement can still be assimilation if it is assembled from admissible flows in the existing algebroid. Conversely, a small textual edit can be proto-accommodative if it proposes a new state variable, skill type, or interaction law; it becomes accommodation only after finite realization and admission.

Pressure for proto-accommodation becomes credible when persistent evidence cannot be represented by the current anchored skill system. Even then, the diagnostics motivate an extension; they do not specify or admit it. Typical diagnostics include:

- the required theory motion lies outside  $\text{im } \rho$ ;

- available sections fail to generate a needed direction under brackets;
- a learned anchor–bracket residual persists after estimation error and observer misspecification have been checked;
- distinct mechanisms collapse to the same base state or anchor image;  
or
- the Markdown compiler cannot type a skill required by the proposed repair.

The candidate extension is represented by an extended algebroid

$$A^+ \xrightarrow{\pi^+} M^+, \quad \rho^+ : A^+ \longrightarrow TM^+,$$

and transport data consisting at least of a semantic base map  $\bar{J} : M \rightarrow M^+$  and a bundle map  $\phi_A : A \rightarrow A^+$  over  $\bar{J}$ .

Here  $J : \mathbb{S} \rightarrow \mathbb{S}^+$  is the presentation map and  $\bar{J}$  its induced or supplied semantic transport; they are not the same typed object. The anchor comparison

$$\rho^+ \circ \phi_A \stackrel{?}{=} T\bar{J} \circ \rho$$

tests whether old skills retain their visible meaning under the proposed extension. Bracket preservation supplies a second comparison whenever the relevant sections are  $\bar{J}$ -related. If the standard compatibility conditions hold,  $(\phi_A, \bar{J})$  is a Lie algebroid morphism. Requiring a morphism is appropriate for conservative transport, but not every creative extension should preserve all old interactions strictly; any failure must instead be recorded and justified.

**Definition 4.2** (Admitted Lie algebroidal accommodation). An admitted accommodative skill update is a typed revision of the base  $M$ , skill bundle  $A$ , anchor  $\rho$ , bracket, Markdown language, or compiler, motivated either by a persistent obstruction that no admitted rollout in the maintained algebroid accounts for, or by a registered opportunity whose value cannot be expressed in that algebroid. Its audit record includes the extended algebroid, the transport pair  $(\phi_A, \bar{J})$ , the comparison defects, and an independent admission result that warrants retaining the revision.

Piaget’s conservation example has a particularly simple anchored form. Once quantity is represented by an observer  $q : M \rightarrow Q$ , pouring skills  $s$  should satisfy

$$Tq(\rho(s(x))) = 0.$$

Before accommodation, the state space or observer family may encode only height and vessel shape, so this invariant direction is not expressible. Accommodation refines the base or its observers so that

quantity-preserving motions become identifiable; later pours are then assimilated as trajectories tangent to the level sets of  $q$ . For object permanence, accommodation similarly enlarges  $M$  with a persistent latent object state and lifts visible and occlusion-handling skills into  $A^+$ . Subsequent occlusions can then be assimilated within the enlarged geometry.

#### Design principle

Search first for an assimilatory rollout in the current Lie algebroid. Propose a finite extension only after testing whether an apparent nonclosure, missing direction, or anchor defect is caused by finite data, an inadequate observer, or a compilation error. When extension is necessary, preserve the old skill geometry through explicit transport wherever the evidence warrants it.

### 4.5 Latent skill variation

The kernel  $\ker \rho$  contains skill directions with no immediate visible displacement in  $M$ . These directions are not necessarily irrelevant. Two executions can produce indistinguishable Markdown artifacts while differing in tool routing, evidence provenance, intermediate countermodels, or future composability. Such latent differences can interact with later skills and become observable.

This explains why a final text diff is an inadequate record of creative optimization. A reproducible experiment must preserve at least

$$(m, E, s_m, x, \text{trace}, \text{probes}, \text{admission status}).$$

In particular, identical Markdown files compiled under different model, parser, or tool versions need not denote the same section. The environment is part of the semantics.

### 4.6 Two coupled optimization problems

The framework separates two levels that are easily conflated.

*Object-level exploration.* For a fixed script language, compiler, and Lie algebroid, the agent selects and composes sections to move through  $M$ . Continuous coefficients, when the skill fiber supports them, can be optimized by geometric or natural-gradient methods. Discrete skill choices can be screened by rollouts and bracket proxies.

*Meta-level skill optimization.* The system revises the Markdown program

$$m \longrightarrow m^+$$

using observed failures, traces, and evaluation outcomes. Within a benchmark version, validation outcomes drive edits; released admission outcomes may become training evidence only in a new, explicitly versioned cycle. This is a controlled program transformation, not an infinitesimal vector in raw token space unless an additional continuous relaxation and its semantics have been declared. A revision can change the selected section, its domain of availability, or the permissions under which it acts.

The creative optimization landscape is consequently not a scalar function over prompts. It is the structured package

$$\mathcal{L}_{\text{cre}} = (M, A, \rho, [\cdot, \cdot]_A, \text{Md}_{\text{adm}}, \text{compile}_E^{\text{sec}}, \Lambda, \mathcal{R}, \mathcal{A}),$$

where  $\Lambda$  contains observers,  $\mathcal{R}$  records typed residuals, and  $\mathcal{A}$  is the family of admission procedures. Scalar scores may rank candidates inside this package, but they do not define its types, legal moves, or certificates.

#### 4.7 DIAL-SkillOpt: training a discovery skill

The curve-object semantics of Chapter 3 turns the preceding static landscape into a trainable process. SkillOpt treats a skill document as external state for a frozen agent: a separate optimizer uses scored rollouts to propose bounded textual edits, retaining an edit only when it improves held-out validation performance [Yang et al., 2026]. DIAL-SkillOpt adopts that optimization discipline but changes both the state being maintained and the meaning of success.

**Definition 4.3** (DIAL-SkillOpt discovery skill). A DIAL-SkillOpt discovery skill is a versioned, typed control artifact  $m$  whose compilation determines a state-dependent policy over assimilatory skills, proto-accommodative probes, pairwise-order tests, finite theory-extension proposals, and abstention. Training revises  $m$  using bounded edits and replayable discovery traces. The trained artifact is successful only to the extent that it improves independently certifiable theory construction, not merely the score of its generated prose.

The object being optimized is therefore a *policy for discovery*. It is not creativity represented as a scalar, nor a differentiable coordinate for the finite sketch map  $J$ . The policy learns how to navigate, diagnose, and propose changes to maintained theories; proof, simulation, or new evidence still decides whether any proposed change is admitted.

##### 4.7.1 Typed state and actions

At decision time, let the discovery state be

$$z_t = (\mathfrak{R}_t, x_t, H_t, b_t),$$

where  $\mathfrak{R}_t$  is the versioned representational package containing the current sketch, DIAL geometry, observers, and compilers;  $x_t$  is its current semantic state;  $H_t$  is the provenance-bearing history; and  $b_t$  is the remaining experiment and computation budget. Compiling a discovery-controller document in environment  $E$  uses the policy interface

$$\text{compile}_E^{\text{pol}} : \text{Ob}(\text{Md}_{\text{pol}}) \rightarrow \Pi_{\text{adm}}$$

and yields a partial policy

$$\pi_m(\text{cdot} \mid z_t) = \text{compile}_E^{\text{pol}}(m)(z_t)$$

over the typed action family

$$\mathcal{U}(z_t) = \{ \text{Assimilate}(s), \text{Probe}(r), \text{Compare}(s, r), \\ \text{Propose}(Y; J, \bar{J}, \phi_A, \chi), \text{RequestEvidence}(e), \text{Abstain} \}.$$

Here  $s$  and  $r$  are locally available sections of the assimilatory and proto-accommodative side bundles. The proposal action first declares the package comparison  $Y$ . When the presentation changes, it also contains a finite sketch map  $J : S \rightarrow S^+$ , its semantic transport  $\bar{J}$ , skill transport  $\phi_A$ , and differential comparison data  $\chi$ ; otherwise those fields are omitted or replaced by the transport data appropriate to the changed package components. Ill-typed or unauthorized actions are absent rather than merely assigned a large penalty.

Within a fixed theory fiber, the anchors realize the selected sections as vector fields. A curve object integrates them, when possible, to partial flows  $\Phi_{\text{asm}}$  and  $\Phi_{\text{acc}}$ . For a small probe scale  $\epsilon$  and observer  $O$ , an empirical mixed witness is

$$\hat{\Delta}_\epsilon(s, r; x) = O(\Phi_{\text{acc}}^\epsilon \Phi_{\text{asm}}^\epsilon(x)) - O(\Phi_{\text{asm}}^\epsilon \Phi_{\text{acc}}^\epsilon(x)).$$

The normalized small-loop behavior estimates an observer-relative bracket or interchange effect. DIAL-SkillOpt learns when such a comparison is worth its cost and how to respond to a persistent, uncertainty-qualified defect. It is not trained to regard every nonzero finite-sample loop as accommodation.

#### 4.7.2 One discovery episode

An episode operates on one problem drawn from a registered task population. It may terminate with an assimilatory solution, a typed extension dossier  $\Xi$ , a request for evidence, or an explicit abstention. When an extension is proposed, the episode uses Chapter 2's status-bearing dossier

$$\Xi = ((\mathfrak{R}, D), \omega, \gamma, Y, (\mathfrak{R}^+, D^+), \tau, \Pi, \sigma),$$

where  $Y$  records the changed package components, including  $J$  when the presentation changes, and the skill-specific transport record  $\tau$  contains  $(\bar{J}, \phi_A, \chi)$  and the transported evidence. The episode may set  $\sigma = \text{proposed}$ ; it cannot assign admitted status itself.

---

**DIAL-SkillOpt discovery episode**

- 1: Load  $z_0 = (\mathfrak{R}_0, x_0, H_0, b_0)$  and type-check the compiled policy  $\pi_m$ .
  - 2: **while** budget remains and no terminal record has been produced  
  **do**
  - 3:   Select an admissible action from  $\mathcal{U}(z_t)$ .
  - 4:   Execute its partial flow, paired-order probe, or evidence request;  
  append the complete trace to  $H_t$ .
  - 5:   Estimate uncertainty and test whether observed defects persist  
  under controls, repeat probes, and changes of presentation.
  - 6:   **if** the current geometry supports an adequate repair **then**
  - 7:     assimilate within  $\mathfrak{R}_t$  and continue or terminate;
  - 8:   **else if** the obstruction is localized but not repairable in the  
  current geometry **then**
  - 9:     assemble the status-bearing dossier  $\Xi$  with  $\sigma = \text{proposed}$ ;
  - 10:   **else**
  - 11:     request evidence or abstain without inventing an unsupported object.
  - 12:   **end if**
  - 13: **end while**
  - 14: Return the terminal record, trace, resource use, and internal  
  precheck vector.
- 

The policy does not issue the final scientific verdict. It can run inexpensive internal checks and construct the evidence package, but an admission procedure withheld from the episode assigns its epistemic status.

#### 4.7.3 The training loop

Let  $\mathcal{E}_{\text{train}}$ ,  $\mathcal{E}_{\text{val}}$ , and  $\mathcal{E}_{\text{adm}}$  be disjoint problem sets. Training rollouts produce traces for proposing edits. Validation accepts or rejects edits. The admission set remains locked until the skill document has been frozen. A benchmark may expose admission outcomes only after the training run is over; using those same cases for another update creates a new benchmark version.

For a skill document  $m$ , define a validation profile

$$Q_{\text{val}}(m) = (q_{\text{type}}, q_{\text{loc}}, q_{\text{proposal}}, q_{\text{precheck}}, q_{\text{abstain}}, -q_{\text{cost}}).$$

The components measure type safety, obstruction localization, proposal quality, internal precheck success, calibrated abstention, and resource

use. They remain visible separately. A declared preorder  $\succ_{\mathcal{G}}$  first enforces hard gates  $\mathcal{G}$ —including false-extension and regression limits—and only then compares eligible profiles. A bounded edit  $m \rightarrow m'$  is accepted exactly when

$$\mathcal{G}(m') = 1 \quad \text{and} \quad Q_{\text{val}}(m') \succ_{\mathcal{G}} Q_{\text{val}}(m).$$

This generalizes SkillOpt’s strict held-out improvement rule without hiding scientific validity inside a weighted sum.

---

#### Training DIAL-SkillOpt

- 1: Initialize frozen discovery agent  $F$ , typed skill document  $m_0$ , optimizer  $G$ , rejected-edit buffer  $B$ , and disjoint task splits.
  - 2: **for** epochs  $k = 0, \dots, K - 1$  **do**
  - 3:   Run  $F[m_k]$  on  $\mathcal{E}_{\text{train}}$ ; record DIAL states, flows, mixed witnesses, proposals, abstentions, and resource use.
  - 4:   Let  $G$  propose bounded add/delete/replace edits using the traces, current validation profile, and  $B$ .
  - 5:   Type-check each candidate and evaluate it on  $\mathcal{E}_{\text{val}}$  with matched budgets and fresh seeds.
  - 6:   Accept the best candidate satisfying the hard gates and strict profile-improvement rule; place rejected edits and counterexamples in  $B$ .
  - 7:   Apply a slow meta-update to the edit strategy without changing the frozen discovery agent or reading  $\mathcal{E}_{\text{adm}}$ .
  - 8: **end for**
  - 9: Freeze  $m_*$ , then evaluate it once on  $\mathcal{E}_{\text{adm}}$  using independent proof, simulator, or empirical certificates.
  - 10: Publish  $m_*$ , all versions, traces, split hashes, admission vector, and failures.
- 

When a fixed local skill basis admits continuous coefficients  $w$ , an inner optimizer may use a geometry-aware step such as

$$w_{k+1} = w_k + \eta G(w_k)^{-1} \nabla_w \tilde{U}_{\text{val}},$$

where  $G$  is a declared metric and  $\tilde{U}_{\text{val}}$  is a validation surrogate. This is legitimate for motion inside the maintained skill geometry. Bounded Markdown edits, new skill types, and finite sketch extensions remain discrete typed operations; the equation does not make them ordinary gradient steps.

**Admission contract**

The optimizer may inspect training traces, validation profiles, rejected edits, and provisional checks. It may not query the locked admission cases, alter their simulators or proof obligations, or treat its own confidence as a certificate. The final report separates proposal success, admission success, false extension, calibrated abstention, transfer, and cost.

*4.7.4 A preregistered evaluation design*

A minimal evaluation trains one external discovery skill across registered synthetic worlds. Training worlds expose exact or simulator-checkable certificates; validation uses fresh worlds and seeds; admission withholds both the structural defect and its certificate until the skill is frozen. Transfer to categorical symbolic regression or simulator-grounded theory extension is a separate, stronger claim rather than part of the in-distribution result.

The minimum comparison includes:

1. the frozen agent without a skill document;
2. a manually written discovery skill;
3. ordinary SkillOpt using only terminal task score;
4. fixed-geometry LASKO skill optimization without accommodation;
5. DIAL execution with an untrained discovery skill;
6. DIAL-SkillOpt without the mixed witness; and
7. complete DIAL-SkillOpt.

Primary outcomes are locked-set admission rate, false-extension rate, correct localization, experimental cost, and abstention calibration. Secondary outcomes include time to a valid proposal, preservation of old results after transport, transfer to a new task family, and the fraction of admitted extensions that remain useful in later episodes. Shuffled-witness and equal-budget controls test whether improvement comes from the double geometry rather than extra information or computation.



**Design principle**

Train the skill that conducts discovery, not the admission rule that declares truth. Curve-object flows make local exploration reproducible; DIAL makes the two repair modes and their interaction observable; SkillOpt makes the external controller iteratively improvable. A finite theory extension remains a typed proposal whose scientific standing comes from evidence the optimizer could not edit.

#### 4.8 *Boden's modes as changes of control geometry*

The construction provides a common language for Boden's three modes without collapsing their differences.

*Combinational creativity.* Registered scripts or skill sections are composed. Their bracket reveals whether ordering introduces a genuinely new effect.

*Exploratory creativity.* Sections and their parameters are optimized inside a fixed  $(M, A, \rho)$ , Markdown grammar, and semantic compiler.

*Transformational creativity.* The system changes the grammar or compiler, enlarges the admissible skill bundle, or proposes a finite base change  $M \rightarrow M^+$ . Chapter 3's comparison maps must then explain how old probes and skills transport to the new world.

This classification draws the decisive boundary. Optimizing a workflow inside a fixed skill geometry is exploratory, even when the resulting theorem or mechanism is novel. Changing what counts as a legal skill or a meaningful theory state is transformational. A Markdown revision can do either; its creativity mode is determined by its semantics, not by the number of edited tokens.

#### 4.9 *Scope and experimental consequences*

**Boundary**

The LASKO construction models controlled creative exploration, not creativity in its entirety. It does not claim that every concept is a script edit, that every useful jump integrates a local vector field, or that a large bracket is evidence of scientific truth.

The proposal nevertheless makes several claims testable. One can compare flat prompt optimization with state-dependent skill availability; bracket- screened validation with random validation; artifact-only

logging with trace-sensitive logging; and fixed-script exploration with controlled meta-level revision. The AGENTIC testbeds later in the book provide different certificates for these comparisons.

Most importantly, DIAL-SkillOpt turns this construction into an explicit training framework while preserving an optimization landscape that respects the book’s central distinction. Infinitesimal geometry organizes local exploration. Markdown makes the policy inspectable and revisable. A finite theory extension remains an explicit creative proposal whose value is decided only by independent admission. Chapter 5 next defines the algebraic theory and model semantics that such a proposal is intended to construct. Chapter 12 later turns these ingredients into OPTIC, an integrated architecture for creative skill construction, and reports its initial mechanism evidence.

### *Further Reading*

LASKO develops the interpretation of typed skill artifacts as sections of a Lie algebroid and uses anchors, brackets, and kernel directions to diagnose agentic skill optimization [Mahadevan, 2026d]. Mackenzie provides the standard geometric theory of Lie groupoids and Lie algebroids [Mackenzie, 2005]. Piaget’s account of equilibration supplies the developmental distinction between assimilation and accommodation [Piaget, 1985]; Drescher gives an early computational realization in which schema refinement can be distinguished from construction of new representational items [Drescher, 1986, 1991]. SkillOpt is a complementary example of agents improving external, reusable skill documents through iterative evaluation [Yang et al., 2026].

## 5

# *Algebraic Theories as Creative Targets*

## LAWVERE THEORIES, SKETCHES, PROPS, AND TOPOSES

Chapters 3 and 4 described the local geometry of conceptual variation and the skills that explore it. This chapter specifies the finite artifact toward which those local mechanisms point: a constructed theory. Rather than stopping at an artifact or a list of claims, the system attempts to construct a theory with an explicit syntax, a class of models, and a transport relation to what was known before. This is a chosen scope for DIAL, not a claim that every form of creativity must culminate in an algebraic theory.

The target is not the paper or paragraph that describes a theory. It is the typed collection of operations, relations, and preservation requirements that the prose presents, together with the models in which those requirements can be interpreted and tested. This is why the chapter treats sketches as compact presentations rather than as diagrams added only for exposition.

Lawvere’s functorial semantics supplies the decisive idea: an algebraic theory is a category of abstract operations and equations, while a model of that theory is a structure-preserving functor into a semantic category [Lawvere, 1963]. The same presentation can therefore have models in sets, smooth spaces, vector spaces, sheaves, or another suitable category. What changes with the target is the realization; what persists is the theory’s compositional form.

### Design principle

The principal target in this book is a presented theory together with its functorial model semantics. A sketch records the generators and structural obligations compactly; models witness formal realizability in one or more semantic worlds. They do not by themselves establish empirical truth, explanatory value, or scientific acceptance.

“Any target category” always means any category possessing the products, limits, tensor products, or other structure that the chosen theory asks its models to preserve.

### 5.1 Lawvere's functorial semantics

A one-sorted Lawvere theory may be described as a small category  $\mathbb{L}$  with finite products whose objects are the finite powers

$$1, X, X^2, X^3, \dots$$

of a distinguished object  $X$ . A morphism  $X^n \rightarrow X^m$  is an abstract  $m$ -tuple of  $n$ -ary operations. Composition performs substitution, and equality of morphisms records equational laws.

Let  $\mathcal{C}$  be a category with finite products. A model of  $\mathbb{L}$  in  $\mathcal{C}$  is a finite-product-preserving functor

$$M : \mathbb{L} \longrightarrow \mathcal{C}.$$

Natural transformations between such functors are homomorphisms of models. We write

$$\text{Mod}(\mathbb{L}, \mathcal{C}) = \text{FP}(\mathbb{L}, \mathcal{C})$$

for the resulting model category.

For the theory of monoids, for example,  $\mathbb{L}$  contains abstract maps

$$\mu : X^2 \rightarrow X, \quad e : 1 \rightarrow X,$$

and equations expressing associativity and the two unit laws. A product-preserving model in  $\text{Set}$  chooses a set  $M(X)$ , a multiplication, and a unit satisfying those equations. A model in another Cartesian category interprets the same formal operations using that category's products.

The theory and its models must not be conflated. A theory is not one favored implementation, dataset, or neural parameterization. It is the invariant compositional specification shared by a category of realizations. This makes Lawvere semantics a natural target for persistent creativity: the constructed concept can travel between models.

Many-sorted theories replace the single generator  $X$  by a family of sorts. This is usually the appropriate form for scientific domains: particles, fields, observations, interventions, and experimental contexts need not have the same type.

### 5.2 Sketches present theories

A fully completed category of operations is too large and redundant to be a convenient output format. A sketch presents it by generators and relations. Following Ehresmann, a sketch contains a graph of formal objects and arrows, specified commutative diagrams, and designated cones or cocones [Ehresmann, 1968, Barr and Wells, 1999]. A model

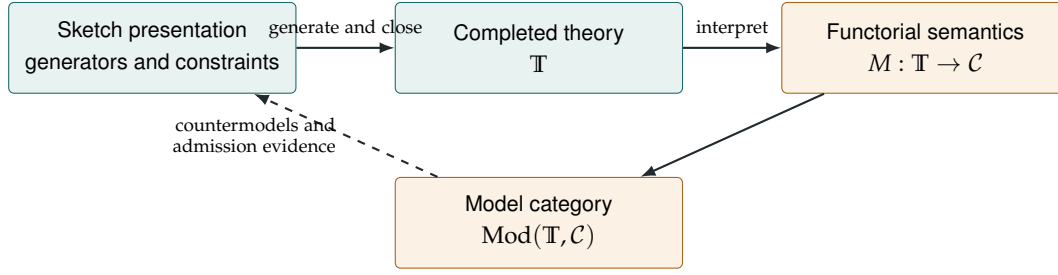


Figure 5.1: A sketch is a compact presentation, not the entire theory. Its completion determines a doctrine-specific category of operations; preserving functors realize the theory in a semantic category.

maps this presentation into a target category and realizes the designated diagrams and universal constructions.

For a finite-product sketch  $S$ , a model

$$D : S \longrightarrow \mathcal{C}$$

must send each designated product cone to a product cone in  $\mathcal{C}$  and satisfy the declared path equations. Closing the presentation under the required finite products and equations yields the corresponding algebraic theory. In this precise sense, the sketch is a finite specification of the theory rather than a second semantic model of it.

This distinction gives synthetic creativity an auditable output format. A frontier model may propose prose, but an admitted theoretical contribution must identify which sorts, operations, equations, cones, or cocones have been added. Its consequences can then be tested across models rather than only in the example that inspired the proposal.

### 5.3 From products to finite limits

Finite products express total, many-input operations. Some theories require more discriminating domains of definition. A finite-limit theory is a small category  $\mathbb{T}_{\text{fl}}$  with finite limits. Its models in a finitely complete category  $\mathcal{C}$  are finite-limit-preserving functors

$$M : \mathbb{T}_{\text{fl}} \longrightarrow \mathcal{C}.$$

A finite-limit sketch presents such a theory by specifying only the equations and finite-limit cones needed as generators.

The extra expressiveness is important. Internal categories, for instance, have an object of arrows, an object of objects, source and target maps, and a composition defined on the pullback of composable pairs. The pullback is not merely a product: it selects pairs whose target and source match. Finite-limit logic can therefore describe essentially algebraic structures with partially defined operations whose domains are themselves specified structurally.

Passing from product preservation to finite-limit preservation is not a cosmetic strengthening. It changes which models count as valid

and which countermodels can refute a candidate theory. A creative system must declare the preservation doctrine it intends; otherwise it can silently move between different notions of model.

#### 5.4 PROPs and monoidal theories

Not every domain permits unrestricted copying and deletion. A Cartesian product automatically supplies diagonals  $X \rightarrow X \times X$  and terminal maps  $X \rightarrow 1$ . Physical, probabilistic, quantum, resource-sensitive, and process-oriented theories often need a tensor product without those automatic operations.

A PROP is a strict symmetric monoidal category whose objects are natural numbers, with tensor on objects given by addition. It is generated under tensor by one object, so a morphism

$$m \longrightarrow n$$

represents an operation with  $m$  inputs and  $n$  outputs. An algebra of a PROP  $\mathbb{P}$  in a symmetric monoidal category  $\mathcal{V}$  is, subject to the usual strictness or coherence convention, a symmetric monoidal functor

$$A : \mathbb{P} \longrightarrow \mathcal{V}.$$

Lack describes PROPs as one-sorted symmetric monoidal theories and shows how distributive laws can compose them [Lack, 2004].

Lawvere theories and PROPs are therefore neighboring doctrines, not stages of one ladder. A Lawvere theory is Cartesian and one-sorted; a PROP is symmetric monoidal and one-sorted. The choice determines which structural maps are free and which must be explicitly generated. In synthetic creativity, discovering that copying is illegal can be as important as discovering a new operation.

| Doctrine                    | Theory structure                                                         | Model contract                                     | What it can make explicit                              |
|-----------------------------|--------------------------------------------------------------------------|----------------------------------------------------|--------------------------------------------------------|
| Lawvere theory              | finite products; usually one or many sorts                               | preserve finite products                           | total operations, substitution, equations              |
| Finite-limit theory         | finite limits                                                            | preserve finite limits                             | typed domains, pullbacks, partial operations           |
| PROP                        | one-generated symmetric monoidal structure                               | preserve tensor and symmetry                       | multi-input/output processes without automatic copying |
| Geometric theory            | finite-limit syntax with arbitrary disjunction and existential structure | interpret geometric logic and its homomorphisms    | varying-context logic and classifying semantics        |
| Site and sheaf presentation | a category equipped with a Grothendieck topology                         | satisfy descent for the declared covering families | local compatibility, gluing, and contextual variation  |

Table 5.1: Several doctrines for presenting theories and their models. They answer different semantic questions and should not be treated as a single inclusion chain.

### 5.5 From presentations to Grothendieck toposes

When a theory must describe compatible local models across contexts, its semantic universe may be a Grothendieck topos. For a small site  $(\mathcal{C}, J)$ , the category

$$\mathcal{E} = \text{Sh}(\mathcal{C}, J)$$

contains presheaves satisfying the gluing conditions declared by the topology  $J$ . It sits inside the presheaf category

$$i_* : \mathcal{E} \hookrightarrow \widehat{\mathcal{C}} = \text{Set}^{\mathcal{C}^{\text{op}}}$$

as a full reflective subcategory. The inclusion has a left adjoint

$$a : \widehat{\mathcal{C}} \longrightarrow \mathcal{E}, \quad a \dashv i_*,$$

and sheafification  $a$  preserves finite limits. Conversely, a left-exact reflective localization of a presheaf category is a Grothendieck topos [Mac Lane and Moerdijk, 1992, Garner and Lack, 2012].

Thus the relevant functor is not an arbitrary  $T : \mathcal{E} \rightarrow \widehat{\mathcal{C}}$ . The geometric embedding consists of the adjoint pair  $a \dashv i_*$ : its inverse-image functor is the left-exact sheafification  $a$ , and its direct-image functor is the fully faithful inclusion  $i_*$ .

Topos semantics enters synthetic creativity when the proposed theory is inherently contextual. Local causal models, source-conditioned arguments, scientific regimes, and simulator interfaces may agree on overlaps without being globally identical. A topology specifies which families are allowed to count as covers; sheaf semantics determines when their local models glue.

A classifying topos goes further: for a suitable geometric theory, models in another Grothendieck topos can be represented by geometric morphisms into its classifying topos. We use that prospect as a semantic destination, not as an automatic consequence of writing down a sketch. Constructing the site, proving the required universal property, and establishing the claimed model correspondence are separate admission obligations.

#### Boundary

A Lawvere theory, a finite-limit theory, a PROP, and a Grothendieck topos are not interchangeable names for categorical structure. The first three specify different doctrines of operations and model preservation. A Grothendieck topos is a semantic universe of sheaves, equivalently a left-exact localization of a presheaf category.

### 5.6 The creative target package

Chapter 2's status-bearing extension dossier can therefore be refined. The following tuple is the *structural extension subrecord* of that dossier, not a replacement for the status-bearing dossier  $\Xi$ :

$$\mathfrak{T}_{\text{str}}^+ = (\mathbb{S}^+, \mathfrak{D}^+, \text{Mod}_{\mathfrak{D}^+}(\mathbb{S}^+, -), \iota, \delta_{\mathfrak{D}}, \iota^*),$$

where:

*Presentation*  $\mathbb{S}^+$ . The new sorts, generators, equations, cones, cocones, symmetries, or covers.

*Doctrine*  $\mathfrak{D}^+$ . The proposed preservation contract: finite products, finite limits, symmetric tensor, geometric structure, or another explicitly defined doctrine.

*Model semantics*. For every registered target category, the category of structure-preserving model functors.

*Theory map*  $\iota : \mathbb{S} \rightarrow \mathbb{S}^+$ . The relation between the old and proposed presentations.

*Doctrine comparison*  $\delta_{\mathfrak{D}}$ . A declaration of whether the old doctrine is retained, strengthened, weakened, or replaced, together with the compatibility needed to compare old and new models.

*Restriction*  $\iota^*$ . When the presentation and doctrine comparisons make it well-defined, the functor sending an expanded model back to its old-theory reduct.

Here the dash in  $\text{Mod}_{\mathfrak{D}^+}(\mathbb{S}^+, -)$  denotes the registered family of target categories for which the doctrine and its change-of-target semantics have actually been defined; it does not assert functoriality in every category. To recover the complete Chapter 2 dossier, take the presentation component of  $\mathbb{Y}$  to be  $J = \iota$ , let  $\mathbb{S}^+$  denote the presented theory specified by  $\mathbb{S}^+$ ,  $\mathfrak{D}^+$ , and its registered model semantics, and let  $D^+$  denote the current realization of that theory. Record any chosen model transport in  $\tau$ , and retain  $\omega$ ,  $\gamma$ ,  $\Pi$ , and  $\sigma$  with their earlier meanings. Models, countermodels, proofs, simulations, and other admission evidence belong to that provenance-bearing dossier. They are not proposer-controlled fields of  $\mathfrak{T}_{\text{str}}^+$ .

When the doctrine comparison permits it, the restriction functor

$$\iota^* : \text{Mod}_{\mathfrak{D}^+}(\mathbb{S}^+, \mathcal{C}) \longrightarrow \text{Mod}_{\mathfrak{D}}(\mathbb{S}, \mathcal{C})$$

is especially informative. Its fiber over an old model records the model's possible expansions. An essentially unique expansion is evidence



compatible with a definitional extension, but does not prove definitionality by itself; that stronger claim requires the registered theory-equivalence criterion. Several inequivalent expansions expose genuinely new choices. An empty fiber means that the new theory excludes the old model and owes an explanation.

*A concrete restriction-fiber example.* Return to the monoid theory above and extend its presentation by an inverse operation  $i : X \rightarrow X$ , together with the two equations saying that multiplying an element by its inverse on either side yields the unit. The expanded presentation is the theory of groups, and restriction forgets the inverse:

$$\iota^* : \text{Mod}(\mathbb{S}_{\text{Grp}}, \text{Set}) \longrightarrow \text{Mod}(\mathbb{S}_{\text{Mon}}, \text{Set}).$$

For a given monoid  $M$ , the fiber of  $\iota^*$  is empty when some element of  $M$  is not invertible. When every element is invertible, its inverse is unique, so the compatible group expansion is essentially unique. The example makes two points visible at once. Adding a generator need not create an arbitrary new degree of freedom, because equations can determine it uniquely; yet the extension is not definitional over all monoids, because it excludes models that do not satisfy the new existence requirement. A creative system must therefore report both the new syntax and what happens to old models.

Presentation changes also require quotienting. Two sketches can present equivalent model theories. Renaming a generator, adding a definable operation, or changing to a more convenient presentation should not automatically count as transformational creativity. The appropriate comparison may be an equivalence of model categories, a Morita-style equivalence, or a stronger doctrine-specific invariant. Which comparison is required must be registered before evaluation.

### 5.7 From a double repair theory to a finite theory change

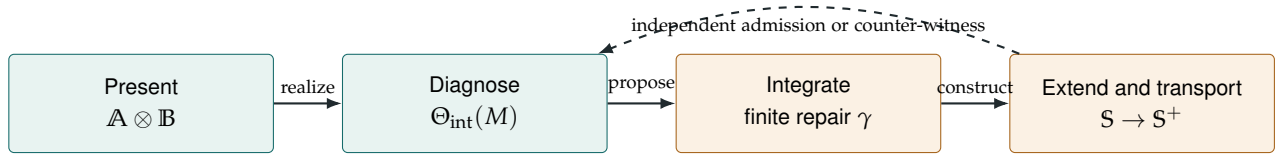
The tensor-product construction of Chapter 3 sharpens the route from local repair to theory extension. Four operations occur on this route, and none should be used as a synonym for another.

1. **Tensor the repair theories.** Enriched sketches  $\mathbb{A}$  and  $\mathbb{B}$  present two typed classes of local variation. Their tensor product  $\mathbb{A} \otimes \mathbb{B}$  presents the side laws together with the squares that compare their interaction.
2. **Diagnose a model of the tensor theory.** A current learning state realizes those declarations only approximately or partially. A typed interchange defect  $\Theta_{\text{int}}$  localizes where the two realized orders fail

to agree. This is a failure in a model relative to a declaration, not yet a change of declaration.

3. **Integrate a proposed infinitesimal repair.** A local direction must extend to a finite path, flow, or groupoid-level action on which the claimed repair remains coherent. A vanishing pointwise residual does not establish this. Monodromy, holonomy, blow-up, path dependence, or loss of compatibility may obstruct finite realization.
4. **Extend the domain theory and transport its models.** Only after a finite proposal  $\iota : S \rightarrow S^+$  is specified can restriction and Kan extension describe how old model semantics are forgotten, freely propagated, or compatibly completed.

The first three concern the *repair calculus*; the fourth concerns the *domain theory being repaired*. In an application they may share objects or semantic categories, but their roles remain different. In particular,  $\mathbb{A} \otimes \mathbb{B}$  does not by itself equal  $S^+$ , and integrating a repair vector field does not automatically choose the generators or equations of a new scientific theory.



This decomposition supplies a useful experimental contract. A computational system need not solve classical smooth integrability exactly in every domain, but it must declare an empirical surrogate: multi-step rollout, path independence, recurrence, continuation across covers, intervention closure, or another finite persistence test. The surrogate should be separated from both the local obstruction score and the final task metric. Otherwise a one-step improvement can be mislabeled as an integrated structural repair.

Figure 5.2: The four semantic levels preceding an admitted DIAL theory change. The dashed return emphasizes that transported models and new evidence can refute the diagnosis, the integration claim, or the proposed extension.

#### Admission contract

For any DIAL-X claim, report separately: the tensor-sketch declaration, the localized interchange defect, the finite-realization test and its registered observer, the proposed domain-theory map, and the independent admission result. If the application has no defensible finite-realization observer, describe the output as an infinitesimal repair candidate rather than a realized creative transition.

### 5.8 Kan extensions as the transport semantics of DIAL

The restriction functor  $\iota^*$  explains how to forget the additional structure of a proposed theory. The converse problem is just as important: given an old model

$$M : \mathcal{S} \longrightarrow \mathcal{C},$$

how should it be transported into the proposed theory  $\mathcal{S}^+$ ? In the ambient functor categories, the canonical candidates are the left and right Kan extensions along  $\iota$ :

$$\text{Lan}_\iota M, \quad \text{Ran}_\iota M.$$

When these constructions restrict to the doctrine-preserving model categories, they form an adjoint triple

$$\iota_! \dashv \iota^* \dashv \iota_*, \quad \iota_! M = \text{Lan}_\iota M, \quad \iota_* M = \text{Ran}_\iota M$$

[Mac Lane, 1998, Riehl, 2016].

Pointwise, when the required colimits and limits exist, their values at an object  $s^+ \in \mathcal{S}^+$  are

$$(\text{Lan}_\iota M)(s^+) \cong \text{colim}_{(\iota \downarrow s^+)} M(s), \quad (\text{Ran}_\iota M)(s^+) \cong \lim_{(s^+ \downarrow \iota)} M(s).$$

The left extension therefore propagates old semantics through the generators and relations directed into  $s^+$ ; the right extension gathers compatible semantics constrained by arrows directed back toward the old theory. It is often useful to read the former as a free or generative transport and the latter as a compatibility envelope. Those phrases describe universal properties, not probabilities, causal interventions, or empirical validity.

The unit and counit

$$M \longrightarrow \iota^* \iota_! M, \quad \iota^* \iota_* M \longrightarrow M$$

make continuity with the old theory explicit. If  $\iota$  is fully faithful, the corresponding ambient Kan extension recovers  $M$  on the old objects. For a genuine sketch doctrine, however, an ambient pointwise Kan extension need not preserve the products, limits, tensors, covers, or tangent structure required of a model. One must prove that  $\iota_!$  or  $\iota_*$  exists inside the relevant model categories, or apply an explicitly justified reflection back into them. In enriched or internal settings the needed construction is an enriched or internal Kan extension [Kelly, 1982]. Existence is an admission obligation, not notation.

#### 5.8.1 Transport does not choose the creative extension

Kan extension begins only after the theory map  $\iota : \mathcal{S} \rightarrow \mathcal{S}^+$  has been supplied. It can determine a universal way to propagate semantics

along that map, but it does not decide that a new sort, mechanism, invariant, or law should exist, nor does it choose among rival extensions. Searching over a registered category of candidate maps may be useful, but remains creative only relative to that previously declared metalanguage.

This locates Kan transport precisely in the DIAL workflow. Double involution diagnoses the interaction between assimilatory and proto-accommodative local variation. A proposal engine turns the localized frontier into a finite sketch map  $\iota$ . Kan extensions then transport old semantics through the proposal. Admission finally determines whether the transported model is coherent, empirically adequate, and useful. The universal construction neither replaces proposal nor certifies admission.

There is also a prior sketch-theoretic layer. If the two local repair directions are presented by enriched limit sketches  $\mathbb{A}$  and  $\mathbb{B}$ , their interaction can be presented by  $\mathbb{A} \otimes \mathbb{B}$ . The equivalence between iterated models and models of the tensor product makes horizontal-first and vertical-first interpretations comparable before any domain theory is extended [MacAdam, 2023]. Thus DIAL's local declaration may itself be an algebraic theory, while  $\iota : \mathcal{S} \rightarrow \mathcal{S}^+$  changes the theory being investigated. These are distinct levels: tensoring presents how two repair calculi interact; Kan extension transports models after a particular theory change has been proposed.

#### Design principle

DIAL diagnoses and helps organize a change of theory; Kan extensions transport semantics across a supplied change of theory. Do not identify the universal transport with the abductive act that proposed its indexing map.

#### 5.8.2 A Beck–Chevalley interpretation of double interaction

There may nevertheless be a deeper relation between DIAL and Kan extension. Suppose a finite extension proposal and a change of context, evidence, or realization form a commutative square of presentations

$$\begin{array}{ccc} \mathcal{S}' & \xrightarrow{J'} & \mathcal{T}' \\ u \downarrow & & \downarrow v \\ \mathcal{S} & \xrightarrow{J} & \mathcal{T}. \end{array}$$

Here the horizontal maps represent a proposed extension and its transported counterpart; the vertical maps represent a second registered change. Whenever the relevant adjoints exist, the commutative square

generates a Beck–Chevalley mate, schematically

$$\beta : J'_! U^* \Longrightarrow V^* J_!.$$

It compares changing context before extending with extending before changing context. If  $\beta$  is invertible, the two composites are naturally isomorphic. If it is not, failure of this particular exactness comparison records a global order dependence relative to the chosen square and adjunctions. Which direction of the mate is relevant depends on the chosen variance and adjunctions; a diagram alone does not supply a preferred learning semantics.

This suggests a research hypothesis rather than an established result. For a DIAL object that integrates to such a square, the internal interchange defect  $\Theta_{\text{int}}$  may be an infinitesimal shadow of the failure of an appropriate Beck–Chevalley mate to be invertible. Under a classical cochain realization, its observer  $\Omega_{\text{cre}}$  would then measure the first-order incompatibility of the two finite transport orders. Symbolically, one would seek a realization in which

$$\Theta_{\text{int}} \rightsquigarrow D\beta \rightsquigarrow \Omega_{\text{cre}},$$

with the arrows defined by an actual integration and realization theorem. No such theorem is assumed here. In particular, the two DIAL side directions are semantic data and need not integrate to theory maps at all.

#### Boundary: DIAL-Kan hypothesis

The proposed relationship is not that DIAL *is* a Kan extension. Kan extension supplies global universal transport once a theory or context map is known. DIAL supplies local two-direction geometry and an interchange diagnostic. A future integration theorem would relate the latter to the infinitesimal component of a Beck–Chevalley comparison.

#### 5.8.3 Left–right asymmetry and admission

In some applications a right Kan extension may encode observational or constraint-based completion, while a left Kan extension propagates semantics after a declared mechanism or theory intervention. This is a useful analogy, not an automatic causal interpretation. A left Kan extension becomes interventional only when the map  $J$  carries an independently specified mechanism-change semantics. Likewise, a right Kan extension becomes observational only relative to a declared evidence interface.

The two universal transports can help expose underdetermination. Agreement between suitable left and right extensions is evidence of

a tightly constrained expansion; divergence can reveal choices not settled by the old theory. Neither outcome decides scientific acceptance. Admission must still test doctrine preservation, transport of trusted consequences, resistance to countermodels, novel predictions, and performance on evidence not used to construct  $J$ .

### 5.9 *Developmental towers of theory*

Piaget's stages of cognitive development suggest a provocative question: could development itself be represented as a succession of increasingly articulated theories? There is no established theorem identifying sensorimotor, preoperational, concrete-operational, and formal-operational reasoning with standard categorical doctrines. Nor do the doctrines in Table 5.4 form a single hierarchy: a PROP and a Lawvere theory make different structural commitments, and a topos is not simply a more mature finite-limit theory. Any correspondence must therefore remain explicitly heuristic [Piaget, 1952, 1985].

The formal core of the analogy is instead a *developmental tower of presentations*

$$S_0 \xrightarrow{i_0} S_1 \xrightarrow{i_1} S_2 \xrightarrow{i_2} \dots$$

At stage  $n$ , the presentation  $S_n$  determines what can be represented, composed, observed, and asked. Assimilation changes or explores a model

$$M \in \text{Mod}(S_n, \mathcal{C})$$

while the presentation remains fixed. A finite extension proposal introduces  $S_{n+1}$ , adding or reorganizing sorts, operations, equations, observers, admissible domains, or the preservation doctrine itself. It becomes an accommodation of the maintained theory only after transport and admission.

Each stage map has the restriction semantics developed in Section 5.6. Its fibers distinguish an incompatible change from an essentially definitional elaboration and from a genuinely underdetermined expansion. The developmental tower therefore adds no new kind of model comparison; it applies the same comparison repeatedly while making the history of admitted presentations explicit.

**Boundary: A lossy functor**

It is useful to call the proposed correspondence from Piagetian development to categorical theory change a *lossy functor*. The phrase is intentionally metaphorical: no source category of cognitive stages and no structure-preserving functor have yet been defined. The mapping forgets psychological, embodied, temporal, and social dimensions while retaining one abstract pattern—alternation between change inside a representational scheme and change of the scheme itself.

| Piagetian capacity            | Tentative categorical reading                                                                                                     | Principal loss in the analogy                                                                  |
|-------------------------------|-----------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------|
| Sensorimotor organization     | Models with actions and state transitions before an explicit symbolic presentation has been extracted.                            | Embodied perception and action are not themselves algebraic theories.                          |
| Preoperational representation | A signature or free sketch of objects and operations with few registered equations or invariants.                                 | Symbol use, language, and perspective taking exceed a bare signature.                          |
| Concrete operations           | Equational or finite-product structure supporting composition, reversibility, and invariants such as conservation.                | Concrete-operational competence is task- and context-sensitive, not merely equational closure. |
| Formal operations             | Hypothetical reasoning across models and controlled extensions of a theory, including comparison of assumptions and consequences. | Meta-level theory variation does not capture the full psychology of abstract reasoning.        |

Two safeguards are essential. First, developmental stages should be attached to semantic capacity, not textual presentation. Presentations related by an appropriate Morita-style equivalence should represent the same stage for the purpose of this abstraction. Second, the developmental relation is at most a partial order. Two theories may support different, incomparable capacities; one need not be globally more advanced than the other.

The proposal therefore does not classify human development. It extracts a formal pattern useful for synthetic creativity: equilibration alternates between fitting experience within a current model category and extending the theory that determines that category. The restriction functor and its fibers make the continuity, incompatibility, and underdetermination of each proposed accommodation auditable.

Table 5.2: A deliberately lossy correspondence between Piagetian capacities and categorical theory construction. The rows are explanatory analogies, not identifications of psychological stages with categorical doctrines.

### 5.10 Boden's modes revisited

The theory/model distinction sharpens Boden's three modes.

*Combinational creativity.* Construct new terms, diagrams, or composites

from registered generators while remaining in the same presented theory. In the double setting, the factors  $\mathbb{A}, \mathbb{B}$  and their interchange law remain fixed; novelty lies in a composite model or artifact. For PROPs, this includes serial, parallel, and symmetric composition.

*Exploratory creativity.* Search the model category of a fixed theory: find models, countermodels, invariants, conjectures, parameter regimes, or unexpected consequences while preserving the doctrine. A claimed trajectory must also pass the declared integrability or finite-persistence observer.

*Transformational creativity.* Change the presentation or doctrine: introduce a sort, operation, equation, limit, tensor generator, topology, or class of admissible models. In DIAL this may change  $\mathbb{A}, \mathbb{B}$ , their tensor-product interchange, the domain sketch  $S$ , or several of these, together with explicit transport from the old theory.

Not every sketch edit is transformational. A definitional extension may make implicit structure explicit without changing the theory's model content. Conversely, a small edit—one new equation or one forbidden diagonal—can radically change the model category. Creativity is assessed semantically, not by textual edit distance.

Chapter 4's Markdown programs have a clear target: their skill sections propose and test controlled edits of  $S$ , its doctrine, and its models. Bracket nonclosure can identify where the current skill or theory declaration is inadequate. It cannot certify that the proposed completion is coherent or scientifically valuable; that is the role of functorial semantics and independent admission.

Chapter 6 next places this target alongside the major traditions in computational creativity. That comparison is essential: novelty, artifact generation, heuristic discovery, symbolic regression, and theory construction are related achievements, but they do not create the same object and cannot be admitted by the same certificate.

#### Admission contract

A synthetic theory is not admitted because its presentation is novel. It must have a declared preservation doctrine, nonempty and interpretable model semantics, explicit transport of prior results, resistance to countermodels, and consequences not used to propose it. Equivalent presentations are counted as one theoretical contribution unless the new presentation itself yields a separately demonstrated computational or explanatory advantage.



### *Further Reading*

Lawvere introduced functorial semantics for algebraic theories [Lawvere, 1963]. Ehresmann introduced sketches as presentations of categorical structure, and Barr and Wells give a broad account of sketches and their models [Ehresmann, 1968, Barr and Wells, 1999]. Lack explains PROPs as one-sorted symmetric monoidal theories and their composition [Lack, 2004]. Mac Lane and Moerdijk develop sheaf and topos semantics, while Garner and Lack state the left-exact localization characterization used here [Mac Lane and Moerdijk, 1992, Garner and Lack, 2012]. Piaget’s accounts of cognitive construction and equilibration motivate the explicitly speculative developmental analogy [Piaget, 1952, 1985]. That analogy should be read through the theory maps, reduct functors, and model expansions defined above, not as a claim that Piaget’s stages are categorical doctrines. Mac Lane and Riehl develop ordinary Kan extensions, and Kelly supplies the enriched setting needed when DIAL models take values beyond ordinary categories [Mac Lane, 1998, Riehl, 2016, Kelly, 1982].



## *Landscapes of Computational Creativity*

### NOVELTY, SEARCH, IMAGINATION, AND SCIENTIFIC DISCOVERY

Chapter 5 fixed the book's principal creative artifact: a presented theory, its models, and an explicit relation to what was known before. Before Chapters 7–9 develop Boden's three modes in detail, it is important to locate that target among neighboring accounts of creativity. Otherwise a recovered equation, a surprising image, a novel search path, and a theory extension can appear to be interchangeable successes even though they produce different objects and require different evidence.

Creativity has been studied as a psychological capacity, a philosophical problem, a property of artifacts, a search process, an information-theoretic event, and a program of automated scientific discovery. No single literature owns the question. The Association for Computational Creativity and its International Conference on Computational Creativity have made this plurality into an interdisciplinary research community spanning artificial intelligence, cognitive science, design, engineering, and the arts [[Association for Computational Creativity, 2026a,c](#)]. The Association also maintains an extensive online bibliography of books, theses, proceedings, conference papers, and other publications. Because that resource can grow with the field, it is the preferred current pointer beyond the selective map developed here [[Association for Computational Creativity, 2026b](#)].

This chapter organizes that broad landscape by asking a concrete question: *what is the object of the creative claim?* It may be an artifact, an observation judged surprising, a path through a conceptual space, a recovered equation, an experimental program, or an extended theory. These are not rival definitions so much as different units of analysis. Once the creative object is fixed, one can ask how it is generated and what evidence warrants calling it creative.

The argument proceeds in four stages. We first compare conceptual-space, developmental, and neurocognitive accounts of what changes. We then separate novelty and surprise from explanation and theory

construction. Next we examine search, imagination, cellular automata, and evolution as mechanisms that generate candidates at different levels of description. Finally, symbolic regression, automated laboratories, and contemporary scientific agents show how those distinctions become experimental. The point is not to force these traditions into DIAL, but to identify what each contributes and what remains missing when the desired artifact is a persistent, revisable theory.

#### Design principle

A creativity claim needs both reference and evidence coordinates. Its reference coordinates identify the object or process being credited, the space relative to which the change is novel, whether the process remained assimilatory or proposed a package change, and which package component changed. Its evidence coordinates record the less-disruptive controls, the proposal's epistemic status, and the certificate warranting admission. "Creative output" without these coordinates is too underspecified for cumulative experimental science.

The chapter is a map, not an exhaustive history. Its purpose is to locate the book's theory-construction objective among complementary research programs and to identify reusable evaluation methods.

### 6.1 *A map by creative object*

Table 6.1 gives a first orientation. The rows overlap: AI Feynman, for example, is simultaneously a search algorithm, a symbolic regression system, and a scientific-discovery benchmark. The classification marks its primary output and certificate rather than forcing each system into one exclusive school.

The final row does not subsume the others. Information-theoretic scores can guide search; computational-creativity criteria can assess process and presentation; symbolic regression can propose laws; and automated laboratories can test them. Synthetic creativity asks what additional structure is needed when the intended output is a reusable theory rather than a single artifact or formula.

### 6.2 *Conceptual spaces and transformations*

Boden's distinction among combinational, exploratory, and transformational creativity is the closest direct precursor to the book's organization [Boden, 1998, 2004]. It shifts attention from the surface novelty of an answer to the conceptual space that makes the answer possible. Wiggins develops a formal framework for describing and comparing creative systems in this spirit, making the rules, traversal mechanism, and evaluation functions explicit [Wiggins, 2006].

| Tradition                  | Creative object                                                            | Generative mechanism                                                   | Typical certificate                                                  |
|----------------------------|----------------------------------------------------------------------------|------------------------------------------------------------------------|----------------------------------------------------------------------|
| Information and surprise   | an unexpected description or repertoire                                    | probabilistic or order-theoretic comparison                            | information gain, improbability, or surprise                         |
| Heuristic discovery        | a concept, conjecture, law, or search path                                 | heuristics over a represented problem space                            | usefulness, rediscovery, proof, or prediction                        |
| Computational creativity   | an artifact and sometimes the producing process                            | generation, transformation, appraisal, and interaction                 | novelty, value, intentionality, and observer judgment                |
| Neuroscience of creativity | a creative cognitive process realized in a brain                           | interacting neural systems supporting generation and evaluation        | behavioral tasks, neuroimaging, stimulation, and convergent evidence |
| Symbolic regression        | an equation explaining numerical data                                      | numerical fitting plus symbolic decomposition and search               | exact or predictive recovery with controlled complexity              |
| Automated science          | a hypothesis–experiment–revision cycle                                     | active experimentation and machine interpretation                      | reproducible intervention and held-out prediction                    |
| Synthetic creativity       | a typed construction or admitted extension relative to a maintained theory | typed composition, local probing, finite theory surgery, and transport | proof, countermodel, simulation, or independent evidence             |

Table 6.1: Traditions organized by their principal creative object. The final column is decisive: different objects require different admission certificates.

This tradition contributes two lasting lessons. First, novelty is relative to a repertoire and a set of generative rules. Second, changing the rules is a different operation from searching them more effectively. Those lessons align with the distinction between models of a fixed sketch and a map  $S \rightarrow S^+$  that changes the presentation itself.

The categorical formulation adds a semantic test to the geometric metaphor. A conceptual space is not only a region of possible artifacts. It can be a presented theory whose admissible realizations are models and whose legal transformations preserve a declared doctrine. A transformation can then be examined through restriction, extension, equivalence of presentations, and the transport of old results. This is particularly useful when two syntactically different spaces generate equivalent model categories: textual novelty alone should not be mistaken for theoretical transformation.

### 6.3 Creativity as a neurocognitive process

A different literature asks how creative cognition is realized in the human brain. Abraham’s *The Neuroscience of Creativity* builds a systematic framework that moves from definitions and assessment to cognitive explanations, global and local brain-based accounts, neuroscientific methods, methodological difficulties, and studies of musical, literary, visual, kinesthetic, and scientific creativity [Abraham, 2018].

That neural scale is only one level of explanation. Sawyer synthesizes psychological, cognitive, developmental, anthropological, and sociological research, emphasizing that creative achievement depends on

collaboration, institutions, and cultural histories as well as individual cognition [Sawyer, 2012]. For this book, the lesson is methodological: an abstract computational architecture and a neural account may describe a capacity without explaining how a community recognizes, teaches, and retains its products.

The developmental lineage begins earlier. Piaget’s assimilation, accommodation, and equilibration motivate the distinction between learning inside a schema and reorganizing the schema itself [Piaget, 1952, 1985]. Chapter 2 introduced the analogy, and Chapter 5 made its categorical component explicit as a developmental tower of presentations. Here it serves only as a coordinate for comparing research traditions, not as a neural implementation or a classification of human development.

Drescher’s Schema Mechanism is the direct computational bridge. It translated Piagetian sensorimotor development into a Lisp-based constructivist learning architecture capable of refining predictive schemas and introducing new representational items [Drescher, 1986, 1991]. It is therefore an important predecessor to systems that treat representation construction itself—rather than parameter adjustment alone—as a learning objective.

This biological level is complementary to the computational landscape. A categorical theory can specify the types of representations, variations, transformations, and admission judgments required by a creative process without identifying the neural mechanisms that implement them. Conversely, evidence about brain networks or regions does not by itself determine the abstract organization of the conceptual space being explored. The two descriptions answer different questions about the same capacity.

The neuroscience perspective nevertheless constrains synthetic models. It foregrounds the separation and interaction of generation, control, assessment, memory, and imagination; it also forces careful attention to what a task or instrument actually observes. These lessons support modular creative architectures and explicit observer models. They do not license a claim that Weil probes, tangent directions, or sketch transformations are literal neural objects. Establishing such a correspondence would require an independent model relating categorical roles to neurocognitive processes.

#### 6.4 *Assimilation and accommodation across the landscape*

Chapters 2 and 3 introduced the Piagetian distinction and its differential realization. This survey does not define that machinery again. It uses only the resulting diagnostic question: does a process work inside a maintained representation, or does it propose a change to the representation, its observers, its generative operations, or its admission

rules? The first is assimilatory. The second is proto-accommodative until independent evidence admits it. A revised admission rule cannot, by itself, certify the revision that introduced it.

This axis resembles Boden’s distinction between exploratory and transformational creativity, but the two classifications are not identical. Boden classifies how a conceptual space is used or transformed. The Piagetian reading additionally emphasizes why reorganization occurs, how the old scheme is related to the new one, and whether the change restores a viable equilibrium with experience. Combinational and exploratory work is often assimilatory, while transformational work often begins with a candidate extension; neither implication is automatic.

| Tradition                        | Assimilatory reading                                                                         | Accommodative reading                                                                                      |
|----------------------------------|----------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------|
| Information and surprise         | Update probabilities or locate novelty inside a fixed repertoire.                            | Change the variables, contexts, repertoire, or sample space that makes surprise measurable.                |
| Heuristic discovery              | Search with registered primitives and heuristics, as in many BACON-style law-recovery tasks. | Introduce a missing variable, concept, operator, or discovery heuristic and transport prior results.       |
| Computational creativity         | Generate and appraise artifacts under a fixed grammar and value model.                       | Revise the grammar, inspiring set, evaluator, or ontology of admissible artifacts.                         |
| Imagination and world models     | Simulate counterfactual variation inside a maintained world model.                           | Repair the state ontology or transition laws when imagined and observed consequences cannot be reconciled. |
| Cellular automata and evolution  | Explore trajectories, initial conditions, or inherited variations under a fixed constructor. | Introduce a macroscopic ontology or alter the effective constructor–variation system.                      |
| Symbolic regression              | Recover expressions in a registered language.                                                | Extend the language with latent quantities, coordinates, operations, or factorization principles.          |
| Automated and LLM-guided science | Optimize hypotheses and experiments inside a supplied scientific world.                      | Revise the objects, mechanisms, admissible interventions, or theory language that defines that world.      |

Three cautions prevent the distinction from becoming rhetorical. First, assimilation is not synonymous with a small parameter update. A long proof, large search, or extensive experimental campaign can remain inside the same representation. Second, accommodation is not synonymous with textual or architectural novelty. A renamed primitive, redundant feature, or Morita-equivalent presentation may leave the model content unchanged. Third, an accommodative *proposal* is not yet cognitive or scientific progress. Only independent admission can establish that the revised package explains the obstruction while retaining the warranted achievements of its predecessor.

Thus the table is a crosswalk, not a second formalization of DIAL.

Table 6.2: A Piagetian reading of computational-creativity traditions. Assimilation and accommodation are relative to the maintained representational package, not intrinsic labels attached permanently to an algorithm or artifact.

Later chapters evaluate the two loops directly: exploration and repair within a declared representation, followed—when necessary—by minimal typed extension, transport of prior competence, and independent admission.

### 6.5 *Prediction, explanation, and structural possibility*

Chomsky, Roberts, and Watumull’s 2023 essay, “The False Promise of ChatGPT,” offers a deliberately sharp critique of the conception of language and knowledge embodied by statistical machine learning [Chomsky et al., 2023]. Its central concern is not that language models lack practical uses. It is that increasingly successful prediction of probable continuations may be mistaken for the human capacities to construct explanations, distinguish possible systems from impossible ones, and reason within causal and normative constraints. On their account, a compact theory that explains why an event can occur is categorically different from a model that becomes proficient at reproducing its statistical appearance.

This criticism converges with McCarthy’s distinction between appearance and reality and with Zahavy’s argument that scientific creativity may require an abductive jump beyond a fixed predictive language [McCarthy, 2007, Zahavy, 2026]. All three positions resist the identification of intelligence with compression or curve fitting alone. They ask how a learner obtains the theoretical entities, mechanisms, and laws that make observations intelligible and support questions not already implicit in the original representation.

The distinction between improbable and impossible is especially important for DIAL. A probabilistic model may assign a construction a small value  $p(x \mid D)$ , but low probability does not establish that the construction is ill-typed, inconsistent, or unrealizable. Relative to a sketch  $S$  and a registered semantic class, structural impossibility means that no model in that class satisfies the declared equations, cones, and preservation doctrine. Failure of a search procedure to find such a model is not by itself a proof of nonexistence. The resulting implications should not be collapsed:

unlikely observation  $\nRightarrow$  poor fit of the current realization,  
 poor fit of the current realization  $\nRightarrow$  violation of the current theory,  
 violation of the current theory  $\nRightarrow$  inadequacy of the theory itself.

The first two conditions may be assimilated by changing beliefs or model parameters. A violation may expose bad data, a failed observer, or an inadmissible proposal. Persistent, localized failure supported by independent evidence may warrant opening a candidate extension; the



declaration changes only after the proposed extension and its transport pass admission.

### 6.5.1 Chomskyan structure and Piagetian development

The Chomskyan and Piagetian perspectives create a productive tension. Chomsky emphasizes antecedent structural constraints that make language learning possible at all. Piaget emphasizes how cognitive schemas are reorganized through assimilation, disequilibrium, and accommodation. DIAL does not require a reasoner to begin as an unconstrained statistical blank slate, but neither does it require the initial declaration to remain fixed. A sketch supplies a current grammar of objects, operations, and admissible compositions; assimilation learns inside it; proto-accommodation probes its possible reorganization; a candidate finite extension proposes a new sketch; and admitted accommodation installs it as the maintained declaration.

The resulting synthesis can be summarized as

- begin with structure  $\rightarrow$  learn within it
- $\rightarrow$  diagnose its boundary
- $\rightarrow$  construct and admit revised structure.

This is not a claim that categorical sketches implement universal grammar or that DIAL models child development. It is an architectural response to two requirements: learning needs an admissible space, and creative learning may need to revise the space that made earlier learning possible.

The connection also clarifies the role of explanation. A DIAL system has not explained a phenomenon merely because it emits a fluent rationale or records an obstruction. An explanatory accommodation must introduce explicit objects or mechanisms, derive relevant consequences, transport what remains valid from the previous theory, and survive countermodels or interventions not used to propose it. Explanation is therefore a property of an admitted theory package, not of persuasive prose.

### 6.5.2 An empirical target, not a settled verdict

The 2023 essay should be read as a motivating position rather than a settled capability theorem. Language-model architectures and evaluations continue to change, and experiments can challenge particular claims about their structural biases. Wang and Ma, for example, compare learning of constructed “possible” and “impossible” languages by small transformer and recurrent models and report architecture-dependent behavior [Wang and Ma, 2026]. Their reported comparisons

are descriptive single runs ( $n = 1$  per condition), so they do not support formal statistical inference. Such evidence can weaken an assertion that all statistical learners are indifferent to structural possibility without showing that their learned preferences constitute explicit causal theories or human-like linguistic knowledge.

The corresponding DIAL experiments must therefore go beyond asking whether a model prefers one string distribution to another. They should test whether a system can declare the relevant constraints, produce countermodels for violations, identify when no admissible realization exists, and decide when to repair a model rather than extend its theory. The strongest test is transfer: does the proposed explanation support new deductions, interventions, or experimental designs outside the data that elicited it?

Chomsky and colleagues also connect intelligence to moral reasoning. DIAL does not derive ethical principles from double involution. Its admission contracts can make normative constraints, responsibility for interventions, and rejected alternatives explicit, but the source and legitimacy of those constraints remain separate philosophical and institutional questions.

#### Boundary: What the critique contributes

The Chomsky–Roberts–Watumull critique supplies a demanding target: move from statistical continuation toward explicit possibility constraints, causal explanation, and responsible judgment. DIAL offers a candidate architecture for declaring, diagnosing, and revising structure. It does not yet establish that present language models lack those capacities in principle or that double involution is sufficient to produce them.

## 6.6 *Novelty, information, and surprise*

An information-theoretic tradition asks how unexpected an event or description is relative to a probability model. Palm broadens this perspective from Shannon's partitions to repertoires and arbitrary covers, separating related notions of novelty, information, and surprise [Palm, 2012]. This is valuable for creativity because real descriptions frequently overlap: an artifact may belong simultaneously to several styles, mechanisms, or explanatory classes.

Barto, Mirolli, and Baldassarre sharpen a distinction that is often blurred in computational and behavioral accounts [Barto et al., 2013]. Novelty is primarily relative to memory or a repertoire: has an item, representation, or relation been encountered before? Surprise is relative to a predictor: does an observation violate an expectation? The two

may co-occur, but a familiar event can be surprising in a new context, while a novel event need not violate a registered prediction. Both can function as intrinsic motivations for exploration and learning.

This distinction also clarifies the Piagetian analysis. Surprise is a natural signal of disequilibrium, but not an automatic trigger for accommodation: the learner may assimilate the observation by repairing parameters or expectations within its current representation. Novelty may likewise enlarge a stored repertoire without changing its organizing concepts. A candidate extension is warranted only when persistent novelty or surprise exposes an inadequacy that cannot be repaired within the current representational language. The proposed change must still pass independent admission.

The connection to this book is richer than assigning a scalar novelty score. A cover resembles a family of partial observational contexts. One may ask not only whether a candidate is rare, but in which repertoires it is novel, which contexts distinguish it, and whether local descriptions agree on overlaps. This opens a path from information-theoretic novelty to sheaf-like contextual analysis.

Yet surprise and creativity must remain distinct. An outlier can be highly surprising and scientifically worthless; a powerful abstraction may appear simple once expressed in the right language. Moreover, the probability model used to score surprise is itself part of the inherited conceptual space. A transformational contribution may change the variables, descriptions, or sample space on which that score was defined.

#### Boundary

Information theory can quantify novelty or surprise relative to a registered repertoire. It does not, by itself, construct a new ontology, determine which structural promise failed, or certify that a proposed theory extension is explanatory. Those require typed representation and domain-specific admission.

### 6.7 *Discovery as heuristic search*

A long AI tradition treats discovery as search through a structured space of concepts, hypotheses, or programs. Lenat's AM explored elementary mathematics using a growing collection of concepts and heuristic rules [Lenat, 1976]. The BACON family sought empirical laws by applying heuristics to numerical regularities, and the broader computational account of scientific discovery linked representation, search, and theory formation [Langley et al., 1987]. Genetic programming and later symbolic regression continued the program by searching executable

expression spaces; Schmidt and Lipson showed that compact free-form laws could be recovered from experimental data without supplying the target equations [Schmidt and Lipson, 2009].

### 6.7.1 BACON: discovery from numerical regularity

BACON is the indispensable classical precedent for the experiments developed in this book. Beginning from tables of measurements rather than a supplied target equation, the program searched for simple regularities among variables. Its heuristics considered relations such as constancy, proportionality, and inverse proportionality, introduced derived quantities when useful, and applied the same discovery operations recursively. Across successive versions, BACON rediscovered forms of the ideal-gas law, Kepler's third law, Coulomb's law, Ohm's law, and other historical relations in physics and chemistry [Langley et al., 1987].

This achievement matters for two reasons. First, it demonstrated more than forty years ago that scientific law discovery could be studied experimentally rather than treated as an unanalyzable flash of inspiration. Second, the program made the representation of the search space visible. Its successes depended on a carefully designed vocabulary of variables and discovery heuristics. Later systems in the same research program—including DALTON, GLAUBER, and STAHL—expanded the kinds of quantitative and qualitative regularities that could be represented.

The present experiments revisit BACON's question with very different machinery. Frontier models provide broad proposal capabilities; categorical presentations specify typed objects, operations, and equations; infinitesimal probes localize failures; and independent admission checks candidate repairs. Our generated cyclic and finite-symmetry worlds are deliberately modest in the same methodological sense as BACON's historical rediscoveries: their purpose is to expose the mechanism of discovery under controlled conditions. The objects sought, however, are not only scalar formulas. A successful package may need to identify a law family, extend a declaration, preserve earlier evidence, derive new consequences, or abstain when the registered theory language is inadequate.

The comparison yields a useful evidence hierarchy. A BACON-level task asks whether a system can recover a compact empirical relation. A structural task asks whether it can also recover the relation's sorts, symmetries, factorizations, and domain of validity. A transformational task withholds a needed primitive and asks the system to enlarge the theory itself. Finally, a persistent-discovery task asks whether the admitted extension is transported and reused in later problems.

BACON therefore supplies both a baseline and a historical warning: success inside a discovery language does not by itself show that the system can revise that language.

This literature makes representation impossible to ignore. A search procedure can discover only what its primitives, grammar, and evaluation functions make expressible. Heuristics may dramatically enlarge the tractable region without changing that language. Conversely, adding one well-chosen primitive can turn an intractable search into a short derivation.

Synthetic creativity therefore divides the search problem in two. An *inner loop* explores expressions, models, and skill compositions inside the current sketch. An *outer loop* proposes changes to the sketch, observer family, or preservation doctrine. The inner loop resembles classical heuristic discovery and is assimilatory relative to the registered discovery language. The outer loop proposes candidate extensions, makes the representation change explicit, and demands a transport account. Chapters on exploratory and transformational creativity develop these loops separately.

### 6.8 *Local novelty, integrability, and persistence*

The differential account adds a distinction that is often absent from creativity benchmarks. A proposal may be locally novel and locally beneficial without defining a realizable creative process. An infinitesimal direction can fail to extend to a finite trajectory; an apparently repaired interaction can accumulate path dependence; a new symbolic rule can work on the chart that suggested it and fail under transport to the next context.

We therefore distinguish three evaluative questions:

*Local possibility* Does the proposal reduce a typed obstruction or open a previously unavailable direction?

*Integrability* Can successive local changes be composed into a finite, coherent path under the declared dynamics?

*Persistence* Does the resulting object or theory retain its value on an untouched cover, recurrence, intervention, proof obligation, or later task?

Classical Lie integration gives this distinction its sharpest mathematical form, but computational applications may use registered empirical observers. Examples include multi-step rollout stability, approximate holonomy, cycle-consistency, continuation under context shift, or reuse after a regime returns. These are not interchangeable tests. Each application must explain why its observer is evidence for finite realization.

This criterion also clarifies what the literature’s artifact/process distinction leaves open. A process can generate an attractive artifact while being structurally unstable, and a stable process can produce an artifact that lacks novelty or value. Integrability is consequently neither a complete definition of creativity nor another name for quality. It is the bridge from a local creative opportunity to a sustained construction.

#### Design principle

Do not credit an infinitesimal creative proposal as a realized transition until it survives a finite composition or persistence test. Report local novelty, integrability evidence, and independent value as separate endpoints.

### 6.9 *Artifacts, process, and evaluation*

Computational-creativity research asks more than whether an output differs from its training examples. Ritchie proposes empirical criteria relating a program, its inspiring set, and its generated artifacts [Ritchie, 2007]. Colton’s “creative tripod” asks a system to exhibit skill, appreciation, and imagination, thereby treating generation and self-evaluation as distinct capacities [Colton, 2008]. Jordanous’s SPECS methodology begins by declaring what creativity means for the system under study and then deriving tests from that declaration [Jordanous, 2012].

Miller’s cross-domain survey provides a complementary historical snapshot of AI-powered practice in visual art, music, literature, poetry, and performance before the present generation of frontier models [Miller, 2019]. Its breadth is useful here because it shows how many different artifacts had already been called creative by 2019. It is a map of systems and practices, rather than a single computational theory by which their creative mechanisms can be identified.

These approaches anticipate two principles of this book. Evaluation must be declared rather than improvised after seeing the output, and the process can matter as much as the artifact. A system that accidentally emits a valuable formula is not equivalent to one that represents its assumptions, diagnoses a failure, proposes a change, and records why the change survived testing.

The status-bearing extension dossier is consequently both an artifact and a process trace. It records the prior theory, obstruction, proposal, transport, predictions, and epistemic status. The package does not settle the philosophical question of whether the system is “really” creative. It supports an empirical judgment about a specified capacity while preserving the evidence needed for another evaluator to disagree.

### 6.10 *Imagination and counterfactual worlds*

The Blue Sky paper *Imagination Machines* argued for imagination as an overarching challenge for AI and connected generative models to layered causal capacities for observation, intervention, and counterfactual reasoning [Mahadevan, 2018]. The arrival of powerful generative and world models makes that challenge experimentally approachable: a frontier model can now propose mechanisms, scenes, analogies, experimental designs, and formal objects at a scale unavailable to earlier systems.

Imagination nevertheless exceeds unconstrained sampling. A useful imagined world must preserve some invariants while varying others, expose the intervention that produced the variation, and distinguish a possible world from an admitted claim about the actual one. Simulators are especially important because they turn an imagined intervention into an executable test. They also impose a hard limitation: success establishes validity relative to the simulator, not automatically to nature.

This motivates the division of labor adopted here. Frontier models serve as proposal engines; sketches register the structure being preserved or changed; tangent and Weil probes explore local alternatives; and proof, simulation, or empirical evidence controls admission. Imagination supplies candidates, while the theory package prevents a vivid candidate from silently becoming a fact.

### 6.11 *Cellular automata and emergent conceptual structure*

Cellular automata supply a complementary lineage in which complex behavior is generated by the repeated composition of a simple local law. Von Neumann's theory of self-reproducing automata established an early cellular setting for universal construction and reproduction [von Neumann, 1966]. Wolfram's *A New Kind of Science* made the systematic exploration of simple programs—especially cellular automata—the basis of a broad experimental study of complexity, emergence, universality, and computational irreducibility [Wolfram, 2002].

The important contrast with a Turing machine is representational, not a claim of greater computability. Universal cellular automata and universal Turing machines share the same classical ceiling on what is computable. Cellular automata nevertheless expose a different behavioral repertoire: computation is spatially distributed; the same local rule acts everywhere; global patterns arise through neighborhood interaction; and one can vary alphabets, neighborhoods, initial conditions, and rules while retaining an exact executable semantics. Those features make composition and emergence visible in a way that is particularly useful for synthetic creativity.

There is also a precise categorical reading. Let  $G$  be a group of spatial translations and  $A$  a finite alphabet. The configuration space is the function object

$$A^G = \{x : G \rightarrow A\}.$$

A finite neighborhood  $M \subseteq G$  and local rule  $\mu : A^M \rightarrow A$  induce a global evolution map

$$F_\mu : A^G \longrightarrow A^G.$$

For finite discrete  $A$ , the Curtis–Hedlund–Lyndon theorem characterizes such global maps as precisely the continuous maps that commute with the translation action [Hedlund, 1969]. Thus a cellular automaton is an equivariant endomorphism of a configuration object, and the composite of two cellular-automaton updates is again an equivariant endomorphism. Shift spaces and sliding-block codes can accordingly be organized into a category.

This formulation can be presented as a sketch. The declaration contains the cell-state object, neighborhood projections, local update, translation action, and equations

$$F_\mu \circ \sigma_g = \sigma_g \circ F_\mu \quad (g \in G).$$

A trajectory is then a model of the iterated update law. At the microscopic level the theory is complete: the local rule determines every next state. At the explanatory level it may be radically incomplete. Persistent domains, moving structures, collision types, conserved quantities, and effective particles are not normally named in the initial cell alphabet. Inventing those macroscopic objects and their interaction laws is a concrete instance of theory construction from behavior.

This distinction is central to the present book. Simulating another time step is deduction inside a fixed microscopic theory. Searching a fixed rule space is exploratory creativity. Introducing a new macroscopic vocabulary that supports stable predictions is potentially transformational: the system has constructed a higher-level theory whose models organize many trajectories at once. Computational irreducibility warns that no general method need compress every trajectory into a shortcut. It does not imply that useful invariants, causal mechanisms, or mesoscopic theories cannot be discovered for particular families.

Ordinary finite cellular automata have no canonical differential or tangent structure. Their natural probes are initially discrete: flip a cell, change a local table entry, enlarge a neighborhood, perturb an initial pattern, or compare two update orders. Continuous or probabilistic relaxations may support tangent probes, but any resulting derivative must be transported back to the discrete automaton and checked against its exact semantics. This makes cellular automata a useful test of whether synthetic creativity can retain the logic of typed



probing without pretending that every creative space is intrinsically smooth.

### 6.12 *Evolution as a creative constructor*

The cellular-automaton lineage was partly motivated by biology. Von Neumann wanted to understand how a machine could construct a copy of itself while retaining the capacity for heritable change. Evolution supplies the natural example on an extraordinary scale: finite inherited descriptions participate in the development of organisms containing nested molecular circuits, organelles, cells, tissues, and organs, while the reproductive system transmits descriptions from which altered successors can develop.

It is tempting to say that evolution *designed* this architecture. That language must remain metaphorical. Dawkins’s phrase “the blind watchmaker” captures the essential qualification: natural selection is cumulative but has no foresight, plan, or final target [Dawkins, 1986]. Organisms are not steps along a monotone path toward perfection. They are historically contingent solutions to local problems of survival and reproduction under changing environments, complete with compromises, redundancies, and inherited constraints.

The comparison with von Neumann becomes sharper when the description is used in two modes. A self-reproducing automaton must interpret its description to construct a successor and also copy that description without interpreting it. Biological inheritance exhibits an analogous division: genomic sequences are expressed through cellular machinery and are also replicated into descendant cells and organisms. DNA is therefore not a freestanding blueprint for a body. Development also depends on pre-existing cellular organization, regulatory and epigenetic state, intercellular signaling, spatial context, stochastic events, and the environment. The finite string is central, but the constructor is the larger living system in which that string is interpreted.

Schematically, a genotype  $g_k$  participates in a developmental map

$$g_k \xrightarrow[\text{environment } e_k]{\text{Dev}} O_k \xrightarrow{\text{Rep} + \text{Var}} g_{k+1}.$$

At population scale, selection changes which resulting descriptions and developmental systems contribute to the next generation. The recursively important point is that  $g_{k+1}$  may change not only the organism but also parts of the developmental, reproductive, and variation-generating machinery that constructs later organisms. Evolution can thereby alter its own effective space of future possibilities—what evolutionary biology calls the evolution of evolvability.

For synthetic creativity, the resulting artifact is not simply one successful organism. It is a persistent constructor–description system,

embedded in a population and environment, that can generate and test descendants. Its compositional hierarchy may be written suggestively as

genes  $\longrightarrow$  molecular circuits  $\longrightarrow$  cells  
 $\longrightarrow$  tissues  $\longrightarrow$  organs  $\longrightarrow$  organisms.

The arrows are not simple one-way functions: feedback, context, and many-to-many realization occur at every level. Their value is to mark a tower of effective theories. A cell type, tissue boundary, or organ is a stable macroscopic object whose laws are not transparently stated in the nucleotide alphabet, even though the genome participates in its construction.

Boden's three modes can all be recognized in this process. Recombination is literally combinational; mutation and selection explore an inherited developmental space; and major changes to regulation, reproduction, or levels of biological organization transform the space in which later variation occurs. Evolution thus offers a model of creativity whose product is not only a theory of a world, but an embodied constructor capable of continuing the creative process.

This analogy also anticipates recursive self-improvement. The biological loop does not certify itself: differential survival and reproduction under a changing environment supply external selective pressure, not epistemic admission. The narrower lesson for a synthetic discovery system is that it should revise its theory, probes, and constructor only while retaining evidential channels that it cannot silently redefine.

Contemporary research programs explicitly pursue recursive improvement in automated research, system design, and AI hardware [Recursive Superintelligence, 2026, Ricursive Intelligence, 2026, Sakana AI, 2026, Chen et al., 2026]. They are signals of an active engineering direction, not evidence that an autonomous system can already construct and independently admit a new scientific theory.

Wilson brings an evolutionary and humanistic lens to the origins of language, art, and creative expression [Wilson, 2018]. His account broadens the question from biological construction to human symbolic culture, but it is philosophical and conjectural rather than a computational mechanism or an empirical validation of DIAL. It therefore marks a phenomenon to be explained, not an evidential premise of the framework developed here.

### 6.13 *AI Feynman as a controlled testbed*

Symbolic regression is an unusually clean meeting point between machine learning and scientific discovery. Given samples from an unknown function, the system must recover a symbolic expression rather than merely interpolate the samples. Udrescu and Tegmark observe that

scientifically interesting functions often possess simplifying structure—symmetry, separability, compositionality, dimensional constraints, and simple transformations—even though unrestricted symbolic regression is combinatorially hard [Udrescu and Tegmark, 2020].

AI Feynman combines neural fitting with physics-inspired tests that detect and recursively exploit this structure. On the authors' benchmark of one hundred equations drawn from the *Feynman Lectures on Physics*, it recovered all one hundred, compared with seventy-one for the previously available systems in their comparison. The important fact for this book is not only the score. The algorithm succeeds by diagnosing structural properties and using them to factor a hard search into simpler subproblems.

The name also recalls Feynman's atomic-hypothesis thought experiment [Feynman et al., 1963, Section 1–2]. Its ideal is not the isolated recovery of one equation, but a compact theory whose primitives and laws make many phenomena intelligible. Equation recovery is thus a natural controlled baseline, while recovery of the generative theory remains the more ambitious target.

Categorically, a recovered formula can be read as a term in a many-sorted algebraic theory. Units constrain sorts; composition records the expression tree; symmetry identifies invariant actions; and separability exposes a factorization diagram. AI Feynman therefore supplies a concrete benchmark for the claim that compositional structure can guide discovery.

It is not yet a benchmark for full theory construction. The variables are given, the observations are generated by a known closed-form expression, the language of candidate formulas is largely fixed, and success is defined by recovering an existing law. This is primarily exploratory creativity within a carefully designed expression space. A synthetic-creativity benchmark should preserve that controlled core while progressively relaxing those supports.

#### 6.14 A four-level symbolic-discovery challenge

The AI Feynman suite can anchor a more demanding four-level test without replacing its established task.

1. **Equation recovery.** Reproduce symbolic regression from numerical samples, recording predictive error, symbolic equivalence, expression complexity, and computation. This is an assimilatory task relative to the registered expression language.
2. **Structural recovery.** Return not only a formula but a typed presentation of its variables, units, symmetries, separability relations, and compositional factorization. The certificate compares the recovered

structure with a hidden reference sketch. It remains assimilatory when that structural vocabulary is already supplied by the benchmark's meta-language.

3. **Representation repair.** Begin with a deliberately inadequate grammar or ontology. Permit the system to introduce a latent quantity, new operation, quotient, or coordinate system when the obstruction cannot be resolved inside the original language. This tests a proto-accommodative proposal rather than admitted theory change.
4. **Theory-package construction.** Require the extension map, the transport of earlier fits, new predictions, countermodels, and an admission status. Evaluate the package on held-out regimes or simulator interventions that were unavailable during proposal. This tests whether a proposed accommodation survives equilibration with independent evidence.

The benchmark must also control memorization. Famous Feynman equations are likely to occur in frontier-model training data. We therefore need procedurally generated “Feynman-like” families, blinded renamings and coordinate changes, novel compositions of known mechanisms, and withheld intervention regimes. Historical equations remain useful calibration cases, but claims about creative mechanism should rest on worlds whose generating structure the model could not have retrieved verbatim.

#### Experiment: AI Feynman extension benchmark

**Epistemic status:** proposed benchmark design; this card specifies an evaluation target rather than reporting a completed experiment.

**Input:** numerical observations, declared units, and a versioned expression grammar; later levels withhold or corrupt parts of the grammar.

**Proposal:** equation, structural sketch, and any requested theory extension.

**Admission:** symbolic equivalence or predictive fit, recovery of hidden structure, extrapolation under interventions, and audit of transport.

**Creativity label:** equation recovery is Mathematics–E; a new primitive or doctrine is a candidate Mathematics–T result before admission and a benchmark-relative Mathematics–T result only after independent admission.

**Boundary:** recovery of a known formula is personal or system-relative creativity, not evidence of historical scientific discovery.

### 6.15 Large language models in scientific discovery

Moving from law recovery to theory building requires evidence that can answer back. Automated laboratories provide one route because they select and execute experiments rather than merely fit a static dataset. The Robot Scientist “Adam,” for example, coupled hypothesis formation, experiment selection, laboratory automation, and interpretation in a closed functional-genomics setting [King et al., 2009]. Such systems make admission an active process constrained by instruments, time, cost, and biological variability.

The rapid growth of language-model systems for science supplies a second contemporary comparison. Zheng and colleagues organize this literature by an *autonomy axis*: an LLM may serve as a tool for one stage of inquiry, an analyst that integrates information and proposes conclusions, or a scientist that orchestrates substantial portions of the research cycle [Zheng et al., 2025]. Their accompanying, continuously updated bibliography spans literature synthesis, hypothesis generation, experimental planning, data analysis, validation, function discovery, autonomous research agents, and scientific foundation models [HKUST KnowComp, 2026]. This is a valuable operational map of a fast-moving field.

Autonomy, however, is not the same axis as structural creativity. A system can autonomously traverse a hypothesis space, operate tools, and optimize an experimental objective while leaving the objects, admissible operations, and laws of its scientific language unchanged. Conversely, a human-guided system may propose a consequential change to that language. Synthetic creativity therefore adds a *theory-change axis*: whether a system searches within a fixed model class, recovers latent structure, or constructs and admits an extension of the theory itself.

In Piagetian terms, autonomy measures how much of the research loop the system executes without human direction; it does not measure whether the loop is assimilatory or accommodative. A highly autonomous agent may repeatedly assimilate evidence into a fixed ontology, while a human-guided collaboration may perform the decisive accommodation by introducing a new scientific object or mechanism.

The Scientific Discovery Evaluation (SDE) framework makes this distinction empirically concrete. It evaluates frontier models on 1,125 expert-vetted questions drawn from 43 research scenarios in biology, chemistry, materials, and physics, and separately on eight closed-loop projects in which models propose hypotheses, receive simulator or experimental feedback, and revise their proposals [Song et al., 2025]. The study finds a substantial gap between ordinary science-question performance and discovery-grounded performance, diminishing returns from model scaling and test-time reasoning, correlated failure modes

across leading models, and no single model that dominates all projects. It also reports cases of successful optimization despite weak scenario-level knowledge, suggesting that guided or serendipitous exploration can sometimes compensate for an incomplete explicit account of the governing structure.

The authors also delimit the benchmark: its 43 scenarios and eight projects reflect a finite contributor cohort, omit several scientific and engineering fields, and use one evolutionary search strategy and prompting protocol for the more expensive project-level evaluations. Commercial API variation and the restriction of those evaluations to a subset of models further limit reproducibility and breadth. SDE is therefore a strong, explicitly bounded baseline rather than a universal assay of scientific discovery [Song et al., 2025].

These results make SDE a strong baseline for this book, but its closed-loop projects generally begin with a supplied hypothesis space, computational oracle or simulator, and selection rule. The principal output is a successful candidate within that registered world. Our experiments ask a complementary question: can a system recover the world’s latent factors, generators, relations, invariants, and domain of validity, and can it enlarge that presentation when the registered language is inadequate? High objective value and accurate answers remain evidence, but neither alone certifies the construction of a reusable theory.

#### 6.15.1 *Biological foundation models, world models, and co-scientists*

Current biological systems illustrate three components that a synthetic discovery architecture may combine. Evo 2 is a genomic foundation model trained on trillions of DNA base pairs across the domains of life. It predicts effects of genetic variation, represents biological features learned from sequence alone, and generates genome-scale sequences [Brix et al., 2025]. It therefore supplies a powerful proposal distribution over a biologically meaningful sequence space, but generation within that space is not by itself an explicit theory of the mechanisms that make a sequence viable.

A complementary program seeks multimodal foundation models of human cells, connecting DNA, RNA, proteins, structures, images, literature, cell states, and higher biological scales. Biomni addresses the accompanying workflow problem: it combines language-model reasoning, retrieval, code execution, and a large environment of biomedical tools, databases, and protocols to carry out heterogeneous research tasks without a fixed task template [Huang et al., 2025]. A Stanford overview places Evo 2, virtual-cell modeling, and Biomni side by side as, respectively, a biological generative model, a prospective simulation

substrate, and an agentic co-scientist [Itoi, 2026].

For synthetic creativity these roles should remain distinct. A domain foundation model proposes structured candidates; a virtual cell can act as a world model for simulated interventions; and an agent such as Biomni can orchestrate literature, databases, tools, and protocols. The theory layer must still declare what the system takes the biological objects and mechanisms to be, record which compositional obligation failed, and decide whether evidence from the simulator or laboratory admits a proposed extension. Together these systems make the experimental infrastructure far more powerful; they do not remove the need for an explicit and persistent object of scientific change.

The distinction is echoed by contemporary evaluations of idea generation. The Stanford report describes language-model proposals judged highly novel but often infeasible, with humans regaining the advantage when execution was included in the assessment [Itoi, 2026]. In the terminology of this book, proposal novelty is not admission. A creative scientific system must connect an imaginative variation to executable intervention, evidence, and a theory package that can survive subsequent use.

The long-term target for synthetic creativity combines strengths from across this chapter: Palm’s context-relative account of novelty; Boden and Wiggins’s explicit conceptual spaces; Lenat’s and BACON’s heuristic discovery; the computational-creativity community’s process-sensitive evaluation; imagination machines as proposal generators; AI Feynman’s structural decomposition; and automated science’s experimental loop. Its distinctive wager is that these components can be coordinated around a persistent algebraic theory whose changes are explicit and whose models can be tested.

That claim remains an open empirical hypothesis. Neither a high novelty score, an impressive generated artifact, a recovered equation, nor a successful simulator intervention alone demonstrates field-building creativity. The achievement sought by this book is cumulative: a new presented language that organizes several results, survives independent criticism, and enables other agents to pose questions that were previously unavailable.

#### Admission contract

Each experiment must state whether it evaluates artifact novelty, search, equation recovery, experimental discovery, or theory extension. Scores may be combined only after their semantics are recorded. In particular, predictive accuracy does not certify ontology change, and novelty does not certify scientific validity.

The landscape therefore supplies comparisons and evaluation instru-



ments, not a single definition to be adopted wholesale. Its output is a set of questions that every experiment should answer: what changed, relative to which representation, by what mechanism, and under which certificate? The following table makes the handoff to the next three chapters precise. Their difference is not how impressive an output looks, but what remains fixed, what persistent artifact is produced, and what additional evidence that artifact owes.

| Mode             | What remains fixed                                                                                                     | Persistent artifact                                                                                                          | Admission burden                                                                                                                          |
|------------------|------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------|
| Combinational    | source objects and their declared theories; an invented interface is recorded as an extension if it adds a type or law | a typed interface and reusable composite                                                                                     | legality, preserved obligations, nontrivial interaction, and transfer                                                                     |
| Exploratory      | the entire representational package and its legal moves                                                                | a path, construction, proof, policy, or consequence inside that package                                                      | validity, novelty relative to the registered space, independent value, and held-out consequence; search efficiency is reported separately |
| Transformational | no component is exempt, but every edit must be named and versioned                                                     | a versioned package comparison with transport, loss, and admission records; a theory extension when the presentation changes | failure of adequate in-language controls, independent discrimination, conservativity, and calibrated abstention                           |

Table 6.3: The handoff from the literature landscape to Boden's three modes. The modes classify changes relative to a registered representation; they are not rankings of historical importance.

The modes label segments of a versioned trace; they are not mutually exclusive classifications of whole systems. One episode may discover a typed interface, explore consequences made reachable by that interface, and later encounter evidence that forces a change of language. Boden's mode and the episode's Piagetian status are separate annotations: a combination may be assimilatory when its interface is already expressible, or proto-accommodative when it requires a proposed extension. The audit record should therefore mark where the maintained representation changes rather than assigning one creativity label retrospectively to the entire episode.

Chapters 7–9 now restate these modes in the common language of typed theories, infinitesimal probes, and admission. Part II then asks how those distinctions survive contact with executable DIAL algorithms.

### *Further Reading*

Boden and Wiggins develop conceptual-space accounts of combinational, exploratory, and transformational creativity [Boden, 2004, Wiggins, 2006]. Palm and Barto and colleagues treat novelty, information,



surprise, and intrinsic motivation; Abraham surveys cognitive and neuroscientific accounts, Sawyer connects individual cognition to collaborative and cultural processes, and Wilson offers a broader evolutionary and humanistic perspective [Palm, 2012, Barto et al., 2013, Abraham, 2018, Sawyer, 2012, Wilson, 2018]. Miller surveys computer-enabled creative practice across artistic media before the frontier-model era [Miller, 2019]. Piaget’s developmental theory and Drescher’s constructivist implementation supply the assimilation–accommodation lineage [Piaget, 1952, 1954, Piaget and Inhelder, 1974, Piaget, 1985, Drescher, 1986, 1991]. Ritchie and Jordanous address the evaluation of creative programs [Ritchie, 2007, Jordanous, 2012].

For computational scientific discovery, see Langley and colleagues, Schmidt and Lipson, AI Feynman, and the Robot Scientist [Langley et al., 1987, Schmidt and Lipson, 2009, Udrescu and Tegmark, 2020, King et al., 2009]. The Association for Computational Creativity maintains a broad bibliography and ICCC proceedings [Association for Computational Creativity, 2026b,a]. Von Neumann supplies the foundational self-reproducing-automata construction; Wolfram and Hedlund develop the simple-program and shift-equivariant viewpoints [von Neumann, 1966, Wolfram, 2002, Hedlund, 1969]. Dawkins’s “blind watchmaker” supplies the anti-teleological biological comparison [Dawkins, 1986]. For LLM-enabled science, see Zheng and colleagues, the SDE benchmark, Evo 2, and Biomni [Zheng et al., 2025, HKUST KnowComp, 2026, Song et al., 2025, Bixi et al., 2025, Huang et al., 2025]. Chomsky, McCarthy, and Zahavy motivate the structural critique; Wang and Ma supply a constructed-language counterpoint [Chomsky et al., 2023, McCarthy, 2007, Zahavy, 2026, Wang and Ma, 2026].



## 7

# Combinational Creativity

Combinational creativity constructs a new relation among existing objects, concepts, skills, or theories. Its central difficulty is not producing pairs. It is finding an interface through which the parts can interact without losing their types, and then showing that the resulting composite does something that neither part did alone. Boden's category is therefore best understood as a problem of *typed composition*, not as unconstrained association [Boden, 1998, 2004].

This is the first chapter in a three-part analysis of Boden's modes. Its distinctive artifact is an admitted *interface and composite*. The next chapter holds the generative language fixed and studies paths through it; the third asks when the language itself must change. Keeping those artifacts separate prevents every unfamiliar output from being called transformational.

### Design principle

A creative combination should expose four things: its ingredients, the interface that connects them, the obligations preserved by the connection, and a consequence that is not merely a restatement of either ingredient.

Here "new interface" does not automatically mean "new theory." An interface already expressible as a term, model element, or diagram enlarges the retained repertoire while the language remains fixed. Changing the doctrine, observer, semantic universe, operation, or admission interface begins a candidate package transformation. If that proposal adds or changes a sort, generator, equation, or other presentation component, it is a candidate theory extension. Independent tests

of transport and new consequences govern admission.

### 7.1 From juxtaposition to composition

Use the differential representational package introduced in Chapter 6:

$$\mathfrak{R}_{\text{diff}} = (\mathcal{S}, \mathfrak{D}, \mathcal{M}, F, \Lambda, \mathcal{G}, \mathcal{A}),$$

where  $\mathcal{S}$  is a sketch or theory presentation,  $\mathfrak{D}$  its preservation doctrine,  $\mathcal{M}$  a semantic universe,  $F$  the selected Weil action,  $\Lambda$  the observer family,  $\mathcal{G}$  the generative operations, and  $\mathcal{A}$  the admission procedures. Suppose  $X$  and  $Y$  are already represented in this package. Placing descriptions of  $X$  and  $Y$  next to one another is a juxtaposition. A combination requires a registered way for information, constraints, or action to cross between them.

In the simplest case, the interface is already present as a pair of maps

$$X \xleftarrow{i} I \xrightarrow{j} Y.$$

The object  $I$  names what the two domains share: a boundary state, a common variable, a unit convention, a protocol, or an abstract pattern. If the ambient doctrine admits the relevant colimit, a pushout may construct a glued object  $X \amalg_I Y$ . In another application the correct construction may be a pullback, tensor product, profunctorial composite, or ordinary substitution. The categorical form is not chosen for decoration; it states what is shared and which universal property the combination is expected to satisfy.

**Definition 7.1** (Typed creative combination). A typed creative combination of  $X$  and  $Y$ , relative to  $\mathfrak{R}_{\text{diff}}$ , consists of an explicit interface, an admissible composition operation in the declared theory, and an observer-level claim about the composite. It becomes an admitted combination only when the composition obligations and the claimed consequence pass the registered tests  $\mathcal{A}$ .

This definition separates four increasingly demanding cases:

1. *Juxtaposition*: two items are mentioned together, with no map between them.
2. *Legal composition*: a registered interface makes the construction well typed.
3. *Nontrivial interaction*: the composite has an observable property not reducible to the two isolated descriptions.
4. *Productive combination*: the interaction transfers, predicts, or reorganizes later constructions.

Only the last three are compositional achievements, and legality alone is a weak creativity criterion. A type checker can reject nonsense; it cannot certify significance.

## 7.2 *Interface discovery*

Creative reuse often depends on discovering the interface rather than merely selecting the ingredients. Two theories may use different names for the same role, the same name for incompatible types, or different observational scales. An interface proposal must therefore answer three questions:

*Alignment.* Which objects, operations, or invariants are being related?

*Variance.* Which directions are preserved, reversed, forgotten, or aggregated by the relation?

*Obligation.* Which diagrams must commute after the two domains are connected?

A useful interface may be a map already expressible in  $S$ . It may also require an intermediate object that neither source description contains. For example, two simulator modules may become composable only after introducing a common state variable and conversion law. In that case the creative act is not exhausted by composing old components: it includes proposing the mediating type through which they can communicate.

### Challenge

When does an invented interface count as a combination inside the current theory, and when does it count as a transformation of that theory? The answer is relative to the declared representational package, not to the surface form of the output.

## 7.3 *Assimilation and accommodation*

Piaget's distinction cuts across Boden's. If  $X$ ,  $Y$ , their interface, and their mode of composition already belong to  $\mathfrak{R}_{\text{diff}}$ , then forming and testing the composite is assimilatory. The learner has used the current schema to organize a new case. The result can be large, surprising, and historically important while remaining assimilatory at the chosen level of description.

Proto-accommodation begins when the proposed combination requires a change to what can be represented or composed. This may

involve adding an interface object, splitting an overloaded type, introducing a new conversion, or revising an equation that previously prohibited the interaction. Schematically,

$$S \xrightarrow{I} S^+$$

records that the combination became legal only in an extended presentation. The episode is combinational in Boden's sense because of how its ingredients are joined, but proto-accommodative in Piagetian terms because it changes the maintained schema. It becomes an accommodation only if the interface extension and its composite are independently admitted.

This cross-classification prevents a common mistake. Combining concepts from two fields does not automatically constitute transformational creativity. Conversely, an apparently modest bridge can be transformational if it creates a new sort of object or a new law of composition that persists in later work.

#### 7.4 *Differential tests for interaction*

The differential framework supplies a local test of whether two proposed ingredients interact. Let  $A, B \in \text{Weil}_1^{\text{tan}}$  be probe shapes realized by  $F(A)$  and  $F(B)$ . The monoidal structure compares the joint probe

$$F(A \otimes B)X \cong F(A)F(B)X$$

with the isolated responses. Depending on the application, an observer may test additivity, order sensitivity, mixed variation, bracket closure, or failure of a comparison square.

Write  $o \in \Lambda$  for an observer. An *interaction witness* is any registered comparison

$$\Delta_{A,B}^o(X)$$

whose null value means that the observed joint response is explained by the declared independent-composition law. A non-null value says that the two directions interact relative to that observer and law. It does not yet say that the interaction is valuable, causal, or globally integrable.

First-order probes ask whether each ingredient has a locally visible effect. Paired or higher-order Weil probes ask whether their combination introduces a mixed effect. Order-sensitive probes can distinguish  $F(A)F(B)$  from  $F(B)F(A)$  when the semantics permits such a comparison. This gives combinational creativity a diagnostic sequence:

1. verify that each ingredient has a typed local interpretation;
2. realize the joint probe through the proposed interface;

3. compare joint and isolated responses under selected observers;
4. localize any interaction or obstruction to the interface;
5. integrate the local candidate and test its finite consequences.

#### Boundary

A mixed derivative, commutator, or nonzero interaction residual is evidence of coupling, not a certificate of creativity. Its meaning depends on the chosen observer, baseline composition law, and semantics of the probes.

### 7.5 *Combination inside a tensor-product theory*

When the selected sketch doctrine admits the required tensor product, its semantics makes the fixed space of combination explicit. Let  $\mathbb{A}$  and  $\mathbb{B}$  present two registered families of operations. A model of  $\mathbb{A} \otimes \mathbb{B}$  then supplies both families and the specified interchange squares that determine how their composites are compared. Combinational creativity searches for a new term, diagram, policy, adapter, or mechanism inside this presented double theory; it does not yet alter either factor or the interchange law.

This separates two superficially similar outcomes. Discovering a useful composite already generated by  $\mathbb{A} \otimes \mathbb{B}$  is combinational. A proof, exhaustive finite analysis, or other registered nonrealizability argument showing that no model in the selected semantic class can realize the observed interaction without a mixed generator or revised interchange equation motivates a transformational proposal. Failure of a finite search does not. The distinction is semantic rather than textual: a long composite may remain inside the fixed theory, whereas one small new generator may change its model category.

Finite composability is still an admission issue. A novel local pairing may reduce the interchange defect at one state but fail when repeated, transported, or composed around a loop. Accordingly, combinational DIAL experiments should test whether the proposed composite integrates into a stable finite operation, not only whether its first-order interaction is favorable.

#### Admission contract

Call a DIAL result combinational only when the tensor factors and interchange presentation remain fixed, the constructed composite is novel relative to the registered repertoire, and its benefit persists under a finite composition or transport test. A required edit to either factor or to interchange begins a transformational trace.

### 7.6 *Combinations across the AGENTIC testbeds*

The testbeds expose interfaces under different evidence regimes.

*Mathematical theory construction.* The ingredients and their types can be formal. A proposal might connect two algebraic structures through a functor, adjunction, invariant, or shared presentation. Proof assistants can check legality and derive consequences, although mathematical interest remains an additional observer.

*Simulator-grounded discovery.* The ingredients are modules, mechanisms, or state variables. A proposed interface can be executed. Interventions on one module test whether predicted effects cross the boundary into the other, and held-out regimes test whether the coupling is more than a fitted coincidence.

*Scientific corpus to theory.* The ingredients and interface must first be extracted from language. A model may suggest that two mechanisms share a mediator or invariant, but textual co-occurrence is not enough. Source provenance, incompatible claims, and an external experiment or database must remain visible in the admission record.

*Generative visual languages.* The ingredients include visual objects, attributes, relations, prompt constructions, and image controls. Combining registered ingredients can yield novel images, but ARTISTIC asks a stronger question: whether the combination becomes a productive, counterfactually grounded constructor that transports across scenes and viewpoints. A new token that merely abbreviates an existing prompt is weaker evidence than a new visual-language operation.

### 7.7 *A combinational experiment protocol*

An experiment should compare at least four conditions: isolated ingredients, naive juxtaposition, a supplied interface, and a proposed interface. The last condition tests whether the system can discover the bridge rather than merely use one supplied by the evaluator.



**Experiment: Typed interface discovery**

**Epistemic status:** proposed benchmark design, not a completed result.

**Input:** two registered theories or modules, partial observations, and an interface meta-language in which the benchmark's target bridge is expressible. A separate transformational control may deliberately omit a required interface type.

**Proposal:** an interface object and maps, the resulting composite, and a predicted interaction; any extension of the registered meta-language must be tagged as a transformational proposal.

**Probes:** isolated and paired Weil profiles, order-sensitive tests when meaningful, and counter-witnesses targeted at the proposed interface.

**Admission:** type correctness, preservation of declared invariants, held-out interaction, transfer to a second instance, and calibrated abstention.

**Primary ablation:** remove the interface description and ask whether surface association alone yields the same prediction.

The evaluation should report more than success rate. It should measure the rank of the correct interface among proposals, the number of probes required, false combinations admitted under confounding, and whether the composite is reused in a later task. Reuse is especially important: it distinguishes a one-off association from a new compositional resource.

## 7.8 Failure modes and admission

Combinational systems face characteristic errors:

- *type erasure*: rhetorically similar concepts are treated as interchangeable;
- *interface hallucination*: a mediator is proposed without evidence that either domain supports it;
- *double counting*: overlapping mechanisms are composed as if they were independent;
- *observer capture*: novelty is visible only under the score used to generate the candidate;
- *nontransport*: the bridge works for one example but cannot be reused elsewhere.

**Admission contract**

Admit a combination only after recording the source objects, interface maps, preserved diagrams, interaction observer, counter-witnesses, and transfer test. If a new interface type was required, record the candidate theory extension and transport of prior models, and do not admit the combination until that presentation change has itself passed admission, rather than presenting the result as composition in the original language.

Combination asks whether relevant pieces can be connected through an admissible interface. Chapter 8 changes the creative object while holding the declaration fixed: it asks how active probes can reveal useful but previously unvisited paths and consequences within that space.

*Further Reading*

Boden's account distinguishes combinational creativity from exploration and transformation of a conceptual space [Boden, 1998, 2004]. Wiggins gives a formal framework for creative systems in terms of conceptual spaces and traversal rules [Wiggins, 2006]. Piaget's assimilation and accommodation provide the complementary distinction between using and revising a schema [Piaget, 1952, 1985]. The categorical and differential machinery used here is developed in the literature on tangent categories and Weil actions [Cockett and Cruttwell, 2014, Leung, 2017, 2018].

## 8

# Exploratory Creativity

Exploratory creativity searches a conceptual space without changing its generative rules. It includes much of theorem proving, program synthesis, symbolic regression, optimization, and evolutionary search. Optimization is one important instance, but exploration is broader: the artifact may be a proof, program, mechanism, or reachable construction rather than a minimizing parameter vector. A fixed space may contain consequences that no reasoner has encountered, and a short path to them may be extremely difficult to find.

Chapter 7 asked whether existing components could be joined through a new typed interface. Here the interface and generative language are already registered. The creative artifact is instead an admitted *path, construction, or consequence* within that language. Failure to find such a path is evidence about search only; it becomes evidence about the language only after the controls developed at the end of this chapter.

### Design principle

Exploration is meaningful only relative to a declared space. State the objects, legal moves, observers, stopping rule, and resource budget before claiming that a system has explored rather than retrieved or improvised.

### 8.1 The boundary of a conceptual space

Use the differential refinement of the representational package introduced in Chapter 6 and held fixed in Chapter 7:

$$\mathfrak{R}_{\text{diff}} = (\mathcal{S}, \mathcal{D}, \mathcal{M}, F, \Lambda, \mathcal{G}, \mathcal{A}).$$

The sketch  $\mathcal{S}$  declares the available types and operations;  $\mathcal{D}$  states what models preserve;  $\mathcal{M}$  interprets them; the Weil action  $F$  specifies local probes;  $\Lambda$  makes selected behavior visible;  $\mathcal{G}$  supplies legal moves; and  $\mathcal{A}$  determines which results may be retained. Exploratory creativity

varies a state, term, proof, program, or model while this package remains fixed. Adaptive selection may choose among registered probes or moves, but it may not silently add a probe shape, observer, or operation to the package. The newly reached state, theorem, or program may be retained as content of a model or repertoire; retaining it is not by itself a revision of the package that made the construction expressible.

One may represent legal moves by a directed graph, a transition category, or a space of skill rollouts. Let  $\mathcal{E}_{\mathfrak{R}_{\text{diff}}}(X)$  denote the states reachable from  $X$  by legal finite compositions of those moves. The notation is deliberately schematic: reachability may mean proof derivability, execution, simulator evolution, or integration of a local field.

**Definition 8.1** (Exploratory transition). A candidate exploratory transition  $X \rightsquigarrow Y$  is a path from  $X$  to  $Y$  using only operations and search controls declared in  $\mathfrak{R}_{\text{diff}}$ . It becomes an admitted creative transition relative to an evaluator when the path is valid, the path or its consequence is novel relative to the registered repertoire, the result is valuable under a registered observer, and it passes the registered admission procedure. These are separate obligations.

The second sentence matters. Reachability gives legality, not interestingness. A proof system can derive countless trivial variants; a grammar can generate an enormous number of worthless expressions. Exploratory creativity needs both a generative boundary and an evaluative boundary. Compression progress, surprise, and search efficiency may be useful observers, but none substitutes for the novelty-value judgment or the admission record.

## 8.2 *Assimilation as exploration*

Exploratory creativity is the clearest computational analogue of Piagetian assimilation. New observations and constructions are interpreted through an existing schema. Search may update a model, add a theorem to a library, or traverse a long trajectory, but the types and legal moves used to describe that achievement remain unchanged.

Assimilation is not necessarily small. A million-step proof search and a decades-long investigation can both remain inside a fixed mathematical language. Conversely, a one-line proposal can be proto-accommodative if it adds a primitive that changes which expressions exist. Size of the trajectory and level of representational change are independent variables.

Piagetian equilibration suggests a disciplined outer control. The system first attempts to absorb a discrepancy using paths in  $\mathcal{E}_{\mathfrak{R}_{\text{diff}}}(X)$ . Only persistent, independently reproduced failure supplies evidence that the current space may be inadequate. Even then, failure of search

is not proof of impossibility: the path may exist but be computationally inaccessible.

#### Boundary

Exhausting a search budget is not the same as exhausting a conceptual space. Accommodation requires evidence of representational inadequacy, not merely a timeout, a poor heuristic, or a weak proposal model.

### 8.3 Local exploration with Weil probes

The Weil profile from Chapter 3 turns the local neighborhood of a theory state into an explicit experimental object. For a finite probe family  $\mathcal{B} \subseteq \text{Weil}_1^{\text{tan}}$ ,

$$\mathcal{P}_{\mathcal{B}}(X) = \{F(A)X, F(u)_X \mid A, B \in \mathcal{B}, u : A \rightarrow B\}$$

records selected directions and relations among them. Exploration chooses a probe, observes a response, updates its local picture, and chooses again.

Different probe shapes support different questions. A first-order probe asks for a sensitive direction. Paired probes reveal interactions. Repeated or higher-order probes test curvature-like change, order effects, or failure of a linear approximation. Null probes identify variations invisible to the current observer. The collection is a typed experimental vocabulary, not a claim that all creative search is literally continuous.

The local-to-finite passage requires care. A tangent direction may not integrate, and a locally promising direction may terminate at a boundary or enter a region where the approximation fails. An exploratory trace should therefore separate:

1. the probe that proposed a direction;
2. the finite operation used to realize it;
3. the observations collected along the path; and
4. the admission test applied at the endpoint.

### 8.4 Exploration as an integrable path

Exploration is not merely a sequence of locally legal moves. In the DIAL setting, the current conceptual space is a model of a tensor-product repair theory  $\mathbb{A} \otimes \mathbb{B}$ . A tangent probe proposes a direction; an

exploratory trajectory requires those directions to compose into a finite path that remains inside the declared model space.

The relevant obstruction depends on the domain. A smooth path may encounter monodromy or fail to extend for the required time. A learned policy may be one-step improving but unstable under rollout. A local adapter law may fail when transported across linguistic charts. A causal proposal may cease to close under sequential interventions. These are all failures of finite realization, although they need not share one mathematical integrability theorem.

An explorer should therefore maintain two records: the local tangent evidence that chooses its next move and a finite-realization observer that audits the path already traversed. In this setting it is an integration observer: approximate holonomy, path-order disagreement, recurrence, multi-step regret, or held-out continuation can instantiate the latter. If the observer fails, the system should localize whether the problem is a poor move, a poor integration model, or a genuine limitation of the current theory.

#### Design principle

Exploratory creativity follows an admitted finite path inside a fixed theory. Local probe success chooses candidate directions; it does not certify that those directions compose into a coherent journey.

### 8.5 Active probe selection

Uniform exploration wastes experiments on redundant directions. An active system selects the next probe using its current uncertainty, the candidate it is trying to distinguish, and the cost of intervention. A generic policy has the form

$$A_{t+1} \in \arg \max_{A \in B_t} \left[ \text{Discrim}(A; H_t) + \lambda \text{Novel}(A; \mathcal{H}_t) - \mu \text{Cost}(A) \right],$$

where  $H_t$  is the current hypothesis set and  $\mathcal{H}_t$  the probe history. This is a decision rule, not a universal scalar definition of creativity. Each term must be supplied by the application.

Proposal-aware selection asks a sharper question than generic information gain: which available experiment most directly separates the leading structural proposals? In a simulator this may be an intervention. In mathematics it may be a small countermodel. In a corpus it may be a targeted search for a claim that distinguishes two mechanisms. Such probes connect exploration to admission rather than rewarding novelty for its own sake.

## 8.6 *Novelty, surprise, and interestingness*

Barto, Mirolli, and Baldassarre distinguish novelty relative to memory from surprise relative to prediction [Barto et al., 2013]. Both can guide exploration, but neither is sufficient for scientific value. A novel state may be irrelevant; a surprising result may expose sensor error; and an important result may become unsurprising once expressed in the right coordinates.

Exploratory systems should therefore maintain several observers rather than a single creativity score:

*Validity* asks whether the construction belongs to the declared space.

*Novelty* compares it with the retained repertoire and equivalence rules.

*Surprise* compares observation with prediction.

*Structural yield* measures new consequences, factorizations, or distinctions made available by the result.

*Transfer* asks whether the discovered path or object helps in a second problem.

Palm's information-theoretic treatment is useful for the novelty and surprise observers [Palm, 2012]. The other observers prevent an information measure from silently becoming a complete theory of creativity.

## 8.7 *From exploration to obstruction*

Exploration can reveal three different kinds of difficulty, which should not be collapsed:

1. *search failure*: no path was found under the available resources;
2. *model failure*: a candidate in the current language conflicts with evidence;
3. *language failure*: the registered objects and operations cannot express the distinction demanded by the evidence.

Only evidence for the third directly motivates opening a transformational proposal. Language failure is always relative to a registered syntax and semantic class; it should be certified by benchmark construction, an expressivity result, exhaustive finite analysis, or another explicit nonrealizability argument. Counter-witnesses still matter. One can vary the search policy, enlarge the computational budget, test alternate parameterizations, or ask whether a known solution is representable. A

robust obstruction survives the registered controls and localizes to a declared limitation of  $\mathfrak{R}_{\text{diff}}$ .

Nonclosure is one particularly informative signal. If admitted local moves compose to a direction not represented by the current vocabulary, the system has detected a mismatch between its local dynamics and declaration. Omitted structure is one hypothesis; observer error, estimation error, hidden state, or failure of a side law are alternatives. These must be audited before a new generator is proposed.

#### Creative transition

Exploration warrants opening a candidate transformational trace when it produces an auditable obstruction: a reproducible failure that survives the registered in-language controls and for which an explicit argument rules out the relevant class of ordinary repairs. This warrants proposal, not admission.

### 8.8 Exploration across the testbeds

*Mathematical theory construction.* Search traverses expressions, diagrams, lemmas, models, and countermodels. Legality is comparatively strong because proof or computation can often be checked. Interestingness and choice of definitions remain open-ended.

*Simulator-grounded discovery.* Search includes both hypothesis selection and experiment selection. The simulator supplies controlled interventions, but conclusions remain relative to its mechanisms and numerical regime. An exploit of the simulator is not a scientific discovery about the target world.

*Scientific corpus to theory.* Search traverses claims, entities, mechanisms, and sources. The space is only partly observed, provenance is heterogeneous, and absence of a statement is not evidence of impossibility. Exploration must therefore preserve uncertainty and source boundaries.

*Generative visual languages.* Search traverses prompts, seeds, layouts, scene-preserving edits, and rendering controls already available in the declared visual language. A visually unusual image may be a valuable reachable consequence without changing that language. When no registered path can satisfy a persistent semantic obligation, the trace should localize the failure and hand it to transformation rather than quietly adding a token or control.



**Experiment: Active exploratory creativity**

**Epistemic status:** proposed benchmark design, not a completed result.

**Input:** a versioned theory, an initial state, a finite probe budget, and hidden targets or held-out consequences.

**Policy:** choose sequential probes using uncertainty, proposal discrimination, novelty, and cost.

**Output:** a reachable construction, its path, probe history, observer scores, and any localized obstruction.

**Admission:** validity, registered novelty and value, held-out consequence, transfer, and calibrated abstention.

**Search comparison:** sample and compute efficiency against random and passive policies, reported separately from creative admission.

**Transformational control:** include instances whose target is not expressible in the supplied language; a sound explorer should diagnose or abstain rather than fabricate an in-language answer.

### 8.9 *The exploratory audit record*

A result is reproducible only if the conceptual space and search process are versioned. The audit record should contain the initial theory, legal move language, equivalence relation, proposal model, probe-selection policy, resource budget, random seeds where relevant, observer traces, failed paths, and final admission status. Without this record, retrieval, exploration, and post-hoc explanation are difficult to distinguish.

**Admission contract**

Call a result exploratory only when its derivation stays inside the declared representation. If the system adds a primitive, observer, probe shape, or preservation rule, split the trace at that point and treat the later search as exploration inside a proposed package transformation whose admission is still pending. Reserve *theory extension* for a change to the presentation or generative language that has passed the package-level admission contract; before that verdict, call it a candidate theory extension.

That split marks the boundary of exploration. Once the system must change the package in which paths and outcomes are expressed, the creative claim becomes a transformational proposal. Chapter 9 makes this transition explicit and states the additional transport and admission obligations created by package change, including the stronger case of theory extension.

*Further Reading*

Boden supplies the distinction between exploring and transforming conceptual spaces [Boden, 1998, 2004]; Wiggins formalizes spaces and traversal rules [Wiggins, 2006]. Classical discovery systems make the role of a registered search language explicit [Langley et al., 1987, Lenat, 1976]. Barto, Mirolli, and Baldassarre distinguish novelty from surprise, while Palm develops a broader information-theoretic account [Barto et al., 2013, Palm, 2012]. Piaget’s theory of equilibration motivates the passage from assimilation to accommodation [Piaget, 1985].

## 9

# Transformational Creativity

Transformational creativity changes the rules that generate a conceptual space. It is not necessarily harder than every exploratory achievement: a proof or search inside a fixed language may be extraordinarily difficult. Transformation instead carries the broadest representational burden because the proposal may alter not only which answers are available but also the language in which questions, evidence, and evaluation are expressed. The central artifact is therefore not a surprising output. It is a versioned theory-package revision together with a transport and admission record.

Relative to a frozen representational package and a versioned trace segment, the distinction from the preceding two chapters is exact. Combinational creativity adds a composite through an admissible interface; exploratory creativity adds a path or consequence while the representational package remains fixed. Transformational creativity changes that package. The burden of proof consequently rises: the system must say what changed, why ordinary search or repair was insufficient, and which earlier results survive.

### Creative transition

Transformation is not unconstrained mutation. It is an explicit intervention on a maintained theory, motivated by an obstruction, compared with less disruptive alternatives, and tested on evidence not used to invent it.

### 9.1 What can be transformed?

Let the current representation, with its differential semantics exposed, be

$$\mathfrak{R}_{\text{diff}} = (\mathcal{S}, \mathcal{D}, \mathcal{M}, F, \Lambda, \mathcal{G}, \mathcal{A}).$$

A transformation can act on any component:

*Presentation.* Add or split a sort, introduce an operation, weaken or add an equation, or designate a new universal construction.

*Semantic universe.* Move to models that support distinctions or internal constructions unavailable in  $\mathcal{M}$ .

*Differential semantics.* Add probe shapes, alter their interpretation, or replace the tangent action  $F$  by  $F^+$ .

*Observers.* Introduce a measurement or invariant, or revise a quotient, thereby changing which failures can be detected.

*Generative operations.* Add, remove, or retype legal moves, constructors, proposal operators, or rules of composition.

*Admission.* Revise the tests, evidence types, or preservation demands by which proposals are retained.

The audit record must identify the level. Otherwise ordinary parameter repair can be advertised as ontology change, while a genuine change in observability can be hidden as preprocessing. A proposed change to admission procedures must also be assessed under frozen prior criteria or an independent external procedure; it cannot validate itself merely by redefining success.

**Definition 9.1** (Transformational proposal). A transformational proposal is a versioned comparison

$$Y : \mathfrak{R}_{\text{diff}} \longrightarrow \mathfrak{R}_{\text{diff}}^+$$

that changes at least one component governing what may be represented, varied, observed, composed, or admitted. It specifies the motivating obstruction and the intended interpretation of the new structure. The arrow is package-level notation: no category of representational packages, and hence no formal class of morphisms between all such packages, is assumed here.

A package transformation need not alter the theory presentation. Changing an observer, probe semantics, inference policy, or admission interface can change the conceptual space in Boden's operational sense while leaving  $S$  fixed. This book reserves *theory extension* for the stronger admitted case in which the presentation or its generative language changes. Before admission, the same presentation-level change is a candidate theory extension; the audit record must say which claim is being made.

The proposal is not yet an achievement. A random edit to a grammar is transformational in a syntactic sense but not creative in the evaluative sense used here. Admission requires explanatory or constructive gain.

## 9.2 Declaration interventions

At the level of presentations, a common case is a sketch map

$$J : S \longrightarrow S^+.$$

Several forms of theory surgery should be distinguished.

1. *Definitional extension* adds names or operations explicitly definable from the old theory and is conservative under a stated definitional criterion.
2. *Constraint extension* adds an equation or universal property, excluding some old models.
3. *Ontology extension* adds a genuinely new sort, operation, latent object, or interface with several possible realizations.
4. *Resolution change* splits a type, passes to a quotient, or changes which distinctions count as observable.
5. *Doctrine change* alters the class of limits, colimits, exponentials, enrichment, or differential structure that models must preserve.

These operations have different epistemic force. Merely naming a definable term can improve reasoning without enlarging model content, but unique expansion by itself is only evidence for definitionality, not a proof of definitional or Morita equivalence. Adding an equation makes the theory more restrictive. A latent object may give some old models several inequivalent expansions and others none. A doctrine change can alter the kind of semantics available altogether.

## 9.3 Accommodation and equilibration

Piagetian accommodation provides the developmental interpretation. The current schema first attempts assimilation: can observations be reconciled by changing a state or model inside  $\text{Mod}(S, \mathcal{M})$ ? If persistent disequilibrium survives those repairs, the learner may revise the schema itself. In the present framework proto-accommodation means proposing  $J$ , not merely announcing a new verbal description; admitted accommodation additionally requires transport and independent testing.

Equilibration continues after the jump. The new representation must assimilate the evidence that motivated it, preserve or explain prior competence, and survive new tests. Thus accommodation is not a one-way escape from failure; it creates a new assimilatory regime whose stability must be demonstrated [Piaget, 1985].

**Design principle**

Prefer the least disruptive extension that resolves the registered obstruction. Escalate from model repair to observer change, ontology extension, or doctrine change only when simpler repairs fail under explicit controls.

#### 9.4 *From infinitesimal evidence to a finite jump*

Weil probes diagnose the local behavior of the current representation. They may reveal a null direction that should be visible, a mixed response absent from the declared composition law, a bracket that fails to close, or a comparison map that fails to be invertible. Such signals are valuable because they localize where the present geometry is inadequate.

They do not uniquely determine a replacement theory. Many global extensions can share the same local profile, and a finite conceptual jump need not lie on an integrable path in the old geometry. Frontier models, analogical retrieval, program synthesis, or human investigators must therefore propose candidate extensions. Differential diagnosis constrains this proposal process without pretending to derive the new ontology automatically.

Suppose the old and proposed differential worlds are

$$F : \text{Weil}_1^{\text{tan}} \rightarrow \text{End}(\mathcal{M}), \quad F^+ : \text{Weil}_1^{\text{tan}} \rightarrow \text{End}(\mathcal{M}^+).$$

Let  $K : \mathcal{M} \rightarrow \mathcal{M}^+$  be the semantic transport associated with the proposed theory change. A candidate differential theory extension includes comparison maps

$$\chi_A : KF(A) \Longrightarrow F^+(A)K.$$

The map  $\chi_A$  asks whether an old state probed and then transported agrees with that state transported and then probed in the new theory. Invertibility indicates strong preservation for probe  $A$ ; a controlled failure may reveal new directions introduced by the extension.

**Boundary**

Infinitesimal evidence constrains a creative jump but does not cause or certify it. Local defects underdetermine global ontology, and a proposed extension must be compared on finite models, interventions, or proofs.

#### 9.4.1 Transforming a factor, interchange, or domain theory

When the selected enriched-sketch doctrine admits the required tensor product, its semantics identifies several distinct transformational targets. If  $A \otimes B$  presents the current repair calculus, a proposal may:

1. extend or revise the assimilatory factor  $A$ ;
2. extend or revise the proto-accommodative factor  $B$ ;
3. change the interchange generators or equations relating the factors;
4. change the domain sketch  $S$  on which the calculus acts; or
5. change the enrichment, tangent doctrine, observers, or integration semantics needed to interpret the preceding structures.

These interventions have different empirical signatures. A side failure motivates repair of one factor. Two coherent sides with a persistent mixed defect motivate an interchange change. A coherent and integrable repair calculus, together with a registered expressivity argument ruling out the required scientific object, motivates a domain-theory extension. Experiments should not pool these diagnoses into one generic “schema change” label.

Integrability provides a gate between local diagnosis and declaration surgery. When the proposed infinitesimal repair integrates within the existing theory, no declaration change is yet warranted; the resulting work may be ordinary repair, combination, or exploration. When a registered finite-realization observer rejects it after the relevant controls, that failure is evidence for transformation but remains underdetermining: it does not say which factor, interchange law, or domain object should be added. Candidate extensions must still be generated, transported, and independently admitted.

#### Admission contract

A transformational DIAL trace must name the edited level, provide a counter-witness against adequate realization inside the old declaration—including finite integration when that is the relevant control—specify the new presentation and its model semantics, and test a finite consequence not used to choose the edit.

#### 9.5 Transport, forgetting, and conceptual loss

Every transformation creates a continuity question. Given  $J : S \rightarrow S^+$  and a compatible comparison  $\delta_{\mathcal{D}}$  between the old and new preservation doctrines, restriction induces, when well-defined,

$$J^* : \text{Mod}_{\mathcal{D}^+}(S^+, \mathcal{M}) \longrightarrow \text{Mod}_{\mathcal{D}}(S, \mathcal{M}).$$

The fiber test of Section 5.6 now has an epistemic use: it separates a conservative elaboration from a genuinely new choice and from a proposal that rejects or reinterprets an old model. In a transformational trace, that classification belongs in the admission record; it cannot be inferred from the novelty of the proposed syntax.

Not all loss is a defect. A successful theory can abandon an assumption, identify formerly distinct descriptions, or restrict models that conflict with evidence. The requirement is to make that loss explicit. The extension record should state:

- which old objects and equations are preserved;
- which are reinterpreted, quotiented, weakened, or discarded;
- which old results can be transported and by what map;
- which observations become newly expressible; and
- which countermodels distinguish the new theory from the old.

This makes transformational creativity cumulative without requiring every new theory to contain its predecessor literally.

### 9.6 *An obstruction-to-extension workflow*

A controlled transformational process can be organized into eight stages.

1. **Register.** Version the current theory, observers, probes, search policy, and admission criteria.
2. **Stress.** Explore the current model class using targeted Weil probes, countermodels, or interventions.
3. **Localize.** Identify a reproducible obstruction and the smallest declaration on which it depends.
4. **Control.** Test stronger in-language search, alternate parameterizations, measurement error, and hidden-state explanations.
5. **Integrate.** When a local repair is proposed, test whether it persists and composes into a coherent finite path, rollout, or other declared realization inside the old theory.
6. **Propose.** Generate a small family of typed extensions rather than one unqualified narrative jump.
7. **Transport.** Compute what each proposal preserves, changes, and newly predicts.



8. **Admit.** Use independent evidence to accept, reject, or retain the proposal as speculative.

This workflow gives frontier models an important but bounded role. They are powerful proposal engines and analogical retrievers. They do not get to define their own success after seeing the result.

### 9.7 *Transformations across the testbeds*

*Mathematical theory construction.* A definitional extension can reorganize reasoning without changing model content. A stronger transformation may add an axiom, object, equivalence, or doctrine that organizes previously disconnected results and changes the admissible models or constructions. Proof assistants can check derivations; model finders can search bounded classes for realizations or countermodels; relative-consistency arguments require explicit assumptions. None of these alone certifies that the new concept is fruitful enough to found a field.

*Simulator-grounded discovery.* A transformation may introduce an unmodeled mechanism, latent state, coupling, or scale. The simulator can execute the proposed extension and generate interventional predictions. Because the simulator is itself a theory, successful independent prediction can first establish only adequacy relative to that synthetic world.

*Scientific corpus to theory.* A transformation may turn dispersed textual claims into a new mechanistic object or causal relation. The risk of hallucinated ontology is highest here. Every new object must retain provenance, competing interpretations, and an external route to admission.

*Generative visual languages.* A transformation may add a visual relation, semantic control, observer, or reusable operation on the generation process. Repairing an image against a supplied scene theory remains an assurance task. A stronger claim requires showing that the old visual language cannot express the needed operation and that the proposed extension transports across prompts, subjects, and renderers while preserving previously admitted controls.

**Experiment: Withheld-language theory repair**

**Epistemic status:** proposed benchmark design, not a completed result.

**Input:** observations generated by a hidden theory and an initial sketch that deliberately omits a required primitive, relation, or observer.

**Diagnosis:** a localized obstruction plus a benchmark construction or formal argument certifying that the target is not realizable in the selected initial language; stronger-search failures are reported separately.

**Proposal:** several minimal sketch extensions with interpretations and predicted consequences.

**Transport:** restriction fibers, preserved results, conceptual losses, and probe comparison maps  $\chi_A$ .

**Admission:** held-out interventions or proofs that distinguish the extension from both the old theory and plausible alternatives.

**Safety condition:** abstain when the evidence does not discriminate among extensions.

### 9.8 *Evaluating transformational creativity*

Predictive accuracy on the motivating data is necessary but weak: a more expressive theory can often fit what exposed the failure. Evaluation should instead form a vector of criteria:

*Resolution* Does the extension remove the localized obstruction?

*Minimality* Could a less disruptive change do the same work?

*Conservativity* Which earlier results and skills survive?

*Novel consequence* Does the extension entail a test not used to propose it?

*Discrimination* Can an experiment separate it from rival extensions?

*Unification* Does it organize several failures through one construction?

*Persistence* Is the new object or operation reused in later tasks?

*Calibration* Does the system defer admission when evidence is inadequate?

No scalar combination of these criteria is canonical. The vector should remain visible so that a compact but destructive theory is not silently ranked above a conservative one, or a novel vocabulary above an empirically grounded extension. Unification may reduce description

length, but compression alone does not discharge the other admission obligations.

#### Admission contract

Admit a transformation only with a versioned package comparison—including a theory map when the presentation changes—a localized obstruction, comparison against in-language repairs, transport account, independent consequence, and explicit epistemic status. A fluent proposal without these components remains a hypothesis, not an admitted creative jump.

### 9.9 *The boundary of the claim*

The framework does not solve creativity by renaming it theory extension. It specifies a representation in which certain creative acts can be proposed, compared, and audited. It remains possible that the decisive new object is invisible to every current probe, that no available generator proposes it, or that its value becomes evident only after years of use. Those limitations are part of the challenge rather than reasons to collapse transformation back into search over a fixed output grammar.

Part II now turns these distinctions into algorithms. Its starting question is deliberately operational: for a particular domain, what state is maintained, which two local repair directions are realized, what mixed witness and finite test diagnose failure, which less disruptive controls must be exhausted, what typed package change may be proposed, what prior structure is transported, and what independent evidence can admit and retain the resulting construction?

### *Further Reading*

Boden introduces transformational creativity as change to the rules of a conceptual space [Boden, 1998, 2004]; Wiggins formalizes the relation between a space and the rules that traverse it [Wiggins, 2006]. Piaget's accommodation and equilibration provide the developmental analogue [Piaget, 1952, 1985]. Lawvere theories and sketches supply the theory-presentation semantics used for explicit transformations, while tangent categories and Weil actions supply the local diagnostic layer [Lawvere, 1963, Ehresmann, 1968, Cockett and Cruttwell, 2014, Leung, 2017].



## **Part II**

# **Algorithmic Realizations**



## *DIAL as an Algorithmic Family*

### FROM DOUBLE REPAIR GEOMETRY TO DIAL-X

Part I described creative modes and a geometry of interacting repairs; Part II must turn those distinctions into executable and falsifiable methods. A geometry is not yet an algorithm. To obtain one, DIAL must be specialized to a base learning system  $X$ . The specialization must state what  $X$  is trying to construct, which two repair processes are being compared, what makes their interaction observable, which structural changes may be proposed, which less disruptive alternatives must first be controlled, whether a local repair has a coherent finite realization, what would make such a proposal creative, what evidence can admit one of them, and what must persist after the verdict.

We write the resulting family as *DIAL-X*. Examples include DIAL-GIRL, based on Gradient Infinitesimal Reinforcement Learning; DIAL-ALLORA, based on A Lie-Algebraic Low-Rank Adaptation; and DIAL-SKFM, based on Spectral Kernel Flow Matching. The hyphen is important: DIAL does not replace the domain algorithm. It supplies a typed diagnostic, routing, and admission layer around domain operations whose meanings must already be declared.

Those specialization names also preserve the provenance of experiments run before the integrated systems received shorter names. The following chapters use CLIC for causal construction, OPTIC for skill construction, and RELIC for sequential decision structure; their experiment cards retain DIAL-SKFM, DIAL-SkillOpt, and DIAL-GIRL where those are the registered artifact names.

Throughout the book, a hyphen joins DIAL to its domain specialization, while an en dash introduces a registered experimental rung. Thus DIAL-SKFM is the system family and DIAL-SKFM–1C is one experiment; DIAL-URL is the coalgebraic construction and DIAL-URL–3 is one calibration within it.

This chapter inherits the formal-status convention stated at the opening of Chapter 3. The DIAL-X contract is a book-level algorithmic specification. Instantiating its fields yields an auditable method, but

does not prove that the proposed internal category of DIAL objects exists, that a mixed witness has a unique interpretation, or that the method will improve a particular domain system. Those are separate mathematical and empirical obligations.

### 10.1 *The missing algorithmic layer*

A purely formal DIAL object says that two involution-algebroid directions and an interchange comparison coexist. A computational system needs more. The four-phase reading scaffold of Chapter 0 becomes an implementation contract:

*Declare.* What object is being learned or constructed? Which changes count as motion inside the current declaration?

*Diagnose.* Which observer distinguishes compatible interaction from a persistent obstruction? Which registered controls rule out less-disruptive explanations?

*Construct.* Which finite-realization test distinguishes a coherent repair from a one-step improvement? Which typed proposals may be constructed, relative to which frozen reference space, and in which creative mode?

*Admit.* Which independent test can retain a repair or package change? Which declarations, failures, transports, and provenance persist after admission, rejection, or abstention?

Without these answers, a bracket, commutator, or interchange residual is only a number. It does not say whether to continue optimization, collect a new experiment, change a factorization, restore a prior state, or abstain.

*Chapter roadmap.* The first half turns DIAL's geometry into a four-level implementation invariant, target card, artifact dossier, evidential statuses, and generic algorithm. The second half specializes that contract to decision, adapter, causal, and skill construction before stating the boundary of the algorithmic claim.

#### Design principle

Never introduce DIAL-X by naming two operators alone. Declare its target, side semantics, mixed witness, finite-realization gate, creative mandate, proposal language, less-disruptive controls, package comparison, admission contract, and persistence semantics.



## 10.2 The four-level implementation invariant

Chapter 5 separated four operations that an implementation must not collapse. This chapter adopts that separation as its organizing invariant:

$$\underbrace{\mathbb{A}_X \otimes \mathbb{B}_X}_{\text{repair theory}} \longrightarrow \underbrace{\Theta_X(M)}_{\text{localized diagnosis}} \longrightarrow \underbrace{\gamma_X}_{\text{finite realization}} \longrightarrow \underbrace{Y_X : \mathfrak{R}_X \rightarrow \mathfrak{R}_X^+}_{\text{package comparison}}.$$

The arrows display workflow dependencies, not morphisms asserted to exist in one ambient category. When the selected enriched-sketch doctrine admits the required tensor product,  $\mathbb{A}_X$  and  $\mathbb{B}_X$  present the two typed repair families and their specified interchange laws. The maintained state  $M$  is a model of the domain theory  $S_X$ , equipped with a computational realization of those repair operations. The mixed witness  $\Theta_X(M)$  localizes a failure of the declared repair interaction on that model; it is not itself a change of theory.

A candidate local direction must next pass a *finite-realization observer*, also called an *integration observer*,  $\gamma_X$ : a registered finite-horizon test of whether repeated repair forms a coherent path, rollout, continuation, groupoid action, or other application-specific finite realization. Classical Lie integrability is the sharp mathematical model, but an implementation may use a justified empirical surrogate. Only after this test does the algorithm ask which component of the maintained representational package, if any, must change. The resulting proposal is a package comparison  $Y_X : \mathfrak{R}_X \rightarrow \mathfrak{R}_X^+$ . When the presentation or generative language changes, its presentation component is a typed map  $J_X : S_X \rightarrow S_X^+$ ; changes to semantics, observers, probes, doctrine, or admission occupy their own components. Transport and independent admission remain necessary in every case.

This yields three different negative outcomes. A side-coherence failure invalidates the proposed repair semantics. A mixed defect diagnoses an interaction that the current model does not explain. An integration failure shows that a locally meaningful candidate does not survive finite composition. None uniquely specifies a new object, axiom, or mechanism.

**Admission contract**

Every executable DIAL-X method must expose four records: the tensor-product repair declaration, the localized mixed witness, the finite-realization result produced by its registered observer, and—when proto-accommodation is proposed—the package comparison, its domain-theory component when applicable, and their independently tested consequences. A one-step improvement is not a realized repair, and a realized repair is not automatically a theory extension.

*10.3 Where creativity enters*

DIAL-X is not, by itself, a theory of creativity. It is a controlled mechanism for deciding when a maintained construction can be revised internally and when its space of admissible constructions may need to change. If its proposal language contains only familiar parameter updates, resets, and selections from a fixed catalogue, the resulting method may be adaptive and useful without being creative.

Creativity enters through the relation among an obstruction, a proposal language, and a reference space. Let  $\mathcal{R}_X$  record what is already available to the system—known artifacts, previously admitted theories, registered generators, and the current grammar of construction. A DIAL-X *creative mandate* is

$$\text{Cr}_X = (\beta_X, \mathcal{R}_X, \nu_X, \sigma_X),$$

where  $\beta_X$  declares one or more of Boden's combinational, exploratory, and transformational modes;  $\nu_X$  is a novelty observer relative to  $\mathcal{R}_X$ ; and  $\sigma_X$  records surprise or explanatory gain. Value is not folded into novelty or surprise: it is tested independently by the epistemic certificate  $\mathcal{C}_X$ .

The three modes locate creativity at different levels of the DIAL-X contract:

*Combinational creativity.* Construct a previously unavailable composite from declared generators. The declaration may remain fixed, but the artifact must be new relative to  $\mathcal{R}_X$ , not merely resampled.

*Exploratory creativity.* Reach a new and informative region generated by the current rules, often by sustained tangent exploration or a new sequence of typed operations.

*Transformational creativity.* Revise the generators, equations, factorizations, observers, or proposal grammar itself. This is the closest case to accommodation as theory change.

Thus accommodation is neither necessary nor sufficient for every creative episode. A novel recipe or adapter composition may be combinational while remaining inside a declaration. Conversely, switching to a previously stored schema may be accommodation without creativity. The strongest DIAL claim is more specific: a localized obstruction can provide a reason to search for a candidate outside the currently successful repair family, while typed proposal and admission rules keep that search auditable.

#### Boundary

DIAL-X does not turn every nonzero obstruction into a creative leap. The obstruction diagnoses pressure on the current construction; a proposal engine supplies candidates; a novelty observer locates them relative to a declared reference space; and independent admission determines whether any candidate is valuable. Remove any one of these distinctions and “creativity” becomes only a label for model change.

#### 10.3.1 The creative delta

For an application inherited from LINCS, the relevant question is not merely “What does DIAL add?” but “What becomes inventable that was not expressible before?” A useful one-sentence specification has three clauses:

X operates inside a declared language. LINCS-X makes a particular structural failure in that language visible. Creative DIAL-X may enlarge the language by inventing a specified kind of object or rule, whose value is then tested independently.

The phrase *may enlarge* matters. A DIAL-X system must still be allowed to assimilate, intervene, or abstain. Creativity is the additional capacity to construct an admissible candidate beyond the initial expressive boundary, not a requirement to redescribe every ordinary update as invention. In the catalogue below, *invent* names the kind of structure that the creative mandate permits the system to propose. It does not report that an experiment has already constructed and admitted such a structure.

*GIRL*. GIRL lifts Bellman and policy-learning structure to tangent and quotient spaces, exposing how a policy may vary inside the current decision language. Creative DIAL-GIRL may invent a state abstraction, option, action type, reward decomposition, skill boundary, or quotient that the original decision sketch could not express. The creative artifact is an enlarged decision theory, not merely a policy with higher return.

*ALLORA.* *ALLORA* makes the order dependence of typed low-rank adapters visible and tries to enforce their declared commutativity. Creative DIAL-*ALLORA* may invent a new adapter decomposition, shared mediator, routing object, adapter type, or composition law that makes a previously unavailable capability combination possible. The creative artifact is a revised grammar of adapter composition, not merely a better ordering of registered adapters.

*SKFM.* *BRIDGE*—Bracket Residuals for Interventional Discovery and Geometric Estimation—together with Spectral Kernel Flow Matching (*SKFM*) use intervention fields and Lie-bracket closure to diagnose when a proposed causal factorization is inadequate. Creative DIAL-*SKFM* may invent a previously unrepresented variable, latent dimension, mechanism family, factorization, or discriminating intervention. The creative artifact is a causal-theory extension with new experimental consequences, not merely the highest-scoring graph in a supplied class.

*LASKO.* *LASKO* represents typed skills as sections of a Lie algebroid and optimizes their locally available compositions. Creative DIAL-*LASKO* may invent a new skill type, interface, decomposition, control-flow construct, or workflow grammar when the registered skill vocabulary cannot express a successful repair. The creative artifact is an extended language of action, not merely a retuned skill program. Section 4.7 gives the corresponding training realization: DIAL-SkillOpt improves an external discovery skill on training and validation worlds, then freezes it before independent admission.

*LINCS-RLHF.* *LINCS-RLHF* diagnoses failures of a fixed reward or preference representation, including cycles and incompatible local orderings. Creative DIAL-*RLHF* may invent a relational, contextual, minimax, or otherwise nonscalar preference object together with observers that make it testable. The creative artifact is a new language for value, not merely another fitted reward model.

*RADAR.* *RADAR* diagnoses incompatibility between local optimization and global relational constraints. Creative DIAL-*RADAR* may introduce a new decision variable, decomposition, mediating object, or constraint interface that exposes a feasible decision unavailable in the original presentation. The creative artifact must preserve the governing constraints; relaxing them silently is not invention.

*SID.* *SID* detects failures of local records or models to descend to a coherent global instance. Creative DIAL-*SID* may invent a mediating schema object, cross-context map, provenance relation, or ontology refinement that resolves a localized overlap defect. The creative

artifact is an expanded integration theory whose descent can be replayed on independent records.

*CoLT.* CoLT makes inference and declaration repair explicit in an auditable reasoning sketch. Creative DIAL-CoLT may invent a new concept, warrant type, explanatory mechanism, axiom, or argument role when the current reasoning language cannot factor the evidence. The creative artifact is an admitted extension of the reasoning language, not an unusual continuation of a textual trace.

#### Design principle

For every DIAL-X name, publish a one-sentence creative delta: the fixed language, the LINCOS-visible obstruction, the newly inventable structure, and the independent test of value. If the newly inventable structure cannot be named, the creativity claim is not yet operational.

### 10.4 Three targets for every DIAL-X method

The word *target* has three meanings that should remain separate. For a base system  $X$ , define

$$\text{Target}_X = (Y_X, S_X, C_X).$$

*Operational target*  $Y_X$ . The deployable object: a policy, adapter composite, causal model, workflow, preference representation, recommendation, database instance, or reasoning trace.

*Structural target*  $S_X$ . The declaration that types the operational object: its objects, arrows, factorizations, equations, support restrictions, observers, and allowed interventions.

*Epistemic target*  $C_X$ . The certificate required to call the result admitted: held-out return, interventional prediction, descent, proof, safety constraint, simulator test, or a calibrated decision to abstain.

A scalar objective may help construct  $Y_X$ , but it does not by itself specify  $S_X$  or  $C_X$ . This is why the DIAL-2 experiments reported policy regret and structural-action accuracy separately. A method could lower regret by refitting every regime while changing the declaration when no structural change was warranted.

#### 10.4.1 The artifact dossier

The phrase *creative artifact* will therefore not denote an output object alone. A DIAL-X claim is attached to a versioned *artifact dossier*

$$\text{Art}_X = (y, Y_X, \text{Tr}_X, c_X, m_X),$$

where  $y \in Y_X$  is the operational construction,  $Y_X$  the package comparison,  $Tr_X$  the transport record,  $c_X$  the certificate record, and  $m_X$  the persistent provenance and replay state. The presentation component of  $Y_X$ , when changed, is a typed map  $I_X : S_X \rightarrow S_X^+$ ; changes to observers, doctrine, semantics, or admission occupy their corresponding components. Rejected and unsupported proposals retain dossiers recording why they were not admitted.

For combinational or exploratory creativity,  $Y_X$  may be the identity; for transformational creativity, a nontrivial typed  $Y_X$  is part of the claim. Novelty of  $y$  is never sufficient. The dossier keeps separate what was produced, what language changed, what prior knowledge survived, and what independent evidence warrants retention.

#### Design principle

When a chapter names a creative artifact, read it as shorthand for an artifact dossier. If only the operational output has been constructed, the text must say so and must not silently promote it to an admitted theory extension.

#### 10.4.2 Five evidential statuses

The applications use five status classes whose evidential force must not be blurred. A structure is *supplied* or *registered* when it enters the frozen experimental contract. It is *estimated* when its numerical value is recovered from finite observations inside a declared type. It is *learned* when a rule, representation, or routing policy is fitted from data. It is *constructed* when the system assembles an artifact that was absent from the frozen reference library, possibly from registered primitives. It is *admitted* only after the artifact passes the independent certificate declared for its intended use.

These terms form neither a linear pipeline nor a hierarchy of praise. An estimated operator may be rejected; a learned router may select only supplied objects; a constructed program may fail admission; and an admitted repair may be combinational rather than transformational. Conversely, abstention after a well-calibrated estimate can be the correct scientific outcome. Every experiment should therefore report the provenance and status of each object, not only the terminal score.

#### Design principle

Reserve *constructed* for an artifact absent from the frozen reference library, and reserve *admitted* for a decision made with evidence not consumed by proposal generation or ranking. Neither word, by itself, implies autonomous or transformational creativity.

The evidentiary verbs used below follow the same discipline. An experiment *reports* an observed outcome. That outcome *supports* a claim when it is compatible with the claim under stated controls. It *establishes* a claim only relative to a frozen contract whose registered gates have all passed. A component *satisfies* or *passes* a runtime contract; it does not thereby establish the broader theory that motivated the contract. *Proof* is reserved for a formal consequence of stated assumptions. Every stronger reading requires a separately named admission environment.

### 10.5 The DIAL- $X$ contract

An algorithmic specialization is the package

$$\mathfrak{C}_X = (\text{Target}_X, \text{Cr}_X, M_X, \mathbb{A}_X, \mathbb{B}_X, H_X, V_X, \Lambda_X, \Theta_X, \text{Loc}_X, \gamma_X, \mathcal{L}_X, \text{Ctl}_X, Y_X, \text{Adm}_X, \text{Mem}_X). \quad (10.1)$$

Here  $\text{Cr}_X$  may be empty for a purely adaptive DIAL application; when a creativity claim is made, it is a required part of the specification. The object  $M_X \models S_X$  is the maintained realization. The proposed categorical layer uses enriched sketches  $\mathbb{A}_X$  and  $\mathbb{B}_X$  to present the two repair theories; an executable specialization realizes their selected operations computationally as  $H_X$  and  $V_X$ . The observer family  $\Lambda_X$  makes their effects empirically or formally visible. The interchange witness  $\Theta_X$  compares their two orders, and  $\text{Loc}_X$  records the localized mixed support when such a localization is meaningful. The integration observer  $\gamma_X$  tests finite realization, and  $\mathcal{L}_X$  contains well-typed proposals. The control suite  $\text{Ctl}_X$  registers less disruptive in-language, measurement, hidden-state, no-change, and intervention-only alternatives. Writing  $\mathfrak{R}_X$  for the representational package specialized to  $X$ ,  $Y_X : \mathfrak{R}_X \rightarrow \mathfrak{R}_X^+$  records its change, with presentation component  $J_X : S_X \rightarrow S_X^+$  when applicable;  $Y_X$  is the identity for no-change, combinational, and exploratory dossiers.  $\text{Adm}_X$  assigns epistemic status from independent evidence, and  $\text{Mem}_X$  stores admitted declarations, states, failures, and transport maps.

The two sides need not literally be named assimilation and accommodation. They may be task and safety adapters, critic learning and policy transport, or local field estimation and global causal assembly. Assimilation and accommodation describe the eventual structural verdict. The application must say what its two sides mean before their interaction is interpreted.

When subtraction is meaningful, a computational proxy for interchange is

$$\Theta_X(M) = V_X(H_X(M)) - H_X^+(V_X(M)).$$

In a general categorical realization this expression denotes a comparison map rather than vector subtraction. The superscript on  $H_X^+$  records

that the horizontal operation may itself require transport after a vertical change. Only an admitted observer may turn the comparison into a numerical residual.

### 10.5.1 The target card

Before an experiment, publish the following target card: it is the interface between theory and experiment, and changing a field changes the algorithm.

| Field              | Question                                             | Required record                                                                                                                                   |
|--------------------|------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------|
| Operational target | What is deployed or retained?                        | Type, interface, and evaluation semantics of $Y_X$ .                                                                                              |
| Declaration        | What structural promises type it?                    | Versioned $S_X$ , observers, and intervention contract.                                                                                           |
| Creative mandate   | In what sense could a result be creative?            | Boden mode, reference space $\mathcal{R}_X$ , novelty observer $\nu_X$ , and surprise measure $\sigma_X$ .                                        |
| Repair theory      | Which two typed calculi interact?                    | Presentations $\mathbb{A}_X, \mathbb{B}_X$ , their tensor-product interchange, and realizations $H_X, V_X$ .                                      |
| Mixed witness      | How are the two orders compared?                     | $\Theta_X$ , observability gate, equivalence margin, and localization rule.                                                                       |
| Finite realization | Can the candidate survive composition?               | Integration observer $\gamma_X$ , horizon, uncertainty, and failure semantics.                                                                    |
| Proposal language  | What may change?                                     | Typed candidates in $\mathcal{L}_X$ , including the no-change proposal.                                                                           |
| Controls           | What must be ruled out before extension?             | Registered in-language, measurement, hidden-state, no-change, and intervention-only controls in $\text{Ctl}_X$ .                                  |
| Package comparison | How does a declaration or its interpretation change? | $Y_X : \mathfrak{R}_X \rightarrow \mathfrak{R}_X^+$ , its presentation component $J_X$ when applicable, model transport, and preserved structure. |
| Admission          | What evidence is independent?                        | Certificate family $\text{Adm}_X$ , splits, thresholds, and abstention rule.                                                                      |
| Persistence        | What survives a revision?                            | Memory, transport, regression tests, and provenance in $\text{Mem}_X$ .                                                                           |

The target card is a common interface, not a claim that every real-



ization has already reached its final creative target. The four principal implementations developed next occupy different points in the evidential sequence:

| System  | Supplied or registered                                             | Estimated or learned                                                          | Constructed and admitted                                                                        | Present boundary                                                               |
|---------|--------------------------------------------------------------------|-------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------|
| CLIC    | Repair templates, intervention library, unsupported alternatives   | Intervention operators, uncertainty, residual posteriors, active probe choice | Exact studies instantiate and admit typed repairs within the registered metalanguage            | No autonomous variable or mechanism type outside that metalanguage             |
| OPTIC   | Hypotheses, probes, renderer, skill vocabulary, typed subskills    | Routing and composition of supplied subskills                                 | No new interface or grammar has yet been independently admitted                                 | Mechanism evidence for factorization, not transformational skill invention     |
| RELIC   | Primitive actions, stage contracts, admissible-command channel     | Predicates, filters, proposal rankings, and policy behavior                   | A temporal composition and a missing task-stage operator, admitted up to behavioral equivalence | Construction remains inside registered primitives and contracts                |
| AGENTIC | Packet schemas, probe and action vocabularies, workflow interfaces | Finite-sample causal, skill, and decision packets; active evidence allocation | A response-conditioned policy absent from the baseline declaration                              | Synthetic proof of composition; no invented sensor, action, or packet language |

This ledger prevents evidence from being promoted as it moves downstream. For example, an admitted RELIC policy does not retroactively admit the causal explanation that motivated its probe, and wiring CLIC, OPTIC, and RELIC into AGENTIC does not upgrade registered component vocabularies into invented ones.

## 10.6 The generic DIAL-X algorithm

---

**DIAL-X structural diagnosis, repair, and admission**

**Require:** A DIAL-X contract  $\mathcal{C}_X$ , maintained realization  $M$ , evidence  $E$ , a creative mandate when creativity is claimed, and a registered resource budget.

- 1: Initialize the candidate-dossier set  $\mathcal{P} \leftarrow \emptyset$ .
- 2: Validate the declaration, observer semantics, and intervention contract for the current regime.
- 3: Estimate the horizontal and vertical side operations separately, including uncertainty and side-coherence certificates.
- 4: **if** either side certificate fails **then**
  - 5:   reject its proposed semantics; repair the estimator, gather a typed probe, or abstain before interpreting mixed structure.
  - 6:   persist the failed side certificate and **return** an evidence request or auditable abstention.
- 7: **end if**
- 8: Estimate the interchange comparison  $\Theta_X$  from the two routes; record its observable localization  $\text{Loc}_X$  when possible.
- 9: **if** the mixed witness is practically equivalent to zero **then**
  - 10:   close the structural-repair branch and allow  $X$  to continue ordinary learning or search inside the current declaration.
  - 11:   add any independently proposed in-declaration dossiers to  $\mathcal{P}$ ; the zero witness supplies no novelty claim.
- 12: **else if** the witness is persistently nonzero and localized **then**
  - 13:   construct the typed in-declaration candidate set  $\mathcal{L}_X^{\text{in}} \subseteq \mathcal{L}_X$  and execute the applicable registered controls in  $\text{Ctl}_X$ .
  - 14:   Test each candidate with the registered finite-realization observer  $\gamma_X$  over a finite path, rollout, or persistence horizon.
  - 15:   **if** at least one candidate has a coherent finite realization inside the current declaration **then**
    - 16:     add the realized candidate dossiers to  $\mathcal{P}$  for independent admission.
    - 17:     mark them as in-declaration proposals; do not infer theory change from the original defect.
    - 18:     if creativity is claimed, classify them separately as combinational or exploratory relative to  $\mathcal{R}_X$ .

```

19:   else if the registered control suite has ruled out its specified less disruptive alternatives and an
    expressivity result, benchmark construction, or exhaustive finite analysis certifies that the relevant
    in-declaration repair class cannot realize the target then
20:       propose typed package comparisons  $Y_X : \mathfrak{R}_X \rightarrow \mathfrak{R}_X^+$ , with a presentation map  $J_X : \mathbb{S}_X \rightarrow \mathbb{S}_X^+$  when
    applicable, including the no-change and intervention-only alternatives.
21:       add every well-typed extension dossier to  $\mathcal{P}$ , with its transport obligations and transformational
    status still pending.
22:   else
23:       persist the unresolved integration record and return an evidence request or abstention without
    naming an accommodation.
24:   end if
25: else
26:     select a discriminating experiment and return its request, or persist an auditable abstention when no
    such experiment is available.
27: end if
28: if  $\mathcal{P} = \emptyset$  then
29:     persist a no-change record and return the current admitted state or an auditable no-op.
30: end if
31: For each candidate presented as creative, record its Boden mode and evaluate novelty against  $\mathcal{R}_X$ ; do not
    infer novelty from obstruction magnitude.
32: Include current, no-change, and previously admitted candidates whenever their types permit comparison.
33: Evaluate proposals using evidence not used to generate or rank them.
34: Admit, reject, defer, or request information; do not force a binary choice when the certificate is unresolved.
35: Persist the repair declaration, mixed and integration records, package comparison, learner-state transport,
    control results, evidence, rejected alternatives, and epistemic status.
Ensure: A status-bearing terminal record: an admitted, rejected, or deferred candidate dossier; a discriminating-
    experiment request; an auditable abstention; or a no-change result.

```

---

The ordering of the gates is part of the algorithm. A mixed residual computed after a malformed side has no DIAL semantics. Failure to enumerate a repair under a finite search budget is not proof that the language is inadequate; only the registered nonrealizability gate opens the declaration-extension branch. Stronger-search controls may expose proposal failure, but their failure alone does not establish expressive failure. Likewise, a repair selected on the same finite data that generated it has not undergone independent admission.

A zero mixed witness closes only DIAL's structural-repair route. It does not imply that the domain has no combinational or exploratory opportunities: the base system may still construct a new composite or traverse a new path inside the fixed declaration. Such an artifact receives its own novelty and value tests, but its creativity cannot be attributed to a nonexistent obstruction.

#### 10.6.1 Four control actions

The algorithm exposes four actions rather than a forced update:

*Assimilate.* Update  $M_X$  inside the current declaration.

*Propose proto-accommodation.* After finite diagnosis and a registered nonrealizability gate, propose a typed change  $J : S_X \rightarrow S_X^+$  and transport compatible evidence.

*Intervene.* Acquire a probe designed to separate currently admissible explanations or repairs.

*Abstain.* Record that the observer or intervention contract does not identify an admissible action at the current budget.

These actions make DIAL-X closer to an experimental controller than to a regularizer. A regularizer may be one proposal inside the assimilatory branch; it is not the complete algorithm.

#### 10.7 DIAL-GIRL: the target is an admitted decision state

GIRL supplies tangent diagnostics and quotient-aware learning for sequential decision systems. Its DIAL specialization targets

$$\text{Target}_{\text{GIRL}} = ((Q, \pi), S_{\text{dec}}, C_{\text{decision}}).$$

The operational object is a critic-policy pair. The declaration records the state and action types, transition support, Bellman and policy-evaluation diagrams, intervention semantics, and quotient geometry. Admission combines structural routing evidence with independent policy return, coherence, and effective-sample-size diagnostics.

In the first realization,  $H_{\text{GIRL}}$  is a within-declaration advantage-quotient critic and actor update. The second operation  $V_{\text{GIRL}}$  transports, forks, restores, or revises the learner state when the regime or its declaration changes. The mixed question is:

transport after learning  $\iff$ ? learn after transport.

A failure of this comparison does not automatically authorize reset or rollback. It creates a candidate-routing problem.

The DIAL-2 evidence sequence illustrates the distinction. Single-change experiments separated reward and parameter drift from mechanism and topology change. In A–B–A worlds, exact restoration of A was sometimes worse than continual GIRL because B had supplied transferable learning. The paired routing study therefore compared saved, current, and transported states with disjoint proposal and admission trajectories. It selected saved A under antagonistic forgetting and transported learning under positive transfer.

The four implementation levels are consequently distinct. The GIRL tangent and quotient laws present the repair theory; disagreement between learn–then–transport and transport–then–learn supplies the mixed witness. Its integration observer is a finite rollout or recurrence test: does the proposed learner-state transport remain coherent over several Bellman and policy updates, including a return to a previous regime? Only a reproducible failure of this finite test places pressure on the decision sketch itself. Candidate extensions may then introduce a new option, quotient, state abstraction, or reward component and transport prior learner states along that explicit map.

#### Admission contract

DIAL-GIRL targets neither maximal return nor recurrence recognition alone. It targets a critic–policy state whose declaration route and deployment value are independently admitted. A remembered state is a proposal, not a rollback command.

The completed DIAL-2 experiments are therefore calibrations of the repair and memory mechanism, not yet demonstrations of creativity. They supply a precursor to RELIC, whose embodied experiments in Chapter 13 move from supplied learner-state alternatives to the construction and admission of temporal decision structure. A stronger DIAL-GIRL experiment would invent a state abstraction, option, skill boundary, reward decomposition, or quotient absent from  $\mathcal{R}_{\text{GIRL}}$ . Reusing a saved policy is adaptation; constructing and admitting a new decision vocabulary is a candidate transformational episode.

A remaining mathematical requirement is quotient-aware functional transport. Raw addition of tabular learner states is coordinate de-

pendent and therefore serves only as a control, not as the intrinsic DIAL-GIRL construction.

### 10.8 DIAL-ALLORA: the target is an admitted adapter composition

ALLORA studies ordered adapter composition. Its operational target is a deployable composite of task, domain, capability, or safety adapters. Its structural target is the adapter incidence and factorization sketch, including which adapters share a backbone, which quotient directions are identified, and which orders are promised to agree. Its epistemic target combines held-out task behavior, safety constraints, order robustness, and resource cost.

For two typed adapter updates  $U_i$  and  $U_j$ , the two DIAL sides are their separately coherent specialization processes. The first mixed proxy is the order defect

$$\Theta_{\text{ALLORA}} = U_j \circ U_i - U_i \circ U_j,$$

augmented by cross- and nested-bracket observers when the application declares them. Small or quotient-null interaction may be assimilated by an ALLORA regularizer or order schedule. Persistent localized interaction may instead motivate a finite composition test. The system applies the proposed composite over a registered sequence of contexts or layers and tests path-order stability, capability retention, and safety persistence. If the candidate remains coherent, it is an integrated repair inside the current adapter theory. Reproducible failure triggers stronger-search, observer, and representability controls. Only a registered argument ruling out the relevant in-theory repair class opens a typed proto-accommodation proposal: refactor the adapters, introduce a shared module, change the quotient, split capability from safety, or revise the declared composition graph.

The target is not a zero commutator at any cost. Two adapters can commute by becoming behaviorally useless. Conversely, a nonzero interaction may be admissible when it is typed, localized, and improves the declared task without violating safety. DIAL-ALLORA therefore retains both the structural witness and the behavioral certificate.

Here the creativity distinction is especially concrete. Searching for a new useful ordering or composite of known adapters is combinational creativity if the composite is absent from the reference library. Exploring a large typed composition space is exploratory creativity. Introducing a new shared module, adapter type, or composition rule is transformational. Merely selecting the best registered ordering is optimization, even when DIAL supplies the routing signal.

### 10.9 CLIC: creative causal learning

CLIC—*Causal Learning with Infinitesimal Creativity*—is the DIAL-SKFM realization developed in Chapter 11. Its target is an interventionally admitted causal structure rather than merely the highest-scoring graph. One repair direction estimates local intervention fields and their uncertainty; the other assembles or revises their global causal factorization. Lie-bracket nonclosure supplies a localization signal, but not by itself a causal interpretation.

CLIC makes the architectural creative mandate explicit. After a registered nonrealizability gate rules out the relevant current-sketch alternatives, it may propose a missing directed mechanism, latent common-cause bundle, altered latent dimension, or revised factorization. Unsupported alternatives participate in inference and may request new evidence or veto construction, but they are not silently converted into admissible repairs.

The controlled ladder in Chapter 11 keeps the repair metalanguage registered while isolating four burdens: exact closure repair, operator uncertainty, active counter-witness interventions, and explicit unsupported alternatives. It tests typed construction and guarded admission inside that metalanguage; it does not demonstrate invention of a new causal variable or mechanism type.

### 10.10 OPTIC: creative skill optimization

OPTIC—*Optimizing Programs Through Infinitesimal Creativity*—is developed in Chapter 12. Its maintained artifact is an executable typed skill program together with the workflow sketch in which that program is meaningful. LASKO supplies the local Lie-algebroid geometry, and DIAL-SkillOpt supplies a versioned training discipline. OPTIC adds the explicit mandate to construct a new skill, interface, macro, factorization, or workflow rule when a registered expressivity test rules out the relevant repairs inside the current skill theory.

At the proposed theoretical layer, the primitive infinitesimal structure is an involution algebroid internal to a tangent category rather than only a realized vector-field bracket. Current paired-execution experiments expose a section-level computational shadow of its coherence; they do not establish the internal construction. Their mixed witness tests interaction between skill optimization and workflow accommodation, supplying initial mechanism evidence while exposing the limits of monolithic skill editing.

### 10.11 Further members of the DIAL-X family

The same contract yields additional algorithms, but only after their targets are stated. The following table is a research specification, not a claim that all of these systems have already passed a complete DIAL ablation.

| Method     | Operational and structural target                                                                 | Two repair processes and mixed witness                                                                                           | Admission target                                                                                               |
|------------|---------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------|
| DIAL-RLHF  | Scalar reward model when admissible, otherwise a relational or minimax preference representation. | Preference fitting and relation/observer repair; cycles and incompatible local orders form the obstruction.                      | Held-out preference prediction, exploitability, subgroup consistency, and calibrated refusal of scalarization. |
| DIAL-RADAR | Feasible recommendation or decision at a relational apex together with its constraint sketch.     | Local-chart descent and global relational constraint repair; disagreement on overlaps is mixed evidence.                         | Exact feasibility, chronological utility, constraint preservation, and robustness.                             |
| DIAL-SID   | A descended, provenance-bearing global foundry instance.                                          | Local equalizer repair and cross-context data migration; overlap and cocycle defects localize failure.                           | Descent, source preservation, contradiction reporting, and replay on independent records.                      |
| DIAL-CoLT  | An auditable reasoning trace and, when required, a revised reasoning sketch.                      | Inference within a declaration and typed declaration extension; observation–intervention comparison detects incompatible routes. | Grounded warrants, transported evidence, counterexamples, and explicit constructive or speculative status.     |

These targets also reveal when two names describe different algorithms. For example, DIAL-LASKO with a fixed skill vocabulary is not the same method as DIAL-LASKO permitted to invent new skill types. Their proposal languages and structural targets differ even if they share an optimizer.

They also reveal different creativity claims. DIAL-LASKO can merely tune registered skills or invent a new typed skill; DIAL-RLHF can fit a



reward model or replace scalarization with a new relational representation; DIAL-CoLT can search within a reasoning sketch or extend the language of warrants. The second alternative in each pair requires a declared reference space and a transformational mandate rather than retrospective creative language.

### *10.12 How DIAL-X sharpens an experiment*

Once the target card is fixed, the ablation is almost forced. A credible DIAL-X evaluation separates:

1. fixed-declaration learning;
2. one typed side;
3. a capacity-matched but untyped two-operator learner;
4. two typed sides without the mixed witness;
5. mixed diagnosis without structural repair;
6. local repair without a finite-realization observer;
7. realized repair without an explicit package comparison;
8. package comparison without independent admission; and
9. the complete DIAL-X workflow.

Negative controls shuffle the mixed witness, corrupt one side, change the presentation, suppress a required observer, and supply an exact witness as a diagnostic ceiling. Resource matching includes interventions, validation samples, memory, and compute. Otherwise DIAL can appear to win merely because it receives a richer experiment or maintains more learner states.

A creativity experiment requires one further contrast: the same DIAL controller with a fixed proposal vocabulary versus an expandable one. The evaluation must report how often each system produces a genuinely new candidate relative to the frozen reference space, how far that candidate extends or transforms the space, and how often independent admission retains it. This separates the ability to detect disequilibrium from the ability to construct a valuable response to it.

**Experiment: Minimum DIAL-X experimental record**

**Epistemic status:** reporting template; the fields below are required for a DIAL-X claim but do not themselves constitute experimental evidence.

**Target:** publish  $(Y_X, S_X, C_X)$ .

**Creativity:** when claimed, publish  $Cr_X$ , freeze the reference space, label the Boden mode, and score novelty separately from value.

**Sides:** type  $H_X$  and  $V_X$ , test each coherence law, and report its uncertainty separately.

**Interaction:** estimate  $\Theta_X$  only after both side gates pass; test localization and presentation invariance.

**Integration:** register the finite path, rollout, or persistence observer  $\gamma_X$ ; report local success separately from finite coherence.

**Action:** report assimilation, proto-accommodation proposals, admitted accommodations, interventions, and abstentions separately.

**Proposal:** include current and no-change controls; record the exact language from which repairs were drawn, the registered control suite, the package comparison for every declaration change, and its presentation map when applicable.

**Admission:** use evidence not consumed by proposal generation or ranking.

**Persistence:** measure transport, regression, and reuse rather than counting storage alone.

This discipline turns a vague application claim into a falsifiable algorithm. For DIAL-GIRL, it exposed the difference between recognizing an old regime and benefiting from an old state. For DIAL-SKFM, it separates detecting nonclosure from identifying a causal explanation. For DIAL-ALLORA, it prevents reduced order sensitivity from standing in for a useful adapter composite.

### 10.13 *Boundary of the algorithmic claim*

DIAL-X is an architectural schema whose exact realization depends on  $X$ . The generic algorithm does not prove that every application admits a double involution algebroid, that its numerical residual is invariant, that a local repair integrates, or that its proposal language contains the required creative repair. Nor does tensor symmetry make the two realized orders equal: the interchange law is a declaration whose satisfaction must be observed. Those are application-level obligations.

This boundary is also the answer to the chapter's motivating question. The creative contribution is not the residual and not the word *accommodation*. It is the construction and admission of an artifact or

theory that was unavailable in the declared reference space. DIAL contributes a disciplined trigger, a typed site of revision, and an audit trail for that event. The proposal engine may be symbolic search, a frontier model, a simulator-guided scientist, or a human collaborator; its creative reach must be tested rather than assumed.

Nor does DIAL-X manufacture observability. If no probe distinguishes two declarations, the correct output is an intervention request or abstention. Finally, a successful repair in one coordinate presentation does not establish intrinsic transport. The DIAL-GIRL state experiments make this caveat concrete: raw parameter addition can be an informative control while still failing to define a presentation-independent learner-state geometry.

The remaining chapters of Part II now have a precise role. CLIC, OPTIC, and RELIC instantiate target cards and domain-specific side operations; AGENTIC then tests whether their certified outputs can be composed without collapsing their evidence types. Part III carries those constructions into controlled testbeds. None tests DIAL in the abstract: each must distinguish local diagnosis from finite realization and compare the resulting package changes, including theory extensions when present, with registered ablations.

### *Further Reading*

LINCS develops the categorical foundations and six-stage diagnosis-and-repair workflow [Mahadevan, 2026e]. LASKO supplies the Lie-algebroid skill-optimization setting used by DIAL-LASKO [Mahadevan, 2026d]. Tangent categories and involution algebroids provide the one-direction categorical semantics, while classical double Lie algebroids and their Weil-algebra realization motivate the two-direction comparison [Cockett and Cruttwell, 2014, Burke and MacAdam, 2019, Mackenzie, 2005, Meinrenken and Pike, 2021].



## *Causal Learning with Infinitesimal Creativity*

### CLIC: FROM CLOSURE DEFECTS TO ADMITTED CAUSAL STRUCTURE

Many causal-discovery procedures select a graph or equivalence class from a declared model family. CLIC asks a stronger question: what should a discovery system do when intervention geometry supplies evidence that the declared causal language itself is inadequate? The answer developed here is to treat nonclosure as a typed diagnostic, actively seek counter-witness interventions, and admit a structural extension only when uncertainty and unsupported alternatives have been made explicit.

CLIC abbreviates *Causal Learning with Infinitesimal Creativity*. It is the DIAL-SKFM member of the algorithmic family specified in Chapter 10. Its experiments form the causal counterpart to RELIC: RELIC studies structural accommodation in sequential decisions, whereas CLIC studies structural accommodation in intervention-driven causal inference.

In one sentence, CLIC asks whether a reproducible failure of intervention fields to fit the current causal declaration can be localized, challenged by a discriminating intervention, and resolved by a finite causal proposal that survives evidence withheld from its construction. Each clause is a separate gate; nonclosure alone answers none of them.

#### 11.1 *The CLIC target card*

BRIDGE/SKFM estimates intervention-induced vector fields and uses Lie-bracket geometry to screen causal structure before downstream graph scoring. The corresponding CLIC target is

$$\text{Target}_{\text{CLIC}} = ([G]_{\mathcal{I}}, \mathcal{S}_{\text{causal}}, \mathcal{C}_{\text{int}}),$$

where  $[G]_{\mathcal{I}}$  is an interventional equivalence class rather than necessarily one privileged graph. The declaration records observed and latent variable types, intervention targets, admissible mechanisms, support

and factorization assumptions, and the observer making a closure defect visible.

Operationally, a CLIC run receives a versioned causal sketch, intervention records or estimated intervention operators with uncertainty, and a registered language of experiments and repairs. It returns a status-bearing causal packet: a repair proposal, a request for a discriminating intervention, or an abstention, together with the evidence used to reach that status. It does not take an observational data table and emit a certified ground-truth graph.

In the contract of Chapter 10,  $\mathbb{A}_{\text{CLIC}}$  presents local intervention-field estimation and  $\mathbb{B}_{\text{CLIC}}$  presents global causal assembly. Their computational realizations estimate the local operators and revise the factorization, respectively. The observer family records operator, support, and overlap evidence;  $\Theta_{\text{CLIC}}$  is the registered closure or overlap defect;  $\gamma_{\text{causal}}$  tests finite consistency; and  $\text{Adm}_{\text{CLIC}}$  assigns status using evidence withheld from proposal construction. Thus CLIC specializes the generic contract rather than introducing a second causal workflow beside it.

Lie-bracket nonclosure, failure of a latent basis, or an inconsistent overlap supplies a mixed witness. The localization may suggest a missing edge, latent common cause, incorrect latent dimension, nonlinear mechanism, or inadequate intervention. It does not identify one of these explanations by itself.

The CLIC creative mandate freezes the initial causal sketch, repair metalanguage, intervention library, and reference family of supported structures. A result is combinational when it assembles a new causal repair from that registered vocabulary and transformational only when it changes the registered package. Novelty is measured relative to this frozen reference; interventional value is tested separately.

DIAL-SKFM must therefore choose experiments as well as graphs. Its registered control suite  $\text{Ctl}_{\text{CLIC}}$  includes the current and no-change structures, stronger in-sketch factorization search, operator and support diagnostics, alternate latent dimensions, nonlinear or unmodeled intervention explanations, and equal-cost random probes. Observational equivalence calls for abstention or a discriminating intervention. A causal extension proposal is evaluated through held-out interventional prediction, equivalence-class coverage, and those registered controls. When the causal presentation changes, the presentation component is

$$J_{\text{causal}} : \mathbb{S}_{\text{causal}} \longrightarrow \mathbb{S}_{\text{causal}}^+$$

of a package comparison  $Y_{\text{CLIC}} : \mathfrak{R}_{\text{CLIC}} \rightarrow \mathfrak{R}_{\text{CLIC}}^+$ . Changes to intervention semantics, observers, or the model doctrine occupy their own components rather than being hidden in  $J_{\text{causal}}$ . The package comparison remains a proposal until its added variable, mechanism,

factorization, or observer makes interventional consequences that survive independent tests.

Before such a map can support admitted accommodation, a causal integration observer  $\gamma_{\text{causal}}$  must test whether the candidate local repair persists across intervention targets, overlapping charts, and finite intervention sequences. Depending on the experiment, this observer may use exact closure, approximate holonomy, path consistency, or held-out multi-intervention prediction. Without a registered  $\gamma_{\text{causal}}$ , the output remains a localized causal-repair candidate rather than a coherent finite causal construction.

The experiment artifacts retain their original DIAL-SKFM identifiers. Within the book, the complete architecture is called CLIC.

In the dossier terminology of Section 10.4.1, the operational object is a causal repair, equivalence class, experiment request, or abstention. The structural delta is  $Y_{\text{CLIC}}$ , with typed causal map  $J_{\text{causal}}$  when its presentation changes. Transport records which variables, mechanisms, and interventional evidence survive that comparison. The status-bearing admission record separately records held-out interventional consequences and unsupported alternatives.

The evidentiary status of the present evidence is narrower than that architectural target. Its repair families, unsupported alternatives, and intervention libraries are registered in advance. The experiments test typed diagnosis, active evidence acquisition, uncertainty propagation, and guarded admission inside that metalanguage; they do not yet show autonomous invention of a causal variable or mechanism type absent from it.

The four studies should therefore be read as one controlled decomposition of the causal problem. DIAL-SKFM-1A admits typed construction only with exact operators and a registered repair vocabulary. DIAL-SKFM-1B estimates both operators from finite noisy samples but abstains under its complete contract. DIAL-SKFM-1C shows that an actively chosen counter-witness is more informative than an equal-cost random probe, yet also abstains. Finally, DIAL-SKFM-1D represents unsupported alternatives explicitly and nearly eliminates unsafe accommodation, but remains too conservative to pass its overall gate. Thus the positive conclusion is not that CLIC has solved causal theory invention. It is that the successive burdens—typing, estimation, experiment selection, alternatives, and admission—can be measured separately rather than hidden inside graph accuracy.

### 11.2 *Exact closure and typed repair*

An exact mechanism study isolates the structural step before finite-sample estimation. DIAL-SKFM-1A represents infinitesimal mecha-

nisms by integer linear vector fields and computes their commutator directly. Projection onto the declared generator span separates assimilable coefficients from a closure residual. Within the registered synthetic semantics, an observed residual support may type a missing directed mechanism; a central residual supported on one latent coordinate may type a latent common-cause bundle. Residuals requiring two latent coordinates or a noncentral extension lie outside the registered one-generator repair budget and must elicit abstention.

In this finite study, projection outside the declared span together with the exhaustive registered repair table supplies the nonrealizability gate of Chapter 10. The gate is valid only for that exact finite language: it does not prove that an arbitrary causal representation lacks an in-theory explanation.

Across eighteen admission and eighteen confirmation worlds, the typed rule recovers every registered repair, restores closure, preserves the old generators, satisfies Jacobi, produces no false extension in closed worlds, and abstains on every unsupported case. The raw-residual control is the key negative result: adding the commutator always restores closure, but yields only 1/6 exact typed repairs and never abstains. Closure is therefore a diagnostic and integration condition, not by itself a causal interpretation.

#### Experiment: DIAL-SKFM-1A

**World:** exact integer linear vector fields in three observed and two latent coordinates.

**Obstruction:** commutator residual outside the declared generator span.

**Construction:** directed edge or one-latent central extension, typed from residual support; otherwise abstain.

**Result:** 18/18 admission and 18/18 confirmation repairs, with perfect closure, preservation, Jacobi, and unsupported-case abstention.

**Boundary:** exact fully observed operators; no trajectory estimator or unregistered repair type.

### 11.3 *Finite operators and propagated uncertainty*

The finite-operator study, DIAL-SKFM-1B, removes the exact-operator assumption and estimates both intervention operators independently from finite noisy state-velocity samples, estimates their row covariances, and propagates both sources of uncertainty through the commutator by parametric sampling. A frozen classifier then compares the uncertain residual with the registered repair templates and abstains when



neither fit nor posterior separation supports a typed extension. The reported record is the registered v2 execution; v1 stopped before any admission data were opened, and v2 made only a serialization-type correction, so the two records are not independent trials. Across 1,620 preregistered admission worlds, the operator intervals were close to nominal coverage: their empirical 95% coverage on nonzero entries was 93.6%. The uncertainty-aware rule covered 69.1% of supported worlds and was correct on 96.5% of the supported cases on which it acted, with no false extension in closed worlds. It nevertheless failed the complete admission contract: overall action accuracy was 71.7%, below the registered 85% threshold, and unsupported-case abstention was 81.9%, below the registered 90% threshold. No confirmation run was therefore opened.

This failure is diagnostic rather than merely aggregate. At observation noise 0.2, overall accuracy was 89.3% and unsupported abstention was perfect; at noise 1.2, they fell to 49.6% and 62.2%. With structural strength 1.0, accuracy reached 87.0% and unsupported abstention 94.4%; with strength 0.2, they were 55.7% and 65.6%. The difficult unsupported class was the noncentral extension, on which the method abstained 73.7% of the time, whereas it abstained on 90.0% of two-latent extensions. Thus uncertainty propagation successfully protects the precision of admitted repairs, but the registered residual language still aliases weak noncentral structure with closed or supported worlds.

#### Experiment: DIAL-SKFM-1B

**World:** two independently estimated linear intervention operators, with finite samples, observation noise, and supported and unsupported closure repairs.

**Uncertainty:** row-wise least-squares covariance propagated jointly through the commutator by 80 parametric draws.

**Result:** 93.6% operator-interval coverage, 96.5% selective accuracy, and zero false extensions, but only 71.7% overall accuracy and 81.9% unsupported abstention. The preregistered admission decision is ABSTAIN.

**Lesson:** calibrated operator uncertainty is necessary but not sufficient; a complete method must acquire evidence that separates weak noncentral extensions from the registered repair family.

### 11.4 Active counter-witness interventions

An active follow-up tests whether abstention can itself guide experiment selection. DIAL-SKFM-1C leaves every confident first-stage decision unchanged. After first-stage abstention, it selects one operator from

a registered library of off-diagonal matrix units. The chosen operator maximizes the minimum separation between the leading supported residual template and the two unsupported diagnostic alternatives. A fresh trajectory block estimates the resulting nested-bracket response, which is scored jointly with the original commutator. A random-probe control receives exactly the same additional sample budget.

This active counter-witness produces a substantive but incomplete repair. Across the 1,620 admission worlds, overall action accuracy rises from 72.2% without a counter-witness to 80.3%, and supported coverage rises from 68.2% to 83.9%, while selective accuracy remains 95.3% and the Ho false extension rate remains zero. Active choice also dominates random choice at the same 49.7% probe rate: random probes reach 75.9% accuracy, 80.0% supported coverage, and 76.3% unsupported abstention, compared with 80.3%, 83.9%, and 81.1% for active probes. The price is a rise in mean trajectory observations from 149.3 to 185.1 per world.

The study nevertheless receives a registered ABSTAIN. Overall accuracy remains below the 85% admission gate, and unsupported abstention remains below 90%. The reason reveals a design requirement. A probe is currently requested only after abstention, so it can rescue genuinely supported cases but cannot inspect a confident, incorrect supported decision made on an unsupported world. Moreover, 24 initially correct abstentions become false supported admissions after the extra noisy measurement. At sample size 128, the active rule reaches 89.8% accuracy and 90.0% unsupported abstention; at sample size 32 these fall to 71.3% and 72.8%. Active acquisition is therefore useful, but its trigger must depend on the risk of an unsupported alternative, not merely on whether the supported-family posterior has already abstained.

#### Experiment: DIAL-SKFM-1C

**Action:** after first-stage abstention, select one registered operator that maximally separates the leading supported repair from unsupported residual structures.

**Controls:** no counter-witness, random counter-witness at the same sample budget, and oracle.

**Result:** active acquisition improves accuracy from 72.2% to 80.3% and beats random acquisition by 4.4 points, but attains only 81.1% unsupported abstention. The preregistered decision is ABSTAIN.

**Open boundary:** request evidence when an unsupported alternative is plausible even if the supported-family classifier is confident, and require the counter-witness to confirm rather than merely replace an abstention.

### 11.5 *Explicit alternative posteriors*

DIAL-SKFM-1D makes the unsupported alternatives first-class inferential hypotheses without making them admissible repairs. Its posterior contains the four supported structures, the registered two-latent and noncentral alternatives, and an open “other unsupported” outcome. Unsupported posterior risk can trigger a probe even after an apparently confident supported decision. Admission is asymmetric: a previously abstained case needs positive joint confirmation before becoming a repair, whereas an initially admissible repair is retained unless the new evidence strongly contradicts it.

This change nearly eliminates false accommodation. Unsupported abstention is 99.8%, compared with 84.8% for the legacy posterior, while supported selective accuracy is 97.2% and the  $H_0$  false-extension rate remains zero. Active probing recovers substantial usefulness lost by the safer posterior: supported coverage rises from 41.5% with explicit alternatives alone to 65.1%, and overall accuracy rises from 60.7% to 75.4%. At the same 83.5% probe rate, random selection reaches only 58.3% coverage and 70.9% accuracy. The active selector therefore contributes information beyond merely gathering more samples.

The study still fails its overall-accuracy gate and does not open confirmation. Its principal error is no longer a false causal construction. Rather, the flat posterior treats the null closure hypothesis as just another template and rejects too many genuinely closed worlds: only 23.3% of  $H_0$  worlds are retained, although none is falsely extended. By contrast, noncentral worlds are rejected perfectly and two-latent worlds at 99.6%. This supports a hierarchical model: first perform a calibrated closure-versus-nonclosure decision, then conditionally infer a repair family. The null needs its own evidence-accumulation and stopping rule rather than symmetric competition with every nonzero structural template.

**Experiment: DIAL-SKFM-1D**

**Posterior:** four supported structures, two explicit unsupported alternatives, and an open unsupported outcome.

**Admission:** unsupported hypotheses can request evidence and veto a repair but can never themselves become repairs.

**Result:** 99.8% unsupported abstention, 97.2% supported selective accuracy, 65.1% supported coverage, and zero  $H_0$  false extensions; overall accuracy is 75.4%, so the registered decision is ABSTAIN.

**Lesson:** explicit alternatives nearly eliminate observed unsafe accommodation under this registered contract, but a flat posterior is excessively skeptical of genuine closure. Separate null detection from conditional structural diagnosis.

### 11.6 *From local repair to causal admission*

Here integration is not synonymous with choosing a graph. A local field or bracket explanation must extend across intervention targets, overlapping charts, and intervention sequences without acquiring path-dependent contradictions. Approximate holonomy, closure under newly selected probes, and held-out multi-intervention prediction are possible integration observers. Only after this finite compatibility test and the registered controls should the method propose  $Y_{CLIC}$ , with a causal presentation map to a sketch containing a new latent variable, mechanism family, or factorization when applicable. The changed package must then make consequences not used to generate it.

**Boundary**

Bracket nonclosure is a localization signal, not a certificate of hidden confounding. DIAL-SKFM may infer a latent variable only when the intervention contract distinguishes it from missing edges, misspecified mechanisms, and other admissible explanations.

Likewise, selecting one graph from a supplied candidate class is causal model selection, not automatically scientific creativity. The creative DIAL-SKFM challenge begins when the system may introduce a previously unrepresented variable, mechanism family, latent dimension, or factorization and state what new interventions would distinguish it. Such a proposal is transformational relative to the old causal sketch, but it earns scientific value only after interventional admission.

CLIC thus fixes the object of construction—a causal mechanism or factorization—while leaving open how a system acquires the reusable skills needed to propose and test it. Chapter 12 moves to that second problem, replacing causal variables by typed skill programs while

retaining the separation between local interaction, finite construction, and admission.

### *Further Reading*

Pearl’s structural account of causality supplies the standard distinction between observational association and intervention, as well as the model classes within which identifiability is normally posed [Pearl, 2009]. The immediate technical precursor to CLIC is BRIDGE/SKFM, which uses intervention-induced vector fields and Lie-bracket geometry to diagnose latent-confounded causal structure [Mahadevan, 2026a]. Infinitesimal Causality develops the broader tangent-space interpretation of causal mechanisms and interventions on which that construction rests [Mahadevan, 2026c].

For the categorical setting, the LINC framework treats failed composition as a typed structural diagnostic rather than a scalar loss [Mahadevan, 2026e]. Involution algebroids provide an abstract Lie-theoretic semantics inside tangent categories [Burke and MacAdam, 2019], while Mackenzie’s treatment of Lie groupoids and Lie algebroids gives the classical geometric background [Mackenzie, 2005]. The distinction made in this chapter between a local bracket repair and a globally admissible causal construction is closely related to the classical integrability problem for Lie brackets [Crainic and Fernandes, 2003]. These sources motivate CLIC’s geometry; its additional requirements—explicit unsupported alternatives, counter-witness interventions, uncertainty propagation, and conservative admission—are experimental safeguards against treating nonclosure alone as proof of a particular causal explanation.



## *Skill Optimization with Infinitesimal Creativity*

### OPTIC: OPTIMIZING PROGRAMS THROUGH INFINITESIMAL CREATIVITY

LASKO represents locally available skills as sections of a Lie algebroid and uses their anchors and brackets to organize skill optimization. DIAL-SkillOpt adds an external, versioned skill document that can be improved on training and validation traces before its behavior is tested on a locked admission set. Both constructions supply essential pieces, but neither name by itself identifies the full architectural target developed here.

*OPTIC—Optimizing Programs Through Infinitesimal Creativity*—asks when skill optimization should change not only the selected skill or its coefficients, but the language in which skills are expressed and composed. Its operational artifact is an inspectable program, often a typed Markdown workflow. Its intended infinitesimal semantics uses an involution algebroid in a tangent category. Its creative artifact is a newly constructed skill, interface, macro, decomposition, or workflow rule admitted by executions the optimizer was not allowed to edit—when the declared novelty and admission conditions are actually met.

The name marks an important separation. LASKO supplies the local geometry; DIAL-SkillOpt supplies a training procedure; OPTIC is the integrated system that diagnoses an inadequate skill theory and may extend it. The experiments reported below retain their registered DIAL-SkillOpt identifiers. Together they provide bounded mechanism evidence for OPTIC, not a demonstration of unrestricted skill invention.

#### 12.1 *The OPTIC target card*

Let  $S_{\text{skill}}$  be a typed workflow sketch. Its objects may represent skill types, inputs, outputs, tools, resources, preconditions, postconditions, and persistent records. Its arrows represent execution, sequencing, data transport, checking, and rollback. Distinguished diagrams declare which interfaces compose and which executions must agree.

An OPTIC target card is

$$\text{Target}_{\text{OPTIC}} = (P_m, S_{\text{skill}}, C_{\text{exec}}),$$

where  $P_m$  is the executable program compiled from a skill document  $m$ , and  $C_{\text{exec}}$  is a family of execution, safety, regression, cost, and transport certificates. The target is not merely a high-scoring prompt. A skill program must be typed, executable, persistent, and independently testable.

Operationally, OPTIC receives a versioned skill document, its compiler and authorization environment, typed execution traces, and a registered edit and experiment budget. It returns a replayable executable program, a typed workflow-extension proposal, a request for evidence, or an abstention. The return record includes the source, compiler, traces, regressions, and certificate status; an improved piece of prose is not by itself an output of the system.

For a fixed sketch, optimization may select skills, adjust continuous coefficients, reorder registered steps, or specialize a program to a local state. OPTIC calls this *assimilation*. A proto-accommodation proposal changes the registered skill package. Its control suite  $\text{Ctl}_{\text{OPTIC}}$  includes current and no-change programs, coefficient and ordering search, repairs using known workflow constructs, alternate factorizations, authorization and schema checks, and capacity-matched prompt or program-search controls. Only after these controls fail under the registered representability test may OPTIC propose a package comparison

$$Y_{\text{OPTIC}} : \mathfrak{R}_{\text{OPTIC}} \longrightarrow \mathfrak{R}_{\text{OPTIC}}^+.$$

When the workflow presentation changes, its presentation component is

$$J_{\text{skill}} : S_{\text{skill}} \longrightarrow S_{\text{skill}}^+.$$

It may introduce a skill object, interface, factorization, control-flow construct, or composition law. The presentation component does not transport semantic states by itself. A candidate must also supply or induce a semantic transport

$$\bar{J}_{\text{skill}} : M_{\text{skill}} \longrightarrow M_{\text{skill}}^+$$

and a compatible lift for old skills and their evidence. Changes to the compiler, authorization environment, observer, or admission procedure occupy their own components of  $Y_{\text{OPTIC}}$ . A proposal that destroys previously admitted behavior is not a successful creative extension.



**Design principle**

OPTIC optimizes an executable program and may propose a larger theory of executable programs when a registered representability test rules out the relevant fixed-workflow repairs. The larger theory becomes an accommodation only after finite realization and independent admission. Improvement inside a fixed workflow is not the same claim as extending the workflow language.

Its artifact dossier separates the compiled program  $P_m$  from the package comparison  $Y_{\text{OPTIC}}$ , its presentation and semantic transport components  $(J_{\text{skill}}, \bar{J}_{\text{skill}})$ , the regression and transport record for old skills, the locked execution certificates, and the persistent source, compiler, and trace hashes. The experiments in this chapter learn routing and composition over supplied subskills. They provide mechanism evidence for this dossier interface, but no experiment below independently admits a new skill type, interface, or grammar rule.

### 12.2 From vector-field brackets to involution algebroids

CLIC and OPTIC share an algebraic intuition but occupy different semantic levels. CLIC begins with intervention-induced vector fields in a smooth realization and estimates Lie brackets such as  $[v_i, v_j]$ . Their support helps localize a missing causal mechanism. OPTIC begins one level earlier, with an anchored differential bundle internal to a tangent category. Its primitive coherence is the involution on the appropriate prolongation; after passage to sections in a classical realization, that structure induces the familiar Lie bracket.

Let  $(\mathcal{M}, T)$  be a tangent category supporting the required differential bundles, and let

$$A_{\text{skill}} \longrightarrow M_{\text{skill}}$$

be an involution algebroid of locally available skill directions. As in Chapter 4, two partial compilers must remain distinct:

$$\begin{aligned} \text{compile}_E^{\text{sec}} &: \text{Ob}(\text{Md}_{\text{sec}}) \rightarrow \Gamma_{\text{loc}}(A_{\text{skill}}), \\ \text{compile}_E^{\text{pol}} &: \text{Ob}(\text{Md}_{\text{pol}}) \rightarrow \text{Prog}_E(\mathbb{S}_{\text{skill}}). \end{aligned}$$

Here  $\text{Prog}_E(\mathbb{S}_{\text{skill}})$  denotes the typed executable programs authorized by environment  $E$  for the current workflow sketch. The first turns a local skill artifact into a section  $s_m$ ; the second turns a workflow artifact into the executable target  $P_m$ . The anchor realizes  $s_m$  as an infinitesimal change of the maintained skill state. The involution coherence controls how infinitesimal skill operations are exchanged; in a smooth realization, its section-level shadow is  $[s_m, s_n]$ .

When the observer  $O$  takes values in a space with subtraction, a normalized finite paired-order experiment can estimate this shadow:

$$\hat{\Delta}_\epsilon(m, n; x) = \frac{1}{\epsilon^2} [O(\Phi_n^\epsilon \Phi_m^\epsilon(x)) - O(\Phi_m^\epsilon \Phi_n^\epsilon(x))].$$

Under the usual smooth-flow hypotheses, its small- $\epsilon$  limit is the derivative of  $O$  along the anchored section bracket, up to the order convention. In a general tangent-category realization, the paired routes require a comparison map rather than literal subtraction. A nonzero value may reveal order interference, a hidden shared resource, an incorrect interface, or an unavailable composite. It does not by itself say which explanation is correct. As in CLIC, the bracket is a localization signal rather than an admission certificate.

OPTIC adds the second DIAL direction. One side describes changes available inside the current skill theory; the other describes locally represented changes to its interfaces or workflow organization. When both sides admit involution-algebroid semantics, their mixed interaction belongs to the candidate double involution structure introduced in Chapter 3. We write  $\Theta_{\text{skill}}$  for its internal interchange defect and require a declared realization to relate it to paired executions or section brackets. A finite grammar edit remains a sketch map, not an infinitesimal vector merely because local probes helped propose it.

#### Boundary

The tangent-category theory of one involution algebroid is established. The complete internal double structure used by DIAL remains a research target. OPTIC experiments therefore report the concrete realization being tested; they do not treat an observed execution-order difference as a proof of an abstract double-involution theorem.

### 12.3 The OPTIC contract

In the generic DIAL-X contract,  $\mathbb{A}_{\text{OPTIC}}$  presents the locally available skill operations and  $\mathbb{B}_{\text{OPTIC}}$  presents typed interface or workflow revisions. Their realizations execute a local skill and propose a workflow change, respectively. Execution observers form  $\Lambda_{\text{OPTIC}}$ , their mixed comparison is  $\Theta_{\text{skill}}$ , finite reuse is tested by  $\gamma_{\text{skill}}$ , and locked execution evidence supplies  $\text{Adm}_{\text{OPTIC}}$ . The current experiments instantiate only registered fragments of this architectural contract.

An OPTIC implementation must expose the following components.

| Component                | Declaration                                                                                                                                    | Failure it prevents                                                 |
|--------------------------|------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------|
| Operational target       | Executable typed skill program $P_m$                                                                                                           | Optimizing prose that cannot be executed                            |
| Structural target        | Workflow sketch $S_{\text{skill}}$                                                                                                             | Hiding the skill vocabulary and interfaces                          |
| Creative mandate         | Frozen workflow and skill reference language, Boden mode, novelty observer, and independent execution value                                    | Calling every optimized program creative                            |
| Assimilatory side        | Selection and optimization of available sections                                                                                               | Calling every parameter update creative                             |
| Proto-accommodative side | Typed interface or workflow revision                                                                                                           | Treating arbitrary text edits as theory changes                     |
| Mixed witness            | Involution/interchange defect with an execution observer                                                                                       | Inferring structure from terminal score alone                       |
| Finite realization       | Multi-step execution, route consistency, rollback, and transport observer $\gamma_{\text{skill}}$                                              | Promoting a one-step edit that does not compose as a reusable skill |
| Proposal language        | Bounded skill, macro, factorization, or grammar extensions                                                                                     | Unconstrained post-hoc invention                                    |
| Controls                 | Registered in-language, ordering, factorization, authorization, no-change, and capacity-matched search controls in $\text{Ctl}_{\text{OPTIC}}$ | Extending the workflow before ordinary repairs are tested           |
| Package comparison       | $Y_{\text{OPTIC}}$ , with presentation and semantic components when applicable                                                                 | Hiding compiler, observer, or semantic changes inside a sketch map  |
| Admission                | Held-out execution, invariants, regressions, cost, and transfer                                                                                | Letting the optimizer certify itself                                |
| Persistence              | Versioned program, compiler, traces, rejected edits, and hashes                                                                                | Replacing learning by an unreplayable conversation                  |

The creative mandate must name what is absent from the frozen reference language. Constructing a new ordering of registered steps can be combinational creativity. Searching a typed space of registered workflows is exploratory creativity. Adding a new skill type, interface, or grammar rule is transformational. Retuning coefficients or selecting a supplied program is ordinary optimization unless a stronger novelty claim is separately established.

#### 12.4 An executable OPTIC episode

At time  $t$ , let

$$z_t = (S_t, m_t, x_t, H_t, b_t)$$

contain the current workflow sketch, skill document, semantic state, provenance-bearing trace, and remaining budget. The admissible actions include executing a registered section, requesting a paired-order probe, proposing a bounded program edit, constructing a typed sketch extension, requesting new evidence, and abstaining.

---

**OPTIC skill-construction episode**

- 1: Initialize the candidate-program set  $\mathcal{P} \leftarrow \emptyset$ .
  - 2: Type-check  $m_t$ , compile it relative to environment  $E$ , and load the current workflow sketch  $S_t$ .
  - 3: Execute registered skills and record outputs, resources, invariants, and failures in  $H_t$ .
  - 4: **if** the failure is ambiguous under marginal skill diagnostics **then**
  - 5:     select paired-order or active execution probes that separate the registered explanations.
  - 6:     **if** the additional evidence leaves the failure unresolved **then**
  - 7:         persist the alternatives and **return** an evidence request or auditable abstention.
  - 8:     **end if**
  - 9: **end if**
  - 10: Estimate the involution/interchange witness with uncertainty and localize its support to skills, interfaces, resources, or control flow.
  - 11: Execute the applicable controls in  $\text{Ctl}_{\text{OPTIC}}$  and retain their results in  $H_t$ .
  - 12: **if** a controlled repair exists inside  $S_t$  **then**
  - 13:     optimize or recompose the available skill sections and add each eligible program dossier to  $\mathcal{P}$ .
  - 14: **else if** the defect is persistent, the control suite has ruled out its specified alternatives, and registered representability tests find no repair in the current skill language **then**
  - 15:     construct  $Y_{\text{OPTIC}}$ , including  $J_{\text{skill}} : S_t \rightarrow S_t^+$  when the presentation changes, a versioned program  $m_t^+$ , transported evidence, and new predictions.
  - 16:     add each well-typed extension dossier to  $\mathcal{P}$ .
  - 17: **else**
  - 18:     persist the unresolved record and **return** an evidence request or auditable abstention.
  - 19: **end if**
  - 20: Test every candidate in  $\mathcal{P}$  with  $\gamma_{\text{skill}}$  across multi-step execution, relevant route orderings, rollback, and transport; defer any candidate without a coherent finite realization.
  - 21: Freeze the candidate before testing it on the independent admission suite; persist admitted and rejected versions separately.
-

Scalar scores may rank eligible programs, but eligibility is defined by the typed contract. A candidate with higher task reward but a failed invariant, malformed schema, or unregistered tool request is not an OPTIC improvement. Likewise, an extension that succeeds only on the traces that proposed it has not yet been admitted.

### 12.5 *Evidence for skill factorization*

Three empirical claims are carried by four archived protocols under the DIAL-SkillOpt name. The exact study is a held-out developmental calibration; the compound study is a sealed negative result; and the factored study is a separately registered diagnostic whose v1 and v2 records document one strict interface repair rather than two trials. Read together, they reveal why OPTIC must separate a skill program into typed obligations instead of asking one language-model edit to solve an entangled meta-learning problem.

#### 12.5.1 *Exact mixed-witness calibration*

The exact calibration used linear worlds arranged in matched triplets: assimilation was sufficient in one world, accommodation was required in a second world with identical marginal diagnostics, and a third world withheld the mixed probe. GPT-OSS 20B edited only the decision-policy portion of an external Markdown skill, subject to typed-schema and false-extension gates.

This study is a held-out developmental calibration rather than a fully preregistered confirmation. The world split and admission truth were frozen before model calls, but the optimizer-executor interface was repaired using training and validation behavior before the final skill was hashed and the admission set was opened. On six locked admission worlds, the initial skill and ordinary SkillOpt without the mixed witness each solved 4/6. DIAL-SkillOpt with the witness solved 6/6, localized all six correctly, handled both incomplete-probe controls, and produced no false extensions. The two additional successes were exactly the accommodation worlds that marginal evidence could not distinguish from their assimilation partners.

**Experiment: OPTIC-0: exact mixed witness****Artifact identifier:** DIAL-SkillOpt-0.**Comparison:** initial skill, SkillOpt without the mixed witness, and the complete mixed-witness skill.**Result:** 4/6, 4/6, and 6/6 exact locked decisions, respectively, with zero false extensions.**Boundary:** exact registered witnesses, fixed extension vocabulary, a deliberately small admission set, and pre-admission development of the executor interface.*12.5.2 The compound-skill failure*

The preregistered compound-skill study introduced noisy finite evidence, active probe selection under cost, thresholded hypothesis admission, and selection among three typed extensions. One bounded Markdown edit was required to coordinate all three operations. Neither proposed edit passed every validation gate while strictly improving the registered selection vector. The procedure terminated as specified, froze no skill, and never opened the locked admission set.

This is a useful negative result. It does not show that the model lacked every required capability. It shows that a monolithic external skill and one scalar selection event obscured which obligation failed. OPTIC therefore treats factorization as part of the skill theory, not merely as prompt engineering.

**Experiment: OPTIC-1: compound-skill termination****Artifact identifier:** DIAL-SkillOpt-1.**Task:** jointly learn active probing, noisy admission, and typed extension construction through one bounded edit.**Result:** no admissible improving skill; the confirmatory run terminated before admission.**Epistemic status:** registered negative result; no skill was frozen and the locked admission worlds remained unopened. The result localizes failure to the compound-skill contract rather than establishing model incapacity.**Lesson:** failure localization requires persistent subskill traces and separate type and validation contracts.*12.5.3 Factored subskills*

The diagnostic repair represented active probe selection, noisy hypothesis admission, and typed extension rendering as three separately exercised skills under registered validators. Only the probe-selection

policy required an edit: it learned to choose the largest registered proposal-separation score. The admission and rendering policies were already reliable once their inputs and output types were isolated.

The reported result is the separately registered v2 diagnostic. Version 1 stopped at a string-versus-number schema failure; v2 clarified the numeric JSON contract without relaxing the validator, and the records are not pooled as repeated trials.

On fifteen fresh admission worlds, the initial subskills achieved 10/15 end-to-end accuracy and made 3/15 false extensions. The frozen factored pipeline achieved 15/15 end-to-end accuracy and no false extensions. Probe-label imitation itself was only 10/15, yet all selected probes were decision-sufficient. This distinction matters: the correct target is an equivalence class of sufficient, cost-aware experiments, not necessarily one registered argmax label.

#### Experiment: OPTIC-1F: factored skill composition

**Artifact identifier:** DIAL-SkillOpt-1D v2.

**Factorization:** probe selection, uncertainty-qualified admission, and typed extension rendering.

**Result:** end-to-end accuracy improved from 10/15 to 15/15, and false extensions fell from 3/15 to zero.

**Boundary:** the hypotheses, probes, and renderer vocabulary remained registered; this supports robust composition under noise, not autonomous invention.

### 12.6 *What has been learned—and what has not*

These studies support three mechanism claims within the registered synthetic worlds. First, a mixed witness can contain decision-relevant information absent from marginal skill diagnostics. Second, independent admission gates can prevent an optimizer from promoting a malformed or unsupported skill edit. Third, factorizing a compound skill into typed subskills can convert an opaque optimization failure into a learnable and auditable composition.

They do not establish the central creative claim of OPTIC. Every retained experimental artifact used a registered hypothesis and skill vocabulary. The model learned a routing rule and composed supplied subskills; it did not construct a new interface or workflow grammar absent from the reference space. These are evidence for auditable skill factorization and routing, but remain preparatory to the stronger claim of infinitesimal creativity.

**Boundary**

OPTIC begins where fixed-vocabulary skill optimization ends. A system has not demonstrated transformational skill creativity merely because it edits a Markdown file or reaches a higher score. It must construct a previously unavailable typed operation or composition rule, transport prior behavior, and pass independent execution tests.

*12.7 The open construction problem*

The unresolved test is construction rather than routing alone.

1. **Withheld macro.** Supply primitive skills but withhold a rollback-aware compare–repair–confirm macro required by some worlds. Test whether OPTIC constructs the macro after persistent bracket interference.
2. **Withheld interface.** Give two individually valid skills whose types do not compose. Permit a bounded mediator object and require the system to infer its interface from execution counter-witnesses.
3. **Order and sufficiency.** Compare active bracket-based probes with random paired executions at equal cost, scoring equivalence classes of decision-sufficient probes rather than exact labels.
4. **Transport.** Freeze the constructed skill theory and test it on disjoint task families, tool versions, and workflow graphs. Measure retained old competence separately from new competence.
5. **Baselines.** Compare ordinary SkillOpt, fixed-geometry LASKO, flat prompt search, OPTIC without the mixed witness, OPTIC without explicit factorization, and the complete system under matched execution budgets.

The decisive outcome is not terminal reward alone. Reports should include construction accuracy, type validity, false extension, regression, execution cost, abstention, transfer, and persistence. A successful new macro should remain available as a versioned artifact whose future uses and failures can be audited.

*12.8 CLIC, OPTIC, and RELIC*

The three principal algorithmic realizations form a progression.



|       | <b>Maintained structure</b>                              | <b>Mixed witness</b>                            | <b>Creative construction</b>                      |
|-------|----------------------------------------------------------|-------------------------------------------------|---------------------------------------------------|
| CLIC  | Causal sketch and intervention geometry                  | Brackets of realized intervention fields        | Mechanism, latent object, or causal factorization |
| OPTIC | Skill program, workflow sketch, and involution algebroid | Internal interchange and paired execution order | Skill, interface, macro, or workflow rule         |
| RELIC | Decision sketch, schemas, and policy                     | Failure of decision operations to compose       | State factor, option, schema, or decision grammar |

CLIC asks what causal structure should be constructed. OPTIC asks what executable skill language can construct and test such structures. RELIC asks how an embodied agent should retain and deploy structural accommodations over time. Their shared DIAL discipline is diagnosis, typed construction, independent admission, and persistent transport; their objects and evidence remain domain-specific.

Chapter 13 completes this progression by making the temporal problem explicit. It moves from a versioned skill artifact to an agent whose state representation, options, and decision grammar may themselves need revision while behavior continues.

### *Further Reading*

LASKO develops skill optimization over Lie algebroids and supplies the classical geometric foundation for OPTIC [Mahadevan, 2026d]. Burke and MacAdam define involution algebroids in tangent categories, while MacAdam develops their functorial semantics [Burke and MacAdam, 2019, MacAdam, 2023]. Mackenzie’s treatment of Lie groupoids and Lie algebroids provides the classical background [Mackenzie, 2005]. SkillOpt supplies the complementary engineering pattern in which an agent improves an external reusable skill document under evaluation [Yang et al., 2026].



## Reinforcement Learning with Infinitesimal Creativity

### RELIC: FROM POLICY OPTIMIZATION TO STRUCTURAL ACCOMMODATION

In a standard reinforcement-learning formulation, an agent improves a policy inside a declared decision problem. The state variables, action types, transition interface, reward representation, and temporal grammar are fixed by that formulation before policy learning begins. Exploration changes the evidence available to the agent; optimization changes its value function or policy. Neither operation, by itself, changes what a policy is allowed to be.

*RELIC—Reinforcement Learning with Infinitesimal Creativity*—asks a different question. What should an agent do when its failure is evidence that the current decision language is inadequate? The answer developed in this chapter is not to identify creativity with a larger exploration bonus. The agent must detect a typed obstruction, localize the missing composition, propose a structural repair, test it on evidence not used to propose it, and retain it only after admission.

The experiments below provide the sequential-decision bridge from the DIAL-X architecture of Chapter 10 to an embodied learning system, complementing the causal-learning and skill-construction bridges developed by CLIC and OPTIC in Chapters 11 and 12. They form a progression rather than one headline comparison. Generic novelty changes learning without producing a task theory. Isolated goal-relevant actions also fail to compose. A temporally extended pickup-and-place schema, however, can be triggered by an observed obstruction, admitted on disjoint games, and transported to untouched houses. The result is a controlled mechanism demonstration for structural accommodation. It is not yet a claim that the agent autonomously invented the schema.

#### 13.1 What is creative about reinforcement learning?

Let a conventional decision declaration be

$$S_{RL} = (\mathcal{X}, \mathcal{A}, P, R, \gamma_{disc}, \mathcal{O}),$$

*Chapter roadmap.* RELIC first separates policy learning from declaration change and fixes its target card. The ALFWorld ladder then tests persistent artifacts, novelty controls, supplied temporal composition, constructed composition, and finally construction of a missing stage.

where  $\mathcal{X}$  and  $\mathcal{A}$  describe the state and action types,  $P$  is a transition law,  $R$  is a reward interface,  $\gamma_{\text{disc}}$  is a discount convention, and  $\mathcal{O}$  records the observer available to the agent. A policy  $\pi$  and critic  $Q$  are models inside this declaration. Ordinary learning moves the pair

$$(Q, \pi) \mapsto (Q', \pi')$$

while keeping  $S_{\text{RL}}$  fixed.

That is assimilation in the sense used throughout this book. A proto-accommodation proposal changes the declaration or the language in which its policies are assembled:

$$S_{\text{RL}} \xrightarrow{I} S_{\text{RL}}^+.$$

Examples include adding a latent regime coordinate, forming a temporally extended option, separating one action type into two, introducing an object relation, or changing which observations are retained as persistent state. The creative delta is the newly expressible decision structure, not merely a higher-return parameter vector. Installation as accommodation occurs only after finite realization and independent admission.

#### Design principle

Intrinsic motivation changes which experiences an agent seeks. RELIC changes which policy structures the agent may construct, but admits such a change only after an independent test of value.

This distinction does not diminish exploration. Novelty, surprise, and information gain may be excellent selectors of experiments. Their semantic role is nevertheless different from a declaration extension. A count bonus can lead an agent toward an unfamiliar cupboard without teaching it that “find, acquire, transport, place” should be retained as one reusable composition.

### 13.2 The RELIC target card

RELIC specializes the DIAL-X contract to sequential decisions. Its three targets are

$$\text{Target}_{\text{RELIC}} = (Y_{\text{dec}}, S_{\text{dec}}, C_{\text{dec}}).$$

The operational target  $Y_{\text{dec}}$  is a deployable policy together with any persistent state or macro-actions it requires. The structural target  $S_{\text{dec}}$  declares the types, factorization, temporal interfaces, and observable obligations of that policy. The epistemic target  $C_{\text{dec}}$  combines task completion, validity, counter-witnesses, transport, and conservative abstention.

The RELIC creative mandate freezes that decision sketch, its primitive action and observation interfaces, stage grammar, and reference library of admitted policies and schemas. A new composition of registered stages is combinational; a new reachable policy inside the same package is exploratory; and a change to the registered decision package is transformational. Return and novelty remain separate observers.

Operationally, RELIC receives interaction traces, the maintained decision sketch, and the currently registered observation and primitive-action interfaces. It returns a deployable policy together with any persistent state, schema predicate, macro-action, or proposed declaration change that the policy requires. A finite-rollout record and a status-bearing admission record accompany that output; when the evidence cannot distinguish structural inadequacy from ordinary policy error, the correct output is abstention or an evidence request.

In the notation of Chapter 10,  $\mathbb{A}_{\text{RELIC}}$  presents critic, policy, and action updates inside the maintained decision language, while  $\mathbb{B}_{\text{RELIC}}$  presents probes and revisions of state, schema, or temporal structure. Their realizations perform within-declaration learning and propose decision-language change, respectively. The observer family records reward, validity, memory, and factorization evidence;  $\Theta_{\text{RELIC}}$  compares the two orders;  $\gamma_{\text{dec}}$  tests finite execution; and  $\text{Adm}_{\text{RELIC}}$  assigns status on disjoint games or traces.

The finite-realization observer  $\gamma_{\text{dec}}$  is distinct from that final certificate family. It asks whether a local policy repair remains coherent over a finite rollout, recurrence, or temporally extended execution, including the state and memory transport required by the candidate decision structure. A one-step reward gain that fails this observer is not eligible for admission.

For the present studies,  $\text{Ctl}_{\text{RELIC}}$  registers current and no-change policies, stronger fixed-declaration learning, generic intrinsic and goal-relevant exploration, isolated structural actions, saved-state restoration, reordered and type-invalid programs, and observation or memory alternatives. Only after the applicable controls fail under the stated contract may RELIC propose a package comparison

$$Y_{\text{RELIC}} : \mathfrak{R}_{\text{RELIC}} \longrightarrow \mathfrak{R}_{\text{RELIC}}^+.$$

As with skill construction, its presentation component  $J$  does not itself move policy state or memory. A RELIC proposal must also record a supplied or induced semantic transport  $\bar{J}_{\text{dec}} : M_{\text{dec}} \rightarrow M_{\text{dec}}^+$ , together with the policy-state and memory lift over it.

The RELIC artifact dossier consequently contains more than the deployed policy. It records the constructed temporal composition or task-stage operator, its package comparison  $Y_{\text{RELIC}}$ , the state and memory transported into the new policy language, disjoint-game admission

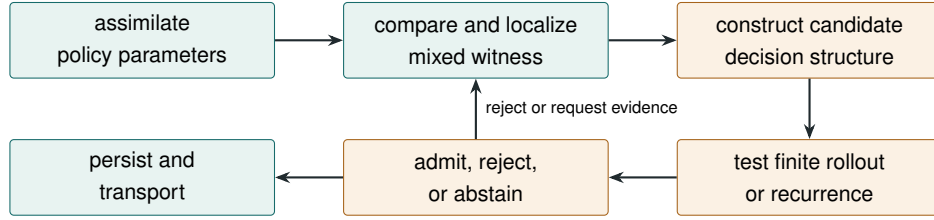


Figure 13.1: The RELIC loop. Creativity is not identified with the obstruction or the proposal. A candidate must first survive a registered finite rollout or recurrence test, and then an independent admission gate, before it becomes persistent policy structure.

and confirmation results, and a replayable version of the resulting program. The strongest experiments below admit such constructions up to behavioral equivalence inside registered primitive actions and stage contracts; they do not establish invention of a new primitive action or observation type.

For the experiments in this chapter, the horizontal direction is learning or acting inside the current policy language. The vertical, proto-accommodative direction probes or proposes a change to the language: a state coordinate, schema predicate, or temporally extended composition. Only a finite proposal that passes realization and admission is installed as accommodation. Their mixed question is operational:

learn and act in the old declaration, then extend it  
 $\iff?$  extend the declaration, then learn and act.

A repeated action, Bellman inconsistency, non-Markov residual, or failed factorization can make the disagreement observable. None of these witnesses automatically licenses accommodation. It localizes pressure on the current model and triggers a proposal, finite-realization, and admission process.

### 13.3 From GIRL to RELIC

GIRL provides the tangent layer of the construction. Bellman and policy relations are lifted to spaces of admissible variation, allowing the learner to distinguish directions that ordinary scalar optimization may conflate. DIAL-GIRL adds a second level: the maintained decision representation can be transported, restored, forked, or enlarged when the evidence warrants it.

The early synthetic experiments progressively removed oracle structure. In switching and partially observed worlds, active probes distinguished fixed parameter drift from a missing regime coordinate. Finite trajectories replaced supplied certificates; prototype families and latent filters were learned; sequential evidence controlled when additional history was justified. One important negative result shaped RELIC: exact restoration of a remembered policy was sometimes worse than continual learning because experience in the intervening regime

transferred positively. Memory supplies a candidate, not a rollback command. Independent admission must choose among saved, current, and transported states.

Within their registered settings, these experiments support three design requirements for creative decision making:

1. the failure must be observable in finite interaction data;
2. the proposed change must have a declared type and support; and
3. numerical improvement cannot replace structural admission.

They do not by themselves establish creativity. Selecting a registered latent coordinate or restoring a stored state remains adaptation inside a supplied candidate family. RELIC begins to approach combinational creativity when the candidate is a newly assembled policy composition, and transformational creativity when the available stage or state language itself can expand.

#### 13.4 *Schema learning in an embodied text world*

The schema experiments move from abstract switching processes to ALFWorld, an environment that aligns textual interaction with embodied household tasks [Shridhar et al., 2021]. An agent may need to find an object, acquire it, operate on it, transport it, and place it in a receptacle. Observations are partial, object locations persist across time, and action validity depends on relations such as containment, co-location, carrying capacity, receptacle state, and tool compatibility.

This setting reconnects RELIC with the Piaget–Drescher lineage. A schema is not merely a label attached to a trajectory. It predicts which action is applicable, what transition should follow, and which persistent relations must be updated [Drescher, 1991]. Failure creates a counter-witness to the current schema; repeated evidence can justify accommodation.

The observation-only studies recovered increasingly relational structure:

- action-family success conditions were learned from textual outcomes;
- source, destination, tool, and receptacle types were separated;
- object containment was represented as a changing location relation;
- acquisition required both object–source and agent–source co-location;
- carrying capacity and persistent tool location were added; and
- frozen schemas were replayed on disjoint games before being used by a planner.

This sequence matters because a strong frozen policy could otherwise conceal the origin of its structure. Every retained predicate has an evidence record and a counter-witness history. Still, the final artifact and the process constructing it must be evaluated separately.

### 13.5 *The ALFWorld benchmark contract*

The benchmark separates two questions. The *competence* question asks whether a policy equipped with structural state solves standard tasks and transports to unseen houses. The *creativity* question asks whether the system detects a withheld obligation, proposes a repair, admits it on disjoint evidence, and reuses it without sacrificing earlier structure.

Three channels must remain distinct:

*Policy-visible.* The natural-language goal, current and past observations, executed actions, observation-derived outcomes, and admitted schemas with their provenance.

*Evaluator-only.* PDDL facts, expert plans, game identity during action selection, and oracle success state before scoring.

*Declared oracle control.* The official ALFWorld TextDQN receives the environment’s admissible-command list. It is therefore an oracle-action-space learned baseline, not a deployable counterpart to a policy that generates its own commands.

This qualification is central. Admissibility removes one source of failure but does not supply planning: a random admissible policy solved none of the 274 validation games. Conversely, comparing a generated-action RELIC policy with an admissible-command DQN without labeling the channel difference would overstate the empirical conclusion.

#### Admission contract

RELIC evidence is partitioned into discovery, admission, confirmation, and transport. A game used to construct or rank a repair cannot certify that repair. Confirmation occurs only after the admitted structure and its thresholds have been frozen.

### 13.6 *Competence of persistent structural artifacts*

The A-series experiments evaluate frozen competence before adding on-line repair. Ao ran proposal-only, frozen-RELIC, and oracle-admissible policies on all 140 seen and 134 unseen TextWorld validation games. On unseen games, the proposal policy solved 0.201, frozen RELIC solved



| Stage | Question                                  | Evidence                                              | Base result              | RELIC result                           | Inference                                                         |
|-------|-------------------------------------------|-------------------------------------------------------|--------------------------|----------------------------------------|-------------------------------------------------------------------|
| A0    | Does a frozen artifact help?              | 140 seen, 134 unseen                                  | 0.201 unseen             | 0.910 unseen                           | Strong competence and transport.                                  |
| A1    | Is A0 a random or tie-seed effect?        | All games; five OOD seeds                             | 0 random; 0.212 proposal | 0.918 mean OOD                         | Validity alone is insufficient; the gain is stable.               |
| B1    | Does generic novelty create structure?    | 24 training episodes; paired evaluation               | 0 success                | 0 success, positive shaped reward      | Learning changed, but no task theory emerged.                     |
| B2    | Do isolated structural proposals suffice? | Checkpoints at 12, 24, 36 episodes                    | 0 success                | 0 success                              | One-step guidance did not compose.                                |
| B3    | Can a macro-schema be admitted?           | 8 admission; 16 confirmation games                    | 0 throughout             | 0.875 admission; 0.625 seen and unseen | A temporal composition can be admitted and transported.           |
| B4    | Can the composition be constructed?       | 24 orders on 6 proposal; 8 admission; 16 confirmation | 0 DQN and shuffled       | 0.750 admission; 0.625 seen and unseen | The useful order is recovered, frozen, admitted, and transported. |

Table 13.1: The RELIC-ALFWorld evidence sequence. A-series results test the value of persistent artifacts; B-series results distinguish shaped learning, isolated exploration, supplied structural accommodation, and combinational construction.

0.910, and the oracle control solved 0.933. Relative to the matched proposal policy, frozen RELIC gained 0.709 success, reduced mean invalid-action rate from 0.708 to 0.083, and reduced median episode length from the 50-step ceiling to 21.

A1 tested simpler explanations. Uniform random generation and uniform random admissible choice both solved zero of 274 games. Across five unseen-house tie seeds, proposal-only success averaged 0.212 (standard deviation 0.034), while frozen RELIC averaged 0.918 (standard deviation 0.012). The structural gain was therefore not a favorable tie-breaking accident.

These are strong competence results, but competence is not the same as creativity. The schemas were already frozen when A0 and A1 began. They establish that structural artifacts can be useful, stable, and transportable; they do not establish how those artifacts were constructed.

### 13.7 Why novelty and one-step guidance are insufficient

B1 compares the official TextDQN under extrinsic reward alone with the same learner receiving ALFWorld’s episodic-count and novel-object rewards. Both arms begin from byte-identical weights and share the same seed, replay budget, action channel, and evaluation games. Neither receives extrinsic training reward or solves an evaluation game. Yet the manipulation is active: the intrinsic arm receives dense shaped return, selects a different repeated action at evaluation, and ends far

from the extrinsic-only checkpoint.

The failure is therefore not “nothing learned.” What was learned did not compose into goal-directed task structure. Novelty values an unfamiliar state or object; it does not declare why that encounter matters to a pickup-and-place factorization.

B2 sharpens the comparison. At episodes 12, 24, and 36 it evaluates three TextDQN arms: ordinary exploration with an epsilon-greedy rule; count/object novelty; and typed goal-relevant exploration. The third arm changes only exploratory proposals, favoring actions such as take, put, open, and task-matched navigation. None of the three solves a training or evaluation game. The structural arm briefly learns a more directed navigation and examination pattern at episode 12, then falls into the same repeated-action attractors as the other arms.

All checkpoints move substantially and diverge from one another. B2 therefore rules out an inactive optimizer and rules in a more specific diagnosis: isolated useful-looking actions do not constitute a temporally extended policy. The missing object is a composition.

### 13.8 *Admitting a supplied temporal composition (B3)*

B3 freezes the episode-36 structural TextDQN from B2 and registers the candidate

LocateTarget  $\rightarrow$  AcquireTarget  $\rightarrow$  LocateDestination  $\rightarrow$  PlaceTarget.

The schema consumes only the natural-language task, observations, action history, and the same admissible-command channel available to the matched DQN. It activates only after the base policy repeats the same action three times. This repetition is the registered obstruction: a local policy loop that makes no progress toward the declared task.

The evidence is divided before execution. Four discovery games document that the frozen DQN repeatedly exhibits the obstruction. Eight disjoint seen games compare the unchanged DQN and candidate schema for admission. Admission requires at least 0.75 candidate success, at least 0.50 absolute gain over the DQN, at least 0.75 trigger coverage, and zero invalid actions. Only an admitted schema may proceed to eight untouched seen and eight untouched unseen games.

The frozen DQN solves none of the discovery, admission, or confirmation games. The candidate activates on all eight admission games, generates no invalid action, and solves seven. It therefore passes every admission gate. Once frozen, it solves five of eight untouched seen games and five of eight unseen games, while the DQN remains at zero in both splits. The Wilson 95-percent interval is  $[0.529, 0.978]$  for admission and  $[0.306, 0.863]$  for each confirmation split.

This is the separately registered 30-step horizon-recovery record. An earlier 15-step execution was invalidated before confirmation and is not counted as an additional trial.

**Experiment: RELIC–ALFWorld-B3**

**Target and obstruction:** complete simple object placement; three consecutive identical actions without completion.

**Candidate:** a four-stage typed pickup-and-place composition.

**Evidence:** four discovery games; eight admission games; eight seen and eight unseen confirmation games.

**Gates:** locked success, gain, trigger, and validity thresholds.

**Result:** 7/8 admission; 5/8 seen and 5/8 unseen confirmation; zero invalid candidate actions.

**Epistemic status:** admitted supplied policy mechanism; experimenter-proposed schema.

The negative and positive rungs belong together. Had B3 followed only A0, its gain could be dismissed as a hand-coded planner. B1 and B2 show why the macro is structurally meaningful: dense novelty and isolated task-relevant actions both change learning, but neither supplies the missing composition. B3 adds exactly that object and subjects it to independent admission.

### 13.9 A DIAL reading of the result

The first involution is assimilatory: update the critic–policy state and act through the current action interface. The second is a policy-language intervention: install a candidate temporal composition and transport the task and observation evidence into it. The B2 failure is a behavioral proxy for an unresolved mixed obstruction. Goal-relevant local motions exist, but their learned order does not close into a successful task process. B2 does not by itself estimate the interchange comparison between learn–then–extend and extend–then–learn.

B3 repairs this interaction by adding a mediating temporal object. The four stages are not four unrelated heuristics; their interfaces constrain the next legal stage. Acquisition changes the persistent inventory state, which changes the target of navigation, which enables placement. Admission tests whether this composite is effective rather than merely compatible with the action types.

This is best classified as *proto-combinational creativity*. Relative to the B2 policy language, a new useful composite becomes available. But the experimenter supplied both its generators and their ordering. RELIC detected the obstruction, conditionally activated the candidate, admitted it, and transported it. It did not search over possible compositions

or invent a new stage type. The stronger creativity claim remains open.

#### Boundary

B3 records admission of a supplied structural repair under the registered contract. It does not show that RELIC autonomously invented pickup-and-place, and its sixteen confirmation games are a mechanism sample rather than a benchmark-scale performance claim.

### 13.10 *Constructing the temporal composition (B4)*

B4 removes precisely one piece of B3's scaffolding. The four typed stage operators remain registered, but their useful order is withheld. All  $4! = 24$  permutations are evaluated on six proposal games. The ranking rule uses success, finite stage progress, validity, and episode length; it has no access to the evaluator's canonical order. The selected presentation map and its enclosing package comparison are written and frozen before eight disjoint admission games are opened.

The search recovered

LocateTarget  $\rightarrow$  AcquireTarget  $\rightarrow$  LocateDestination  $\rightarrow$  PlaceTarget.

It solved all six proposal games with full four-stage progress. On admission it solved six of eight games, compared with zero for both the frozen DQN and a registered shuffled-order control. It generated no invalid action, achieved mean finite progress 3.0, and passed all five admission gates. Once frozen, it solved five of eight untouched seen games and five of eight untouched unseen games. The evaluator-only canonical control produced the same results.

The type-invalid control is especially informative. It solved five of eight admission games and four of eight games in each confirmation split, but incurred 42, 84, and 142 precondition violations respectively. Task success alone would therefore have credited a malformed composition. The typed integration observer rejects that conclusion.

**Experiment: RELIC–ALFWorld-B4**

**Repair theory:** four registered typed stage operators; ordering withheld.

**Proposal:** rank all 24 permutations on six proposal games and freeze one order before admission.

**Controls:** frozen DQN, evaluator-only canonical order, shuffled order, and type-invalid repeated-stage composition.

**Integration:** finite stage progress, precondition coherence, validity, completion, and transport.

**Result:** 6/8 admission; 5/8 seen and 5/8 unseen confirmation; zero precondition violations for the selected order.

**Epistemic status:** admitted combinational construction inside a fixed stage vocabulary.

B4 is the first RELIC study that constructs rather than merely receives the successful composition. Its creativity claim is consequently stronger but still bounded. The system discovers an ordering of known operators; it does not invent a new operator, state predicate, or option interface. This is combinational creativity relative to the frozen vocabulary, not yet transformational creativity.

### 13.11 *Residual failure and declaration scope*

The admitted schema does not solve every confirmation game. The three unseen failures are all salt-shaker tasks; the three seen failures require finding a soap bottle, box, or newspaper. In each case the bottleneck is long target search rather than invalid manipulation. The accommodation has exposed a new obstruction one level down: `LocateTarget` is too weakly structured.

This is how a persistent theory should grow. The failure does not motivate loosening the admission threshold after the fact. It motivates a new declared candidate family: receptacle priors learned from observations, an information-seeking search option, or a relational object-source model with its own counter-witnesses. The existing macro remains admitted where its contract holds; the localized stage becomes the target of a subsequent theory revision.

### 13.12 *Constructing a missing stage*

The three studies in this section progressively refine the observer rather than merely enlarge the search. B5 constructs programs from registered action fragments but cannot separate its selected ordering from a re-ordered control. B5.1 actively acquires disagreement states, yet its raw

syntax space contains programs that remain behaviorally indistinguishable. B5.2 quotients that space by observed action traces and ranks distinct classes with typed coherence before reward and efficiency. The first two studies abstain; the third admits a constructed stage up to the resulting empirical equivalence relation.

Stage construction removes a stronger piece of scaffolding. B4 ordered known stages; B5 deliberately removes `LocateTarget` from the registered task-stage vocabulary. No permutation of the remaining operators can satisfy the acquisition precondition coherently. Because that operator inventory is finite and exhaustively typed, this impossibility is the registered nonrealizability gate for the rung; empirical search failure alone would not have been sufficient. The proposal mechanism instead receives a finite grammar of lower-level action fragments and must construct a new operator of type

`SearchState[target not actionable] → SearchState[target actionable]`.

Its observable codomain witness is an admissible `take` command naming the target. The grammar contains opening an unvisited receptacle, navigating to an unvisited location, navigating away from the destination, and navigating to the destination. Twenty-four three-fragment priority programs are evaluated on five proposal games. The selected program is frozen before eight disjoint admission games are opened.

The proposal phase selected

`Priority(openUnvisited, goUnvisitedAny, goNonDestination),`

with 5/5 task success and 5/5 contract witnesses. On admission it solved 6/8 games, compared with 4/8 for the missing-stage control and 0/8 for the frozen DQN. It satisfied its codomain witness on 7/8 games and incurred neither invalid actions nor precondition violations. A control with an incorrect stop predicate solved 4/8 but accumulated 107 precondition violations. These comparisons show why task reward alone cannot certify a declaration extension.

B5 nevertheless abstained. Reversing the selected fragment order still solved 5/8 games, leaving a gain of only 1/8, below the preregistered 2/8 margin. The untouched confirmation partition was therefore not opened. The result supports the usefulness of adding a typed target-search stage, but does not identify the internal priority structure sharply enough for admission. Several programs tied on proposal success and contract satisfaction; the available worlds did not provide counter-witnesses that separated them.

**Experiment: RELIC–ALFWorld-B5**

**Expressive obstruction:** LocateTarget withheld from the task-stage vocabulary.

**Construction language:** 24 priority programs built from four registered navigation and inspection fragments.

**Frozen candidate:** openUnvisited, then goUnvisitedAny, then goNonDestination.

**Result:** 6/8 admission, zero type violations, versus 4/8 with the stage missing; only 5/8 versus a reordered-program control.

**Epistemic status:** abstained—useful typed stage construction, but insufficient identification of its internal composition.

The relative nature of the creativity claim matters. The B5 proposal is transformational relative to the task-stage vocabulary because it adds a new typed operator. It is combinational relative to the registered fragment grammar because it does not invent a new primitive or its contract. Because admission is withheld, however, the study does not establish a transformational result. This nested reference frame is more informative than labeling the episode simply creative or noncreative.

B5.1 therefore changes the experimental observer rather than enlarging the search. It actively chooses games or interventions that elicit edge-level counter-witnesses between candidate programs, then freeze and retest the surviving extension. Accommodation proceeds only after the system can distinguish programs that happen to agree on coarse task success.

B5.1 implements this observer. A deterministic scout visits a pool of 24 training worlds and, at each state, asks what action each of the 24 candidate programs would recommend. It never observes reward, terminal success, an expert plan, or the evaluator-only canonical program. The acquisition score counts pairwise action disagreement, states with disagreement, and the number of distinct recommendations. Eight maximally discriminating worlds are then selected under a frozen budget; a fixed eight-world sample supplies the passive control.

Active and passive acquisition produce different candidate programs inside the registered fragment grammar. The active sample selects

Priority(openUnvisited, goNonDestination, goUnvisitedAny),

which happens to equal the sequestered canonical program. The passive sample instead places goNonDestination before openUnvisited. On twelve disjoint admission games, the active program solves 12/12, satisfies its registered contract on 12/12, incurs zero type violations, and requires 4.17 mean search steps. The passive program solves 11/12, satisfies its registered contract on 11/12, incurs 14 precondition violations,

and requires 6.25 mean search steps. The active program wins their registered paired utility comparison by two games.

B5.1 nevertheless abstains for two distinct reasons. First, the missing-stage control stumbles to task completion on 11/12 games, despite satisfying no search contract and accumulating 128 precondition violations. Its scalar success is therefore only 1/12 below the selected program, short of the registered 3/12 margin. Second, the syntactic runner-up differs only in a third fallback that is never behaviorally decisive on the admission worlds. It is observationally equivalent to the selected program under all registered utilities, so the paired separation gate fails. Confirmation again remains sealed.

#### Experiment: RELIC–ALFWorld-B5.1

**Acquisition:** reward-blind disagreement scout over 24 worlds; eight actively selected versus eight passive worlds.

**Active proposal:** recovered the evaluator-only canonical fragment program without access to it.

**Admission:** 12/12 success, 12/12 contracts, zero violations, and 4.17 mean search steps; passive proposal 11/12, 14 violations, and 6.25 steps.

**Decision:** abstain—the malformed missing-stage baseline retained high scalar success, while the closest syntactic rival was behaviorally equivalent.

**Lesson:** candidate theories must be quotiented by observable behavior, and typed coherence cannot be subordinated to terminal reward.

The two localized gate failures determine the B5.2 protocol. It ranks equivalence classes of programs rather than raw syntax trees and compares the selected class with the best observably distinct rival. Its primary utility is lexicographic: contract and precondition coherence first, task reward second, and efficiency third. This protocol is registered as a new declaration rather than used to reinterpret B5.1 after observing its outcome.

Fresh reward-blind acquisition over 24 further training worlds selects eight counter-witness worlds. The 24 candidate syntax trees induce only eight observable behavior classes when quotiented by their constructed-stage action traces. The selected class contains six programs and is represented by

Priority(goNonDestination, openUnvisited, goUnvisitedAny).

Its strongest observably distinct rival begins instead with unrestricted navigation. A second member of the selected class is frozen to test whether proposal equivalence transports rather than being assumed.



B5.2 passes every admission gate. The selected representative solves 10/12 games, satisfies its registered contract on 12/12, incurs no invalid action or precondition violation, and requires 7.92 mean search steps. It wins the lexicographic paired comparison with the strongest distinct class on seven games and with the missing-stage control on all twelve. The second member of the selected class has no utility disagreement with the representative. By contrast, the missing-stage and type-invalid controls solve 7/12 but satisfy no contract and each accumulate 204 precondition violations.

Admission opens the sealed unseen confirmation partition. Both representatives of the selected class solve 8/8 games, satisfy their registered contracts on 8/8 games, incur zero violations, and require 11.75 mean search steps. The distinct rival also solves 8/8 coherently but requires 15.0 steps; the evaluator-only oracle requires 13.88. The missing-stage and type-invalid controls fall to 1/8 while each accumulates 201 violations. Thus the quotient preserves coherent behavior and an efficiency advantage under distribution shift, without claiming a unique syntax for the constructed operator.

#### Experiment: RELIC-ALFWorld-B5.2

**Quotient:** 24 syntax trees collapse to eight observed behavior classes; selected class size six.

**Admission:** 10/12 success, 12/12 contracts, zero violations; seven paired wins over the strongest distinct class and twelve over the missing stage.

**Confirmation:** 8/8 unseen success and contracts, zero violations; 11.75 mean search steps versus 15.0 for the distinct rival.

**Epistemic status:** admitted construction of a new typed task-stage operator up to an empirical behavioral equivalence class.

**Boundary:** primitives, the stage contract, and the admissible-command channel remain registered scaffolding.

Stronger tests must remove three kinds of scaffolding: invent or learn a new primitive or stage contract rather than only composing registered fragments; extend the state representation when the fragment grammar is insufficient; and learn the admissible-command model from interaction while preserving generated-action and validity controls. The resulting frozen learner must then be compared with capacity- and interaction-matched pretrained, fine-tuned, planning, and schema-learning baselines at benchmark scale.

This progression gives an operational empirical meaning to “learning by double involution.” Policy learning supplies local improvement. Structural accommodation supplies new decision objects. Creativity becomes defensible only when the system can construct such an object

relative to a frozen reference language and when independent evidence admits it.

The claim remains realization-specific. These experiments implement typed side operations, behavioral proxies for mixed witnesses, finite-rollout observers, and admission gates; they do not prove that the ALFWorld controller is an internal double involution algebroid in an arbitrary tangent category.

CLIC, OPTIC, and RELIC isolate three different creative objects: causal structure, executable skill, and persistent decision structure. The next chapter asks whether their certified outputs can be exchanged and composed without erasing the distinctions among their evidence, uncertainty, and admission conditions.

### *Further Reading*

The Piagetian distinction between assimilation and accommodation motivates the developmental interpretation used here; Drescher gives its classic computational schema realization [Piaget, 1985, Drescher, 1991]. For the ALFWorld environment and its alignment between textual and embodied interaction, see Shridhar et al. [2021]. Standard reinforcement learning is treated systematically by Sutton and Barto [2018]; temporally extended actions by Sutton et al. [1999]; deep value learning by Mnih et al. [2015]; and curiosity-driven exploration by Pathak et al. [2017]. The distinction between novelty and surprise is especially relevant to the B1 control [Barto et al., 2013]. The DIAL-GIRL and schema experiments in this chapter are reported as mechanism studies: readers should distinguish their structural claims from state-of-the-art ALFWorld comparisons based on substantially larger training budgets or pretrained language agents.

## Composing Creative Workflows

### AGENTIC: ACQUIRING GENERALIZABLE NOVEL THEORIES WITH INFINITESIMAL CREATIVITY

CLIC, OPTIC, and RELIC each maintain a different theory and may propose revisions to it. CLIC maintains a theory of how a world responds to intervention. OPTIC maintains a theory of executable skills and their composition. RELIC maintains a theory of which decisions are available and useful over time. A scientific agent needs all three, but merely placing their implementations in one software process does not yet produce an integrated creativity system. Their claims, types, admission records, and revision histories must compose.

This chapter specifies that composition. Its primary object is a *meta-sketch*: a versioned declaration linking a causal theory, a skill theory, and a decision theory. The connections themselves may fail. A causal mechanism may be well supported but not actionable; a skill may execute but not improve decisions; a policy failure may be localized only after the agent performs a new intervention. Such cross-workflow failures provide a new class of obstruction and, potentially, a new source of creativity.

We call the integrated architecture *AGENTIC: Acquiring Generalizable Novel Theories with Infinitesimal Creativity*. “Acquiring” emphasizes active experiment and evidence rather than passive generation; “generalizable” requires transport beyond the discovery setting; “novel theories” identifies structural construction as the target; and “infinitesimal creativity” names the DIAL discipline used to diagnose and coordinate local repair. The expanded name states the architectural target, not the empirical conclusion of this chapter. CLIC, OPTIC, and RELIC are its causal, skill, and decision workflows.

These names denote architectural mandates, not a retrospective claim that every component experiment has already achieved creativity. The evidence in Chapters 11–13 ranges from mechanism calibration and guarded adaptation to bounded combinational construction. AGENTIC must retain those component-level statuses rather than promote them

merely because the workflows have been connected.

### 14.1 Three maintained theories

Let the current integrated state be

$$\mathfrak{M} = (S_C, S_O, S_R; \lambda_{CO}, \lambda_{OR}, \lambda_{RC}; \mathcal{L}, \mathcal{A}).$$

Here  $S_C$ ,  $S_O$ , and  $S_R$  are the current CLIC, OPTIC, and RELIC sketches. The maps  $\lambda$  are typed interfaces, not identifications of the three theories. In compact form, they carry

$$\begin{aligned} \lambda_{CO} &: \text{causal packet} \longrightarrow \text{skill requirement}, \\ \lambda_{OR} &: \text{skill packet} \longrightarrow \text{decision operation}. \end{aligned}$$

and  $\lambda_{RC}$  carries a decision-failure packet or information need to a causal experiment request. The store  $\mathcal{L}$  records versions, provenance, executions, and counter-witnesses. The admission family  $\mathcal{A}$  assigns domain-specific epistemic status. The notation  $O$  in OPTIC is used here to keep it distinct from the causal component  $C$  and the reinforcement-learning component  $R$ .

The integrated target is therefore

$$\text{Target}_{\text{meta}} = (Y_{COR}, S_{COR}, \mathcal{C}_{COR}), \quad Y_{COR} := (Y_C, Y_O, Y_R),$$

where the deployable objects  $Y_C, Y_O, Y_R$  remain separate, while  $S_{COR}$  presents their legal connections. The certificate family  $\mathcal{C}_{COR}$  produces both component admission records and tests of their interfaces. No component may silently promote another component's proposal to an admitted fact.

Operationally, AGENTIC consumes status-bearing CLIC, OPTIC, and RELIC packets, not untyped summaries of their hidden states. It returns a workflow choice, new packet requests, or a versioned proposal for changing the meta-sketch, while preserving the status of every component claim. The interface may coordinate evidence acquisition, construction, and action; it may not replace their distinct admission records with one undifferentiated creativity score.

The maintained representational package is correspondingly broader than the three sketches:

$$\mathfrak{R}_{AG} = (S_C, S_O, S_R; \lambda_{CO}, \lambda_{OR}, \lambda_{RC}; \text{Obs}_{AG}, \text{Adm}_{AG}, \text{Gram}_{AG}).$$

The observer and admission components type interface evidence, while  $\text{Gram}_{AG}$  specifies the registered packet fields, workflow constructors, guards, and feedback laws. The AGENTIC creative mandate freezes this package, the component reference libraries, and the admissible control

*Chapter roadmap.* The chapter defines the three maintained theories and their exchange packets, then studies workflow algebra, cross-workflow obstructions, integrated episodes, ablations, active evidence, and the stronger requirement for genuine meta-accommodation.

suite before an episode. Constructing a new legal graph inside  $\text{Gram}_{\text{AG}}$  is combinational or exploratory; changing a registered packet field, interface, constructor, observer, or admission rule requires a proposed package comparison

$$Y_{\text{AG}} : \mathfrak{R}_{\text{AG}} \longrightarrow \mathfrak{R}_{\text{AG}}^+.$$

Its presentation component may include  $J_{\text{COR}} : \mathcal{S}_{\text{COR}} \rightarrow \mathcal{S}_{\text{COR}}^+$ , but that map is not the entire meta-level change.

The integrated artifact dossier retains all three operational outputs and their component dossiers, plus the package comparison  $Y_{\text{AG}}$ , transport and control records, and the interface admission records that justify their composition. In the experiments reported here, the constructed object is a response-conditioned workflow policy absent from the baseline declaration but assembled inside a registered packet and operation grammar. A genuinely new workflow constructor would be a stronger transformational artifact and is not established by those results.

#### Design principle

Integrate creativity systems by composing their declared interfaces and status-bearing admission records, not by concatenating their prompts or hidden states.

### 14.2 The exchange object

Each workflow communicates through a *structural packet*

$$P = (\tau, \mathfrak{R}, Y, \omega, U, \gamma, h, q, c, s, \kappa, v),$$

consisting of a packet type, prior representational package, proposed package comparison, localized obstruction, uncertainty object, finite-realization record, registered-control record, requested consequence or query, status-bearing admission record, epistemic status, interface contract, and version.

The comparison  $Y$  is the identity when no package change is claimed. When its presentation changes, the corresponding sketch map  $J$  is stored as one typed component rather than substituted for the whole comparison. The uncertainty object  $U$  may be a posterior, confidence region, equivalence class, or explicit set of unsupported alternatives; its interpretation is declared by  $\kappa$ . The record  $\gamma$  states which persistence or integration observer the candidate passed; it is not a substitute for the independent evidence summarized by  $c$ . The status distinguishes at least *proposed*, *admitted*, *rejected*, and *unsupported*. A packet may ask a downstream workflow to construct an experiment or skill without asserting that its motivating explanation is true.

This distinction prevents three common category errors. First, a CLIC nonclosure residual is not automatically a causal mechanism. Second, an OPTIC program that type-checks is not automatically useful. Third, a RELIC reward increase is not automatically evidence for the causal story that proposed an action. The interface transports a question and its provenance; it does not transport certainty for free.

Accordingly, an interface is generally a partial, status-preserving transformation

$$\lambda_{ij} : P_i \rightarrow P_j \sqcup \{\text{abstain}\},$$

defined only when the source packet satisfies its registered contract. It must state how uncertainty is transformed under the change of semantics. Copying a numerical confidence value from a causal packet into a skill packet does not calibrate that value for executability. When such transport is not justified, the target packet retains the alternatives and marks the derived claim as proposed or unsupported.

Here *status preserving* does not mean that the evidential statuses form a numerical ordering. The source packet and its status remain immutable in provenance; the target workflow assigns a new status under its own certificate family. An admitted causal packet may therefore yield only a proposed skill request, and an admitted skill packet may yield an unsupported decision claim. Interface admission certifies the transport, not the source claim under a new name.

Packets also make revision auditable. If RELIC later finds that a skill is useful only in one regime, the system can trace that skill to the causal proposal and intervention evidence that produced it. Retraction can then be localized rather than erasing the entire integrated state.

### 14.3 A small algebra of workflow composition

The first implementation needs only a finite orchestration sketch with four constructors.

*Sequential composition*  $B \circ A$ . The output packet type of  $A$  matches the input contract of  $B$ , and the composite retains both component admission records and a separate admission record for the interface between them.

*Parallel composition*  $A \otimes B$ . Two workflows examine the same versioned state under a declared policy for merging compatible proposals and preserving incompatible ones as alternatives.

*Guarded choice*  $A \triangleleft_g B$ . A registered observer  $g$  selects a workflow without hiding the evidence on which routing depended.

*Feedback  $A^\circ$ .* An admitted or rejected downstream result returns to an upstream workflow as new evidence. The loop has a stopping rule, budget, and version discipline.

These constructors describe workflow syntax. They do not yet prove that creative workflows form a PROP, operad, double category, or another stronger categorical structure. Such a claim requires explicit objects, morphisms, equivalences, and coherence laws. The finite sketch is deliberately weaker: it lets experiments reveal which equations and universal constructions are actually stable.

Component admission does not compose automatically. Even if two packets are individually admitted, their sequential composite is admitted only when an interface observer verifies type compatibility, uncertainty transport, version alignment, and the claimed joint consequence on independent evidence. Overall workflow admission is therefore a separate judgment, not the conjunction of component statuses.

#### 14.4 Three modes of meta-creativity

The workflow level reproduces Boden's three modes with a precise reference space.

*Combinational.* The component workflows and their interfaces are fixed, but the system constructs a new legal composite such as  $R \rightarrow C \rightarrow O \rightarrow R$ .

*Exploratory.* A grammar of legal workflow graphs is fixed, and the system searches their orderings, guards, probe allocations, and stopping rules under a budget.

*Transformational.* The current orchestration sketch cannot express a needed exchange. The system proposes a new packet field, mediator object, interface, workflow, or composition law and independently tests the enlarged meta-sketch.

Thus a newly constructed, independently admitted composition of CLIC, OPTIC, and RELIC may qualify as combinational creativity *over workflow operators*. Merely wiring together a fixed pipeline is orchestration. The claim becomes transformational only when the system changes the language in which those workflows can be connected. This criterion prevents a large agent graph from being labeled creative merely because it contains several specialized agents.

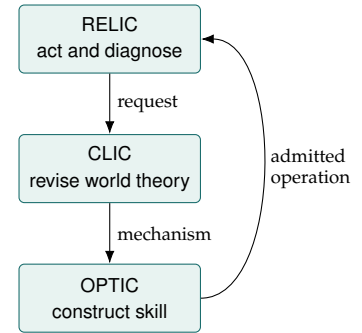


Figure 14.1: A failure-first creative loop. Every arrow carries a typed structural packet; no proposal is admitted merely by traversing the cycle.

### 14.5 *Cross-workflow obstructions*

An integrated system exposes failures invisible to each component alone. Representative obstruction types include:

1. **Actionability defect.** CLIC admits a model-relative causal distinction, or an equivalence class of such distinctions, for which OPTIC has no executable intervention or measurement skill.
2. **Utility defect.** OPTIC admits a reliable skill, but RELIC cannot find a decision context in which its use improves the declared objective.
3. **Explanatory defect.** RELIC discovers a reproducible policy failure that no alternative in the current causal sketch explains.
4. **Realizability defect.** RELIC requests an intervention that the skill theory deems physically, ethically, or operationally inadmissible.
5. **Transport defect.** A causal repair, skill, and policy compose in the discovery environment but fail under a registered change of regime.
6. **Order defect.** Two individually useful workflow calls yield incompatible results when their order is reversed.

Some order defects may admit a DIAL interpretation through paired infinitesimal directions. Others arise from finite data acquisition, irreversible action, or conflicting certificates and should not be forced into Lie-bracket language. The meta-sketch records which semantics applies.

### 14.6 *Canonical compositions*

Different problems call for different registered workflow graphs.

The system should be able to abstain not only on a domain proposal but also on workflow selection. When two orderings are observationally indistinguishable, the appropriate output may be an equivalence class of workflow graphs or a request for a discriminating meta-experiment.

### 14.7 *An integrated episode*

The following protocol instantiates a failure-first loop without presuming that it is optimal.

1. **Freeze the reference state.** Version the three sketches, their interfaces, candidate workflow grammar, budgets, and admission thresholds.



| Composition                                   | Use                                                                               | Characteristic risk                                       |
|-----------------------------------------------|-----------------------------------------------------------------------------------|-----------------------------------------------------------|
| $C \rightarrow O \rightarrow R$               | Begin from an admitted mechanism, construct an operation, then deploy it          | Causal relevance need not imply decision value            |
| $R \rightarrow C \rightarrow O \rightarrow R$ | Let persistent policy failure request causal experiments and a new skill          | Reward failure may be mislocalized as a causal defect     |
| $O \rightarrow R \rightarrow C$               | Explore with newly composed skills, then explain the resulting regime differences | Tool-induced artifacts may masquerade as world mechanisms |
| $C \otimes R \rightarrow O$                   | Jointly use mechanistic and decision evidence to specify a skill                  | Conflicting evidence may be prematurely collapsed         |
| $(C \rightarrow O \rightarrow R)^\odot$       | Iterative scientific action and theory revision                                   | Feedback can amplify an early false accommodation         |

Table 14.1: Registered compositions of the three creativity workflows. The ordering is an empirical choice, not a universal law.

2. **Detect a persistent decision obstruction.** RELIC distinguishes a structural failure from ordinary policy-estimation error and emits a packet requesting information.
3. **Design a counter-witness.** CLIC ranks interventions that separate supported causal explanations while retaining an explicit unsupported alternative.
4. **Construct an executable probe.** OPTIC composes or proposes the skill needed to realize the intervention and checks its type, safety, and execution semantics against the registered contract.
5. **Acquire and propagate evidence.** RELIC or the environment executes a probe admitted for its registered execution contract; CLIC updates the causal posterior without treating execution success as causal confirmation.
6. **Propose structural changes.** CLIC may extend the causal sketch; OPTIC may extend the skill sketch; RELIC may extend the decision sketch. Each extension receives its own status.
7. **Test finite and joint compatibility.** Apply each component's registered integration observer and the applicable controls in  $\text{Ctl}_{AG}$ , then evaluate held-out intervention prediction, skill reliability, decision value, regression, transport, and the interfaces joining them.
8. **Persist or retract.** Admit only the components and connections supported by their certificates. Preserve unresolved alternatives and the complete audit trail.

This procedure permits partial success. For example, a causal extension may be admitted even when no useful policy has yet been constructed. Conversely, a robust skill may be retained without accepting the causal interpretation that first motivated it.

### 14.8 *An integrated testbed*

The cleanest initial world is a partially observed, nonstationary tool-use environment with three withheld structures:

1. a latent regime that changes one causal interaction;
2. a tool macro or mediator interface required to probe or exploit that interaction; and
3. a decision schema that must condition action on the newly recoverable regime.

Ordinary policy optimization can improve within a regime but cannot express the missing state distinction. CLIC can diagnose the causal alternative but cannot execute the required compound probe. OPTIC can construct the probe but does not know when it has decision value. RELIC can detect persistent failure but cannot identify the missing mechanism. In the exact AGENTIC-oA ablation, the complete system succeeded on every challenge world, whereas the best pairwise composition reached at most one half. This is a result about that registered finite construction, not a claim that all three components are necessary in every environment.

A controlled progression begins with a finite exact world, then adds operator estimation, observation noise, imperfect execution, switching regimes, and finally language-mediated tool use. At every stage, discovery, admission, and confirmation seeds remain disjoint. A flat end-to-end learner receives the same interaction and computation budget.

### 14.9 *Ablations and interaction effects*

The base ablation is the full  $2^3$  factorial over CLIC, OPTIC, and RELIC, augmented by controls for the interfaces themselves:

1. no structural creativity and a matched flat learner;
2. each component alone;
3. the three pairwise compositions;
4. all three components with a fixed workflow order;
5. all three with learned guarded routing;
6. the full system without cross-workflow certificate checks;
7. oracle structure, oracle skills, and oracle regime information as separate ceilings.

These alternatives constitute the registered meta-level control suite  $\text{Ctl}_{\text{AG}}$ . It includes the no-change and fixed-order workflows, each component and pairwise composition, a capacity- and budget-matched flat learner, interface-only and certificate-free variants, matched-random acquisition, and the separate oracle ceilings. A nonzero cross-workflow obstruction may route evidence or select among registered graphs, but it licenses  $Y_{\text{AG}}$  only after the applicable controls fail and a registered representability test shows that no graph in the current workflow grammar realizes the target. Otherwise the correct output is an in-grammar construction, an evidence request, or abstention. Let  $V(S)$  be preregistered value for the system containing component set  $S \subseteq \{C, O, R\}$ . A descriptive three-way interaction is

$$\begin{aligned} \Delta_{\text{COR}} = & V(\text{COR}) - V(\text{CO}) - V(\text{CR}) - V(\text{OR}) \\ & + V(\text{C}) + V(\text{O}) + V(\text{R}) - V(\emptyset). \end{aligned}$$

A positive  $\Delta_{\text{COR}}$  suggests complementarity not explained by lower order combinations; a negative value reveals interference. It is not by itself evidence of creativity. The structural record must still show what was constructed beyond each frozen reference space.

Reports should separate task reward from causal coverage, skill construction and execution validity, decision adaptation, false accommodation, abstention, sample and compute cost, catastrophic regression, transport, and persistence. Order sensitivity should be reported directly rather than averaged away.

**Experiment: META-DIAL-0**

**Epistemic status:** proposed factorial benchmark; the controlled demonstration reported next tests only a restricted precursor of this specification.

**World:** finite partially observed tool-use systems with a withheld causal regime, executable probe macro, and decision schema.

**Question:** can CLIC, OPTIC, and RELIC construct and admit a joint extension that none can complete alone?

**Controls:** full factorial component ablation, fixed versus guarded workflow order, interface-certificate ablation, flat learner, and three oracle ceilings.

**Primary outcomes:** joint task success and held-out transport, subject to zero or preregistered-low false accommodation.

**Structural outcomes:** causal-extension recovery, new-skill validity, new decision-schema validity, interface compatibility, and three-way interaction  $\Delta_{COR}$ .

**Admission boundary:** improved reward without a versioned construction and independent component and interface certificates is optimization, not an admitted creative episode.

#### 14.10 *A controlled demonstration: active evidence across the workflow*

The first computational realization of the integrated workflow uses a small family of finite, partially observed worlds. A latent regime changes which of two actions is useful. The regime becomes observable only after a compound probe is executed, and the agent's original decision language does not contain the response-conditioned policy needed to exploit the result. CLIC must support the latent distinction, OPTIC must establish the compound probe's registered execution contract, and RELIC must construct and admit a policy that maps the probe's two responses to different actions. The exact AGENTIC-oA ablation supplies the component-necessity result described above.

The evidence sequence replaces exact packets one at a time with finite-sample estimators; its separately registered rungs are not pooled. AGENTIC-oA verifies exact composition. AGENTIC-oB estimates CLIC and falls short; AGENTIC-oB.1 repairs the boundary cases through active acquisition. AGENTIC-oC localizes three OPTIC failures, and AGENTIC-oC.1 repairs them without relaxing the gate. AGENTIC-oD estimates RELIC. Because AGENTIC-oD.1 contains no eligible downstream boundary case, it does not test active acquisition; the fresh hierarchical AGENTIC-oD.2 split supplies that missing test.

In the final controlled-demonstration stage, all three component

packets depend on empirical trajectories. CLIC actively gathers counter-witness evidence; an independent OPTIC check certifies the proposed probe's registered execution contract; and RELIC estimates four reward cells—two probe responses crossed with two candidate actions. RELIC admits the new policy only when the preferred action in each response context has a Wilson interval separated from its alternative and the two preferences form a genuinely response-conditioned schema. Packet parsing and final execution remain exact. The workflow graph, packet schema, candidate probe, action vocabulary, and response-conditioned-policy form also remain registered. Thus the constructed object is a guarded decision extension inside a supplied meta-workflow; the experiment does not construct or revise the meta-workflow language itself.

For this experiment, overall workflow admission means that the CLIC packet passes its registered causal-evidence gate, the OPTIC packet passes its execution-contract gate, RELIC admits a differentiated response-conditioned policy, and both interfaces pass version, type, uncertainty, and provenance checks. This is a deliberately conjunctive finite certificate for the registered testbed; it is not a general theorem that arbitrary component certificates compose.

The RELIC test uses a preregistered mixture of standard worlds and *boundary worlds*. In the latter, the correct and incorrect actions have reward probabilities 0.64 and 0.36, rather than 0.70 and 0.30. Every response-action cell initially receives eighty trials. Active RELIC allocates batches of twenty additional trials to both actions of an unresolved response context, but only when the current point estimates already form a differentiated candidate policy. It is compared with a fixed budget, a uniform high budget of 120 trials per cell, and a matched-random control that permutes the active method's realized extra budgets across worlds and response contexts.

Because an upstream abstention prevents RELIC from being evaluated, the primary endpoint is RELIC admission conditional on valid CLIC and OPTIC packets. Overall workflow admission is retained as a secondary endpoint. This hierarchical evaluation prevents a CLIC or OPTIC miss from being misreported as evidence against downstream acquisition.

Active acquisition repaired eight of nine eligible fixed-budget failures in the admission split and ten of eleven in confirmation. Matched-random allocation repaired none and one, respectively. The active method also slightly exceeded uniform acquisition while using approximately 30% fewer decision trajectories. Uncertainty-gated RELIC admitted no policy under shuffled rewards and no differentiated policy in forced-null worlds; the corresponding point-estimate control admitted 25.0% and 18.8% of shuffled cases. Conditional interval coverage

| Split        | Acquisition    | RELIC given upstream | Boundary RELIC | Trials per challenge |
|--------------|----------------|----------------------|----------------|----------------------|
| Admission    | Fixed (80)     | 0.812                | 0.625          | 320.0                |
|              | Matched random | 0.812                | 0.625          | 329.2                |
|              | Uniform (120)  | 0.938                | 0.875          | 480.0                |
|              | Active RELIC   | <b>0.979</b>         | <b>0.958</b>   | <b>329.2</b>         |
| Confirmation | Fixed (80)     | 0.771                | 0.542          | 320.0                |
|              | Matched random | 0.792                | 0.583          | 334.2                |
|              | Uniform (120)  | 0.958                | 0.917          | 480.0                |
|              | Active RELIC   | <b>0.979</b>         | <b>0.958</b>   | <b>334.2</b>         |

was 98.9% in both prospective splits.

The two remaining abstentions illustrate the guard rather than an exhausted budget. In admission, one unresolved response had empirical margin 0.075, below the registered 0.08 boundary required to justify further sampling. In confirmation, the provisional policy assigned the same action to both responses and therefore did not constitute a differentiated extension. Active RELIC declined to spend additional evidence in both cases.

#### Boundary

This experiment is a controlled mechanism demonstration, not evidence of general scientific creativity. The worlds are synthetic; the probe and action vocabularies are registered; the proposed extension has a known two-response, two-action form; reward noise is Bernoulli; and packet parsing and deployment are exact. The bounded construction claim tested here is the guarded admission of a response-conditioned policy absent from the baseline declaration. It does not include invention of new actions, sensors, objects, or an unrestricted theory language.

Evidence beyond this mechanism demonstration would have to relax these conveniences separately: hide or generate the candidate vocabulary, enlarge response and action spaces, introduce nonstationary and partial observation, learn packet interfaces, or move to language-mediated and physical environments. Within the registered synthetic workflow, the present result establishes only that the composed architecture can localize an evidential obstruction, acquire data at that obstruction, and retain conservative abstention when its declared preconditions are absent.

Table 14.2: AGENTIC controlled-demonstration results. Active RELIC targets unresolved response-action comparisons and is evaluated conditionally on valid upstream packets. The matched-random control receives the same realized additional sample budget. Admission and confirmation use disjoint registered surface families and random seeds.

### 14.11 *What surprising outcomes would mean*

Composition creates possibilities that single-workflow experiments cannot expose. Several outcomes would be scientifically informative.

*Synergy.* A causal distinction becomes discoverable only after OPTIC constructs a probe that RELIC learns when to deploy.

*Antagonism.* A confident causal proposal narrows OPTIC's search too early and performs worse than an uncertainty-preserving composition.

*Order dependence.*  $C \rightarrow O \rightarrow R$  and  $R \rightarrow C \rightarrow O \rightarrow R$  produce different admitted theories under equal budgets.

*Creative bottleneck migration.* After one missing structure is supplied, the limiting obstruction moves from causal language to skill language or from skill language to decision language.

*Meta-accommodation.* No registered ordering succeeds, but adding a mediator or new feedback edge produces a transportable improvement.

Negative results are equally valuable. If the full system never exceeds the best pairwise composition, the third workflow may be redundant in that testbed. If removing certificate checks improves reward but increases false accommodation, the experiment quantifies the price of epistemic discipline. If a flat learner matches performance and transport, the structural machinery has not yet earned its complexity.

### 14.12 *From orchestration sketch to workflow theory*

The initial meta-sketch is an empirical scaffold. Repeated experiments may support stronger mathematics. Stable sequential and parallel laws would motivate a symmetric monoidal or PROP-like presentation. Multi-input creative operations with substitution may suggest an operadic account. Stateful processes and explicit interfaces may favor double categories or structured cospans. None should be selected for elegance alone.

The transformational endpoint is a system that can propose a conservative package comparison

$$Y_{AG} : \mathfrak{R}_{AG} \longrightarrow \mathfrak{R}_{AG}^+$$

with  $J_{COR} : S_{COR} \rightarrow S_{COR}^+$  when the presentation changes, transport old component histories, state what the new connection permits, and design an independent test. That is creativity applied recursively to the organization of creativity itself.

**Boundary**

A meta-controller that selects among three fixed API calls demonstrates routing, not transformational creativity. Meta-creativity requires a declared workflow reference space, an extension absent from that space, a localized reason for proposing it, and independent evidence that the extension composes without corrupting previously admitted structure.

Part III stress-tests the shared compositional discipline in five non-equivalent evidence regimes: exact mathematical certificates, model-relative simulators, heterogeneous scientific corpora, structural observers of visual assurance, and open-ended admission of reusable artistic operations. The last two share a visual medium but license different claims. Their order reflects decreasing direct executability, not a ranking of epistemic value. The testbeds instantiate different subsets and orderings of the AGENTIC workflow; they do not all claim a fully learned three-component integration. What unifies them is the requirement to preserve typed proposals, provenance, independent admission, and conservative abstention as evidence moves between components.

The progression is deliberate. Chapter 15 begins where admission can be exact but the worlds are controlled; Chapter 16 adds executable mechanisms and interventions; Chapter 17 removes the privileged executable world and reconstructs testable proposals from heterogeneous documents. Chapters 18 and 19 then make the distinction between repair and theory extension visible: the first enforces a declared visual language, while the second asks whether a reusable operation missing from that language can be constructed and admitted.

*Further Reading*

Boden supplies the three modes of creativity [Boden, 2004]. Presented theories follow Lawvere and Ehresmann [Lawvere, 1963, Ehresmann, 1968]; Lack discusses PROPs as a possible, but not claimed, composition language [Lack, 2004]. LINCOS supplies typed obstruction, repair, and admission [Mahadevan, 2026e]; Chapters 11–13 supply the component contracts composed here.



## **Part III**

# **AGENTIC Testbeds and Open Domains**



## Mathematical Theory Construction

### AGENTIC IN EXACT AND GENERATED WORLDS

Part II ended by composing causal diagnosis, executable skill construction, and sequential experiment choice into the AGENTIC meta-sketch. Part III changes the unit of comparison. Its subject is no longer the component workflow but the *admission environment*: what kind of evidence can admit a constructed artifact or a proposed package comparison in mathematics, simulation, scientific literature, or visual generation?

These environments are not ranked versions of one ground truth. Exact proof, executable simulation, source-bearing testimony, and perceptual observation expose different objects and different failure modes. Provenance, uncertainty, counter-witnesses, and abstention must therefore be calibrated anew in each chapter rather than inherited unchanged from the preceding testbed.

Each chapter states which AGENTIC packet fields are registered, estimated, constructed, or independently admitted; a missing component remains visibly missing rather than being supplied by the change of domain. In particular, the common packet schema organizes comparisons across testbeds, but it does not make their evidence interchangeable or imply that every testbed realizes the complete CLIC–OPTIC–RELIC loop.

Throughout Part III, the verbs *reports*, *supports*, *establishes*, and *proves* retain the meanings fixed by the evidence contracts in Section 10.4.2. A stronger reading requires a separately named admission environment.

Table 15 fixes those environments before the individual studies begin. Its last column is as important as its middle one: an internally decisive result may still leave the external scientific or creative question open.

Mathematics supplies the cleanest setting in which a proposed theory can be checked without confusing fluency for correctness. It also supplies historical examples in which a new concept generated a durable field.

For AGENTIC, mathematics is the sharpest admission environment

| Regime                    | Internal reference                                                                       | Strongest internal verdict                                                       | Not established internally                                                   |
|---------------------------|------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------|------------------------------------------------------------------------------|
| Generated mathematics     | Frozen axioms, model finder, proof checker, and withheld constructor                     | Formal correctness and recovery or construction relative to the generated theory | Historical importance, conceptual value, or a new mathematical field         |
| Executable simulation     | Private simulator state and registered intervention semantics                            | Mechanism or theory adequacy under held-out simulated regimes                    | Truth about the corresponding natural system                                 |
| Scientific corpus         | Source-bearing claims, contexts, and locked reports                                      | Corpus-relative support, contradiction, prediction, and a testable proposal      | Causal truth or success of the proposed external experiment                  |
| ARTISTIC visual assurance | Registered or estimated scene declaration and an independent structural observer         | Constraint satisfaction and preservation relative to that declaration            | A unique correct image, universal perceptual validity, or artistic invention |
| DILATE artistic extension | Withheld operations in controlled studies; reuse and critical observers in open settings | Recovery or admission of a reusable operation under registered criteria          | Objective artistic value or a general theory of creative importance          |

Table 15.1: Evidence regimes in Part III. The internal reference is the testbed’s ground truth or admission authority; the final column blocks a stronger external claim.

and the most delicate interpretation of CLIC. Transforming an algebraic object or querying a countermodel is an intervention in a formal world, but it is not thereby a causal intervention. We use the name CLIC only when the generated world has a declared mechanism and intervention semantics. Otherwise the same position in the workflow is occupied by a *structural diagnostician*: it discovers invariants, dependencies, closure failures, and discriminating constructions without making a causal claim.

The controlled evidence in this chapter reaches finite declaration and presentation construction inside frozen formal languages. Its strongest positive certificates concern exact recovery, countermodel rejection, conservative extension, and reuse in generated worlds; every positive, negative, or unresolved outcome is carried by a status-bearing admission record. These results do not establish autonomous field formation or the historical importance of a constructed concept. That unreachd endpoint is why persistence appears as the final level of the hierarchy below rather than as a label attached to every successful recovery.

### 15.1 The AGENTIC mathematical contract

The maintained state contains a presented mathematical theory, an executable construction and proof language, and a sequential policy over conjectures, examples, transformations, and countermodels. The

*Chapter roadmap.* The opening sections define the mathematical contract and study controlled construction. The DIAL specialization of Universal Reinforcement Learning (DIAL-URL) then asks whether sequential model classes can be reorganized coalgebraically, before the Planck episode separates pattern repair from causal explanation.

three workflows specialize as follows.

|                             | Maintained object                                          | Question                                                   | Output packet                                                 |
|-----------------------------|------------------------------------------------------------|------------------------------------------------------------|---------------------------------------------------------------|
| CLIC or structural analogue | Laws, invariants, dependency structure                     | Which transformation separates the current explanations?   | Structural obstruction, candidate law, counterexample request |
| OPTIC                       | Definitions, constructors, proof tactics, executable tests | What new operation or interface makes the proposal usable? | Typed constructor, tactic, recognizer, or proof macro         |
| RELIC                       | Conjecture and experiment policy                           | Which query should be attempted next?                      | Selected transformation, example, theorem, or curriculum step |

The joint creative object is not an isolated theorem. It is a new primitive or presentation together with models, executable operations, nontrivial consequences, and a curriculum showing that later reasoning can reuse it. The meta-sketch records how a structural proposal becomes an OPTIC constructor and how that constructor becomes an available RELIC decision.

In the common artifact-dossier convention of Section 10.4.1, the primitive or presentation is the structural delta, its models and executable operations are the transported realization, proof and countermodel checks supply admission evidence, and the curriculum supplies persistence and reuse. The controlled experiments below fill bounded parts of that dossier. They should not be read as admitting a historically important mathematical field merely because a withheld constructor was recovered in a generated world.

For this testbed, the enclosing package  $\mathfrak{R}_{\text{math}}$  contains the mathematical presentation, model and proof semantics, executable constructor language, query policy, equivalence criterion, observers, and admission rules. A change is recorded by

$$Y_{\text{math}} : \mathfrak{R}_{\text{math}} \longrightarrow \mathfrak{R}_{\text{math}}^+$$

with the theory map as its presentation component when applicable. The registered controls  $\text{Ctl}_{\text{math}}$  include no change, parameter recovery, family selection, stronger search in the current constructor grammar, matched proposal search without the structural diagnostic, shuffled

queries or countermodels, and evaluator oracles used only as ceilings. Each evidence rung declares the applicable subset. Failure of a weaker control is evidence for opening a construction branch, not evidence that the resulting concept is valuable or transformational.

This chapter also fixes the provenance convention for Part III. Each named experiment denotes an archived unit containing a frozen registration or protocol, executable implementation, output records, and a result summary. An admission verdict applies only to that unit's declared world family, observer, candidate language, data split, and gates. The successive evidence sequences withhold progressively more structure; unless a table explicitly pools them, their percentages are not entries in one common benchmark.

#### Boundary

Exact proof can admit a theorem relative to formal axioms. It cannot establish that the theorem introduces a valuable concept, nor that a transformation of a formal object has the semantics of a causal intervention. Those are separate admission questions and require separate evidence records.

### 15.2 *Historical and generated worlds*

Historical controls test reconstruction under carefully restricted evidence. Generated mini-worlds test invention where memorized names and standard proof templates cannot supply the answer directly.

Historical cases such as Galois theory or the development of group theory are motivation, not benchmarks of autonomous rediscovery. Their surrounding languages, failed attempts, and later interpretations are difficult to freeze. They are best used to ask whether AGENTIC's audit vocabulary—obstruction, new primitive, transport, consequence, and reuse—describes the episode without claiming that the historical mathematician literally executed DIAL.

Generated worlds permit a stronger test. The evaluator can sample a finite algebraic presentation, hide one generator, quotient, symmetry, or invariant, and release only typed queries. Novel names prevent direct retrieval, while a machine-held model supplies exact counterexamples. The frozen reference space specifies which concepts and constructors the system already possesses.

### 15.3 *A hierarchy of mathematical construction*

Mathematical success is not a single endpoint. At least five levels should be kept separate:

1. **Parameter recovery** fits constants inside a supplied law.
2. **Family selection** chooses among supplied presentations.
3. **Candidate declaration extension** introduces a typed operation or property missing from the seed language.
4. **Candidate presentation construction** creates a variable-size family of objects, maps, and equations, evaluated up to an explicit equivalence.
5. **Admitted persistent theory construction** makes later definitions, proofs, experiments, or decisions possible and demonstrably easier.

The first two levels can be scientifically useful without being transformationally creative. The third crosses the declared seed-language boundary, but the new declaration may still be a familiar operation reconstructed from a strongly informative probe. The fourth asks the system to determine the size and arrangement of a candidate extension. Admission at either level still requires the applicable independent checks. Only the fifth begins to approach the field-building sense of mathematical creativity that motivates this book. The experiments below move through the middle levels; none demonstrates the fifth at historical scale. Two abbreviated evidence families recur below. *Infinitesimal Probes of Creativity* (IPC) studies hidden mathematical presentations; the *AI-Feynman-inspired* family (AF) studies symbolic-law construction.

| Evidence family                      | Strongest supported construction                                                                               | Boundary not crossed                                                        |
|--------------------------------------|----------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------|
| Lea                                  | Reusable Lean declarations induced by an explicit structural obstruction                                       | Semantic persistence and open-ended concept formation                       |
| IPC                                  | A variable-size latent presentation, scored up to permutation and scaling                                      | Reliable proposal-side abstention outside the registered doctrine           |
| AF                                   | An unnamed executable generator inside a typed semantic grammar                                                | Open proposal coverage once the candidate vocabulary is removed             |
| DIAL-URL                             | A behavior-functor expression and an oriented comparison cell inside a supplied constructor and repair grammar | A new primitive, distributive law, or persistent downstream learning theory |
| Cellular automata and AGENTIC-MATH-o | Proposed tests of macroscopic ontology and integrated reuse                                                    | No empirical result is reported here                                        |

This map is an evidential ordering, not a leaderboard. The families use different worlds, observers, proposal languages, and admission contracts. Their results show where particular construction interfaces work or fail; they do not establish a common percentage of mathematical creativity.

### 15.4 *An AGENTIC mathematical episode*

A failure-first episode begins when RELIC's current conjecture policy repeatedly encounters examples that cannot be separated in the existing vocabulary. The structural component selects transformations whose outcomes distinguish a missing invariant from a bad conjecture. OPTIC constructs a recognizer, canonicalization procedure, quotient operation, or proof macro for the proposed primitive. RELIC then uses the admitted operation to formulate and test a new family of statements. Feedback may revise the definition without silently discarding counterexamples that motivated it.

The component admission records are different. Structural diagnosis is checked by held-out transformations and countermodels. The OPTIC artifact must type-check, terminate under its declared contract, and agree with an independent reference implementation where one exists. RELIC must show improved query efficiency or new solvable problems under a matched budget. The interface certificate asks whether the same mathematical object is transported across all three stages.

### 15.5 *Admission records and formal certificates*

Proofs, countermodels, independence results, conservative extensions, and machine-checked definitions form the principal admission artifacts.

A *certificate* here is positive, checkable evidence for a particular obligation: a proof term, a countermodel, a successful type check, or a verified transport. The enclosing admission record is broader. It names the contract, collects applicable certificates and failures, and returns *admit*, *reject*, or *unresolved*. This distinction prevents an abstention or failed construction from being described as though it possessed a positive certificate.

Admission should be plural. A definition can be conservative but vacuous; a theorem can be correct but unrelated to the obstruction; a useful heuristic can lack proof. Reports therefore separate formal validity, novelty relative to the frozen language, explanatory or generative value, and persistence. An oracle translation into the evaluator's hidden theory is used for scoring, not shown to the proposal system.

Exact source syntax is generally the wrong equality criterion. Two definitions may differ by a renaming, a basis permutation, a common scale, an algebraic identity, or a different but equivalent presentation. Admission must therefore declare its quotient before proposals are scored. This issue arose concretely in the symbolic experiments below: a strict string-matching protocol rejected several mathematically equivalent generator definitions. The corrected protocol preserved that failed record and replaced textual identity by typed semantic equivalence on



hidden points. This is not a favorable scoring adjustment; it is part of specifying what mathematical object the experiment claims to recover.

### 15.6 *Persistence*

A proposed concept will be evaluated not only on its first theorem but on the curriculum of definitions, lemmas, counterexamples, and questions that it makes possible for subsequent agents.

Persistence is tested by freezing the new vocabulary and handing it to a fresh reasoner. The later curriculum should contain problems whose shortest or most reliable solutions use the new primitive, alongside negative controls for which it supplies no benefit. This prevents a one-off abbreviation from being mistaken for theory construction.

### 15.7 *Early calibration: theory extension with Lea*

A first calibration tests this distinction in a small formal environment using Lea as an interactive proof-agent harness around Lean 4. These experiments do not ask whether a frontier model can rediscover a historical mathematical field. They ask a narrower prior question: when several correct statements exhibit a common structure, under what conditions does a model turn that structure into a persistent object in the maintained theory?

The calibration separates four capabilities that are easily conflated.

1. **Proof repair** makes a statement type-check within the existing declaration language.
2. **Proof reorganization** exposes dependencies among existing statements without changing that language.
3. **Declaration accommodation** installs a reusable typed object and transports earlier statements through it.
4. **Semantic transfer** shows that later, previously unseen theorem families can reuse the installed object.

All runs use frozen public interfaces, forbidden-construction audits, independent Lean elaboration, and held-out theorem clients. The proposal model is Qwen3-Next-80B-A3B-Instruct-6bit running on a local exo cluster, except for the initial GPT-OSS interface diagnostic. Table 15.2 records the first evidence sequence.

| Experiment | Structural pressure                                                                                                                                                | Observed outcome                                                                                                                                                                                                      |
|------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Lea-0/0.1  | The target concept and interface were registered in advance.                                                                                                       | The tool-capable replication completed the write-check-repair loop and passed independent Lean and held-out admission. This calibrated the infrastructure, not autonomous concept formation.                          |
| Lea-1      | The abstraction's name, signature, and implementation were withheld, but the prompt explicitly requested a reusable Prop-valued abstraction.                       | The model introduced <code>q_is_fixed_by</code> , used it throughout the public theory, and passed the frozen transfer suite after typed and structural repair.                                                       |
| Lea-2      | Correctness and ordinary held-out consequences were required, with no request to introduce an abstraction.                                                         | The model recognized invariance and equivariance in prose and proved every theorem, but added no persistent concept: fixed-vocabulary assimilation.                                                                   |
| Lea-2B     | A correct Lea-2 theory was rejected by an explicit structural duplication obstruction: score 14 against an admission budget of 2. The repair itself was not named. | The model introduced <code>invariant</code> and <code>homomorphism</code> , reduced the score from 14 to 2, preserved all clients, and passed every registered gate.                                                  |
| Lea-2C     | The scalar score was removed. The model was told only that its theory would be frozen before an undisclosed related family arrived.                                | It recognized semi-conjugacy and reorganized proof dependencies, but installed no Prop-valued concept. It also failed kernel admission at the Stage-A turn ceiling, so the withheld transfer family was not revealed. |

Table 15.2. The first Lea calibration ladder. The comparison varies how a structural need for a new declaration becomes observable while keeping formal admission independent of the proposal model.

#### Experiment: AGENTIC-Lea: early theory-extension calibration

**Maintained theory:** a standalone Lean library of pointwise quiet and aligned transformation laws.

**Candidate construction:** a persistent declaration that reorganizes the laws without changing their mathematical interfaces.

**Admission:** source audit, independent kernel elaboration, frozen held-out consequences, declared persistence, and prospective reuse.

**Positive result:** under an explicit structural obstruction, Lea-2B constructed and reused two admissible concepts.

**Negative controls:** correctness alone in Lea-2 and a qualitative promise of future transfer in Lea-2C did not induce declaration accommodation.

**Boundary:** this is a small formal construction demonstration, not evidence of open-ended mathematical creativity.

The comparison gives a sharper interpretation of the positive Lea-2B result. Correct proofs are insufficient, and verbal recognition of the common concept is insufficient. In this calibration, accommodation occurred only after an independent structural diagnostic made the inadequacy of the current presentation explicit. The model then crossed a genuine, if small, boundary: it moved from naming a latent relationship

in prose to installing that relationship as a reusable component of a formal theory.

The result must remain modest. The disclosed duplication budget strongly favored factoring repeated propositions, the theory was tiny, and the diagnostic measured a syntactic proxy rather than mathematical fruitfulness. Lea-2C also contains a proof-engineering confound: it failed elaboration as well as declaration accommodation. These experiments therefore establish a controlled distinction among proof repair, conceptual recognition, and formal theory extension; they do not establish that the resulting concepts are deep, surprising, or historically novel.

A stronger evaluation would make semantic pressure observable without naming a desired concept or publishing a scalar complexity target. Related theorem families can arrive sequentially while the declaration layer is frozen between rounds. Admission would then ask whether a concept installed before the final family becomes visible enables zero-edit transfer to it. This would test whether recurring mathematical structure, rather than a disclosed linter target, can drive accommodation.

### 15.8 *A worked episode: inventing an intrinsic property*

The IPC-3 generated worlds provide a small but complete example of declaration accommodation. Each world presents mixtures of materials with masses  $m_1, m_2$ , initial temperatures  $t_1, t_2$ , and equilibrium temperature  $t_f$ . The seed theory contains the uniform balance law

$$m_1(t_1 - t_f) + m_2(t_2 - t_f) = 0.$$

This law succeeds when the materials respond uniformly but fails persistently in the positive worlds. The proposed repair adjoins a typed intrinsic property

$$c : \text{Material} \longrightarrow \mathbb{R}_{>0}$$

and replaces the seed equation by

$$c(a)m_1(t_1 - t_f) + c(b)m_2(t_2 - t_f) = 0.$$

The public mixtures identify the values of  $c$  only up to a common positive scale. Consequently, the admitted object is a projective coefficient family, not one privileged numerical vector. A compiler fixes a gauge for execution, but scoring is invariant under rescaling.

This example separates three possible responses to failure. In a uniform control, the old law already works and no new generator is justified. In an intrinsic-property world, pairwise ratios of  $c$  are consistent across mixtures and the extension predicts unseen combinations.

In a contextual negative control, the implied ratios change from one situation to another; forcing a single material property would merely conceal the obstruction. Correct behavior is therefore reuse, extension, and abstention, respectively.

Eight blinded worlds contained four intrinsic-property cases, two uniform controls, and two contextual controls. GPT-OSS 20B selected the correct structural response in seven of eight worlds from the records alone. When given a deterministic diagnostic containing naive residuals and implied pairwise ratios, it was correct in all eight. Every admitted extension transported the public evidence and predicted four held-out mixtures exactly; both contextual controls were rejected in both conditions. Completion-token use fell by 75% and elapsed time by 67% under the diagnostic, even though its prompt was longer.

#### Experiment: IPC-3: Intrinsic-Property Accommodation

**Seed theory.** Uniform mass-weighted balance.

**Obstruction.** Persistent material-dependent residuals across public mixtures.

**Constructed object.** A positive material property and a repaired balance law, interpreted up to common scale.

**Admission.** Evidence transport, exact held-out prediction, uniform reuse controls, and contextual abstention controls.

**Outcome.** Diagnostic-assisted construction was admitted in 8/8 worlds; language-only construction was admitted in 7/8.

**Boundary.** The signature of the possible property and the repair family were registered. The experiment tests typed accommodation, not unrestricted invention of a physical concept.

The construction resembles the role of specific heat in classical scientific discovery systems, but its importance here is logical rather than historical. A failed equation is not repaired by lowering a scalar loss alone. It is repaired by adding an object of a declared type, changing the admissible equations, transporting old observations, and surviving counterfactual controls that make the attractive extension invalid.

### 15.9 From interaction laws to latent presentations

The IPC ladder asks how much of the missing presentation is supplied in advance. Its early worlds contain two opaque finite-state generators  $g$  and  $h$ . A candidate interaction law has the form

$$h(g(x)) = g^e(h(x)),$$

where opaque state names prevent direct arithmetic on labels. Later worlds replace a single exponent by a missing intrinsic property and

then by an unknown collection of latent factors. A typed compiler never chooses the mathematical proposal; it checks public support, constructs the declared package, and refuses unsupported extensions.

| Rung  | Withheld structure                                       | Principal result                                                                                                                                               | Interpretive limit                                                                |
|-------|----------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|
| IPC-1 | Exponent in one supplied conjugacy family                | Compiled proposals were admitted in 6/8 worlds versus 4/8 one-shot packages.                                                                                   | The family was given; the model inferred a parameter and completed a declaration. |
| IPC-2 | Choice among commutation, conjugacy, and no-law families | The initial mixed-family proposer admitted 2/8. In a fresh paired study, an explicit public residual diagnostic yielded 8/8, versus 5/8 for language only.     | Diagnosis, rather than compilation, was the dominant bottleneck.                  |
| IPC-3 | One missing intrinsic material property                  | Diagnostic-assisted construction and controls passed 8/8; language only passed 7/8.                                                                            | The property signature was registered.                                            |
| IPC-4 | Number and partition of one to three latent factors      | Exact count and partition rose from 4/8 to 8/8 with pairwise minor residuals.                                                                                  | Exact, noiseless proportionality made the observer strong.                        |
| IPC-5 | The same variable-size theory under affine noise         | Recovery rose from 0/8 to 6/8; all six supported worlds passed, but both out-of-doctrine worlds were forced into false proposals and rejected by the compiler. | The observer recovered supported structure, but proposal-side abstention failed.  |

Table 15.3. Generated mathematical construction from a supplied interaction family to a variable-size latent presentation. Protocols and budgets differ by rung, so the entries form a diagnostic ladder rather than one pooled benchmark.

The IPC-2 distinction matters for provenance. The 2/8 figure belongs to the original mixed-family proposer-compiler pilot at a 4,000-token ceiling. The later diagnostic study used fresh worlds, paired language-only and diagnostic-assisted conditions, and a matched 2,200-token ceiling; its corresponding figures are 5/8 and 8/8. The diagnostic result therefore supports a paired treatment comparison, while the earlier pilot remains the failure that motivated the instrument. The two studies must not be read as arms of one experiment.

IPC-4 is the first study in which the extension has variable size. Six observables must be grouped into the smallest collection of proportional directions, and neither the factor count nor the partition is supplied. The result is defined only up to permutation and independent scaling of latent directions. The assisted model receives pairwise minor residuals but must form their transitive closure, respect the doctrine on factor size, and emit the complete projection. Hidden rows test whether the proposed relations persist.

IPC-5 removes exact proportionality. Centered-correlation diagnostics recover all supported one-, two-, and three-factor worlds, yet the proposal engine forces two four-direction controls into the declared limit of three factors. The compiler rejects both, so admission remains safe while discovery is incorrect. This is a central negative result: a

closed theory doctrine can create pressure to explain every observation even when the correct conclusion is that the doctrine is inadequate. Constructive creativity therefore requires an explicit representation of unsupported alternatives and abstention, not only a richer generator language.

### 15.10 *Symbolic laws and active mathematical experiments*

The AI-Feynman (AF) ladder moves from latent presentations to symbolic law construction. Generated worlds use opaque variable names and hide coefficients. A deterministic compiler fits numerical parameters only after a structural proposal has been made, then tests fresh points. The scientific question is whether the expression grammar itself is adequate and, when it is not, whether a typed generator can be added conservatively.

AF-0 contains additive, multiplicative-power, inverse, symmetric-interaction, and nested-composition laws. The fixed grammar omits reciprocation. Fixed search recovers 5/10 theories; a typed residual diagnostic recovers and admits all ten, including two conservative reciprocal extensions. Adding an active query changes no decision because the noiseless residual spectrum is already decisive. This negative ablation supports an acquisition rule for these generated worlds: purchase an experiment only while the admissible theory set remains ambiguous.

AF-1 deliberately restricts public observations to manifolds on which pairs of noisy laws agree. Diagnostics alone recover 7/12 families. A random generic query and a maximum-disagreement query both identify the correct family in 12/12; strict admission is 12/12 and 10/12, respectively, because two active responses fail to transport the required grammar repair. Active selection reduces completion cost, but almost any point off the agreement manifold is informative, so it cannot claim an accuracy advantage over random exploration.

| Rung | Language boundary                                                | Result                                                                                                                      | Localized obstruction                                                        |
|------|------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------|
| AF-0 | Select a named family; reciprocation is missing from the grammar | Typed diagnostic 10/10, fixed grammar 5/10.                                                                                 | Unconditional active probing adds cost after ambiguity has vanished.         |
| AF-1 | Resolve two noisy families that initially agree                  | Random query admission 12/12, active query 10/12; both identify 12/12.                                                      | Evidence transport can fail after correct discrimination.                    |
| AF-2 | Write an unnamed unary generator from a complete typed profile   | Typed profile recovers 7/8, raw expandable generation 1/8.                                                                  | The profile localizes a small, compiler-recognized semantic class.           |
| AF-3 | Recover from partial noisy profiles with sequential probes       | Active and two-random-probe policies recover 8/8 and reject 2/2 controls; active uses 0.5 added probes per supported world. | Candidate semantics remain registered.                                       |
| AF-4 | Remove the explicit candidate vocabulary                         | Random probes recover 4/8, active probes 2/8.                                                                               | Compiler identifiability does not imply proposal-engine identifiability.     |
| AF-5 | Select probes from the model's own compiled proposal set         | Random revision recovers 4/8, proposal-aware active revision 3/8.                                                           | A query cannot discriminate a true generator absent from the proposal cover. |

Table 15.4. The categorical AI-Feynman ladder. Recovery means typed semantic admission on hidden points, not merely low error on the visible profile. AF-0, AF-2, AF-3, and AF-4 report their prospectively repaired v2 protocols; the corresponding v1 records are retained as protocol or calibration failures and are not pooled with these values.

AF-2 crosses the clearest declaration boundary in this sequence. It withholds family names and asks for a fresh typed expression from a normalized unary profile. Accepted constructions include reciprocal, absolute value, cube, and base-two exponential. Seven of eight typed-profile cases are admitted, whereas raw expandable generation admits only one. The task remains controlled: the profile is externally localized, the semantic registry is small, and the expression grammar is supplied. Nevertheless, the output is an executable operation rather than a label attached to a supplied family.

The v2 qualifier is substantive rather than cosmetic. AF-0 quarantined its setup record before any model call when fixture replay exposed a serialized variable-role bug. AF-2 v1 compared surface strings and rejected mathematically equivalent expressions; v2 scores semantic equivalence classes. AF-3 v1 assigned the same domain type to every generator and incorrectly rejected valid scalar-valued declarations; v2 registers generator-specific domains. AF-4 v2 prospectively raised the completion budget after the smaller v1 budget truncated twelve of thirty calls. These records expose evaluator and inference-budget defects, not extra model trials from which the best outcome was selected.

AF-3 shows how constructive declaration can be combined with conservative experiment choice. Reciprocal and cubic maps agree on the initial normalized measurements at  $-1$  and  $1$ , while absolute value and exponentiation are already identifiable. The sequential policy

queries only ambiguous worlds, recovers all eight supported generators, rejects sine and sign controls, and uses one additional probe in each ambiguous supported case. It matches a two-random-probe ceiling with one quarter of the added measurements on those worlds.

The open-search rungs expose the unresolved difficulty. In AF-4 the external selector can isolate one class inside its finite enumerator, while the model continues to entertain interpolating polynomials and other expressions outside that privileged space. In AF-5 the query is selected from the model's own compiled candidates, but the true operation is often missing from that set. Proposal-aware selection is then optimal relative to an inadequate cover. These failures give the mathematical version of the observability caveat: no experiment can separate a hypothesis that the construction process is unable to express.

#### Design principle

Active mathematical discovery must close the loop around the proposal language actually used by the reasoner. A discriminating query in an oracle's hypothesis space is not necessarily discriminating for the learner.

### 15.11 Cellular-automaton theory induction

Cellular automata provide procedurally generated worlds with exact local semantics and potentially unfamiliar global behavior. A benchmark generator can choose a new alphabet, neighborhood, equivariant local rule, and initial condition after the proposal model's training cutoff. The system observes selected trajectories and interventions without receiving the hidden rule or the evaluator's macroscopic vocabulary.

The challenge has four levels:

1. **Microscopic recovery.** Infer a local update rule that reproduces the observed trajectories and survives held-out cell interventions.
2. **Structural recovery.** Identify translation symmetries, conservation laws, invariant subspaces, attractors, and compositional factorizations.
3. **Macroscopic invention.** Propose persistent domains, particles, interfaces, collision classes, or other higher-level objects together with maps from cell configurations to those objects.
4. **Persistent theory construction.** Use the invented vocabulary to predict new interactions, compress families of proofs or simulations, and transfer to related automata under changes of coordinates or local rules.



The third level is the decisive creativity test. Naming a visually striking pattern is insufficient. A macroscopic object must have an operational recognizer, transformation laws, boundary conditions, and consequences that are independently executable. The fourth level asks whether the new theory continues to organize later experiments rather than disappearing after the trajectory that suggested it.

In the AGENTIC realization, the structural component chooses cell, rule, and pattern perturbations that discriminate candidate microscopic and macroscopic explanations. OPTIC constructs detectors, coarse-graining maps, collision tests, and reusable simulation macros. RELIC allocates experiments across initial conditions and time horizons, then decides when an admitted macroscopic object should enter its planning vocabulary. The closed loop is essential: a detector without an experiment policy may never encounter a decisive collision, while an experiment policy without a typed detector cannot recognize the event it sought.

#### Experiment: Cellular-automaton emergence benchmark

**Epistemic status:** proposed benchmark design, not a completed result.

**Input:** blinded trajectories and a registered set of cell, pattern, rule, and neighborhood interventions from a newly generated automaton.

**Proposal:** a microscopic rule, structural sketch, and candidate macroscopic theory with explicit abstraction and realization maps.

**Admission:** exact replay, held-out interventions, invariant checks, prediction of collisions and long-horizon behavior, and reuse on later rules.

**Creativity label:** rule inference is Mathematics–E; an admitted, predictively useful macroscopic ontology is a candidate Mathematics–T result.

**Boundary:** success is relative to the generated world and does not establish that the same ontology describes a natural system.

### 15.12 *An integrated benchmark*

The preceding ladders isolate declaration, diagnosis, compilation, and active query selection. A full AGENTIC benchmark must test whether these capabilities compose rather than merely coexist. AGENTIC-MATH–o begins with finite symmetry worlds. The evaluator withholds the number or arrangement of latent factors, a canonicalization operation, and the decision rule that would exploit the resulting orbit structure. The system receives exact transformations and bounded

active queries. It must diagnose the missing invariant, construct an executable normal-form or quotient skill, and use that skill to select shorter discriminating query sequences in held-out worlds.

The full factorial comparison from Chapter 14 separates the structural component, OPTIC, RELIC, their pairs, and AGENTIC. Additional controls receive the hidden invariant, canonicalizer, or optimal query policy separately. Primary measures are exact structural recovery, constructor validity, query cost, false extension, and transport under renamed generators and changed factor counts. The central interaction question is whether the full system constructs a reusable primitive that no component or pair can both discover and exploit.

#### Experiment: AGENTIC-MATH-0

**Epistemic status:** proposed integrated benchmark, not a completed AGENTIC result.

**World:** generated finite symmetry systems with blinded names and withheld latent-factor organization.

**Joint construction:** invariant or quotient, executable canonicalization skill, and an adaptive query schema using it.

**Admission:** exact countermodels and held-out transformations, typed execution, improved query efficiency, and transport under renaming and changed factor count.

**Controls:** full component factorial, flat search, and separate structure, skill, and policy oracles.

**Boundary:** formal success establishes a result in the generated world, not historical-scale mathematical creativity.

### 15.13 *A grand challenge: constructing a coalgebraic theory of learning*

The preceding benchmarks construct operations and presentations inside finite mathematical worlds. A more ambitious test asks whether a learning system can revise the mathematical type of dynamical system through which it understands an environment. Reinforcement learning supplies an unusually clean setting for this challenge. Much of the field is organized around separately named model classes—Markov decision processes, partially observable Markov decision processes, semi-Markov decision processes, and predictive state representations—even though universal coalgebra provides a common language for their unfolding behavior [Rutten, 2000]. Categories for AGI develops this coalgebraic perspective as Universal Reinforcement Learning, or URL [Mahadevan, 2026b]. The creative question considered here is not whether an agent can solve one more supplied coalgebra. It is whether the agent can

recognize that its current behavior functor is inadequate and construct a better one.

Let  $\mathcal{C}$  be a category of state objects and let

$$F : \mathcal{C} \longrightarrow \mathcal{C}$$

specify a behavioral type. An environment of that type is an  $F$ -coalgebra

$$\xi : X \longrightarrow FX.$$

The functor declares what may be observed in one unfolding step: actions, rewards, probability distributions, sets of possible successors, durations, outputs, or combinations of these. A final  $F$ -coalgebra, when available, provides canonical behavior, and coalgebra morphisms and bisimulation compare systems by what they do rather than by the names of their hidden states.

For example, the following equations are schematic finite-state presentations, not assertions that every implementation lives in **Set**:

$$\begin{aligned} F_{\text{MDP}}X &= (\mathbb{R} \times \mathcal{D}X)^A, \\ F_{\text{SMDP}}X &= (\mathbb{R} \times \mathcal{D}(T \times X))^A. \end{aligned}$$

A POMDP adds a hidden-state coalgebra and an observation channel  $X \rightarrow \mathcal{D}O$ . A predictive state representation is more naturally read as an observable realization or quotient of unfolding behavior than as merely another hidden-state tuple. These distinctions must be retained: coalgebra organizes behavior, but reward, control, and policy optimization require additional algebraic structure.

#### 15.13.1 Assimilation and accommodation over behavior functors

**COALGEBRAIC ASSIMILATION.** Fix  $F$ . The learner estimates  $\xi$ , observations, rewards, policies, value semantics, or a behaviorally minimal quotient. Bellman evaluation can then be studied through its algebraic fixed point, while metric coinduction provides a quantitative proof principle for contractive realizations [Kozen and Ruozzi, 2009]. This direction may encompass difficult statistical learning, but it does not change the declared type of system.

**COALGEBRAIC PROTO-ACCOMMODATION.** Persistent failure may instead justify

$$F \rightsquigarrow F^+.$$

The repair could add an observation channel, holding time, nondeterministic branching, termination effect, continuous measure structure, or a law governing how two effects distribute over one another. Sokolova's

grammar of probabilistic behavior functors supplies a principled initial language for such changes: identity and constant functors, action-indexed exponentials, powerset and distribution functors, and products, coproducts, and compositions constructed from them [Sokolova, 2011]. Natural transformations between behavior functors provide typed system translations. A distributive law may be required when the repair combines effects rather than merely adjoining a component. The resulting proposal becomes accommodation only after its realization, transport, and consequences pass the registered admission contract.

The DIAL directions are therefore

$A$  : estimate and optimize inside a fixed  $F$ ,

$B$  : revise  $F$  or its effect-combination laws.

Their mixed obstruction asks whether learning and theory transport agree. If  $\eta : F \Rightarrow F^+$  is a candidate translation, compare learning an  $F$ -coalgebra and then transporting it with transporting the declaration first and learning in the enlarged class. Here transport of  $\xi : X \rightarrow FX$  along an actual natural transformation means the  $F^+$ -coalgebra  $\eta_X \circ \xi : X \rightarrow F^+X$ . A persistent discrepancy in held-out behavior localizes a mixed defect, but it does not by itself show that  $F$  is inadequate: estimation error, optimization failure, or a bad translation may also break the comparison. Only after those alternatives are controlled does the defect support a candidate change of behavioral type.

This comparison is not automatically a Lie bracket. Parameters of coalgebras inside one smooth chart may admit infinitesimal vector fields and mixed brackets. Changing the functor itself may be a finite declaration change. DIAL-URL therefore uses infinitesimal diagnosis within a chart, but finite sketch extension, evidence transport, and typed admission when it crosses a model-class boundary.

#### Design principle

The DIAL-URL objective is not only to learn a policy or an environment model, but to determine when the behavioral type used to formulate those objects must itself be revised.

#### 15.13.2 The admitted artifact

Selecting “POMDP” from a supplied menu is family selection, not yet persistent theory construction. The stronger target is a reusable package

$$\mathcal{T}^+ = (F^+, \xi^+, O^+, \text{Ctrl}^+, \sim_{F^+}, \eta),$$

containing a behavior functor, a realized coalgebra, observation and control interfaces, a declared behavioral equivalence, and a transport

map from the old theory. When  $\eta$  is claimed to be a natural transformation, its naturality and its action on coalgebras belong to the certificate; a typed syntax translation alone is not enough. Its admission contract has six parts:

1. a declared embedding or translation preserves the old models and observations on the subdomain where the old theory remains valid;
2. the proposed extension explains the localized residual on fresh data;
3. behaviorally equivalent hidden presentations receive the same score;
4. downstream prediction or control improves under a matched budget;
5. unsupported extensions are rejected or retained as unresolved; and
6. the new constructor transfers to later environments without rebuilding the entire learner.

The fourth condition prevents coalgebraic elegance from substituting for a learning result. The third prevents recovery of a privileged state syntax from being mistaken for discovery. The fifth is essential because passive trajectories may leave several behavior functors observationally equivalent.

No experiment reported below satisfies this full six-part contract. DIAL-URL-0 through -4 isolate type recovery, finite-data uncertainty, active discrimination, expression synthesis, and a withheld comparison law. An *admit* label inside a rung is a trial-level verdict under that frozen contract; the experiment-level endpoint audit remains separate. In particular, URL-4 did not pass all of its preregistered endpoint families. None of the rungs demonstrates downstream control improvement or persistent transfer of a newly invented constructor to later environments; the package  $\mathcal{T}^+$  remains the grand-challenge target.

### 15.13.3 The DIAL-URL evidence sequence

This ladder moves deliberately from classification to construction.

1. **DIAL-URL-0: exact type recovery.** Exact finite operators are drawn from MDP, POMDP, SMDP, and partially observable semi-Markov families. The learner must reuse the MDP theory, adjoin only the required constructors, and quotient behaviorally redundant hidden states rather than interpreting every state alias as partial observability.
2. **DIAL-URL-1: finite-trajectory accommodation.** Operators and their uncertainty are estimated from trajectories. Admission requires calibrated false-extension control and abstention when the available horizon does not separate candidate types.

3. **DIAL-URL-2: active discrimination.** RELIC selects actions, initial conditions, and observation probes that maximally separate the currently supported behavior functors.
4. **DIAL-URL-3: grammar synthesis.** Named model classes are removed. The system constructs the smallest supported functor expression from the registered probabilistic grammar and supplies a checked syntactic transport chain from the seed expression. These finite certificates stand in for, but do not establish, natural transformations between behavior functors in an arbitrary category.
5. **DIAL-URL-4: withheld composition law.** An environment combines effects whose interaction is absent from the grammar. Success requires a new oriented comparison generator in the finite presentation, not selection from a named catalog. Naturality, coherence, and distributive-law axioms remain outside this rung.

#### 15.13.4 *The synthetic laboratory behind the ladder*

The word *environment* hides an important change across these five rungs. DIAL-URL-0 and DIAL-URL-1 use small sequential simulators with states, actions, rewards, observations, transitions, and holding times. DIAL-URL-2 through DIAL-URL-4 instead use controlled *counter-witness oracles*. They do not render a physical scene or expose raw state trajectories. Given a typed experimental request, an oracle returns a noisy binary indication of whether a structural phenomenon was observed. DIAL-URL-2 through -4 therefore isolate theory revision from the difficult upstream problem of estimating a reliable witness from raw data.

It is useful to write a synthetic laboratory as

$$\mathcal{L} = (\mathcal{T}_\star, \mathcal{Q}, \{P(Y_q \mid \mathcal{T}_\star) : q \in \mathcal{Q}\}),$$

where  $\mathcal{T}_\star$  is the evaluator's hidden theory,  $\mathcal{Q}$  is the registered intervention or probe language, and  $Y_q$  is the observable outcome of applying probe  $q$ . The learner knows the candidate language and likelihood model but not  $\mathcal{T}_\star$ . It chooses probes, updates a posterior over theories, and either admits a typed extension or abstains. In the first two rungs, the distribution of  $Y_q$  is estimated from exact operators or raw episodes. In DIAL-URL-2 through -4 it is specified directly, like a calibrated instrument in a deliberately simplified physics experiment.

| Rung  | Hidden experimental world               | Evidence visible to the learner                                      | Withheld structure                                    |
|-------|-----------------------------------------|----------------------------------------------------------------------|-------------------------------------------------------|
| URL-0 | Exact two-state controlled coalgebra    | Rewards, observation map, transition distributions, duration support | Minimal behavior type                                 |
| URL-1 | Stochastic two-state episodic simulator | Finite records $(o, a, r, \tau, o')$                                 | Whether observation history or holding time is needed |
| URL-2 | Four-hypothesis intervention oracle     | Noisy history and duration counter-witness bits                      | Best experiment and supported registered type         |
| URL-3 | Recursive behavior-grammar oracle       | Noisy lag, time, and mixture witness bits                            | Smallest generated functor expression                 |
| URL-4 | Expression-plus-law oracle              | Marginal witnesses and ordered compound witnesses                    | Constructor pair and orientation of a comparison cell |

Table 15.5. The evidential boundary of the DIAL-URL ladder. Only URL-1 begins from sampled sequential records; URL-2 through URL-4 deliberately factor out witness estimation.

### 15.13.5 The concrete sequential worlds

The exact URL-0 laboratory has two evaluator states  $x_0, x_1$  and two actions, `stay` and `flip`. In the direct world, `stay` preserves  $x_i$ , `flip` maps it to  $x_{1-i}$ , and the observation map distinguishes the two states. Duration is one and reward is zero. This is the smallest fully observed control.

Three modifications create the positive cases. First, the observation map may send both states to the same symbol. Aliasing alone is not enough to require a hidden-state constructor: in the quotient control, both states have identical rewards and induce the same distribution over next observations under every action. By contrast, the POMDP world gives  $x_1$  reward one under `stay` and  $x_0$  reward zero. The same visible symbol now conceals a behavioral distinction. Second, the pair  $(x_1, \text{flip})$  may have duration support  $\{2, 3\}$  instead of  $\{1\}$ , requiring the Time constructor. The composite world contains both defects.

URL-1 replaces these exact maps by stochastic trajectories. An episode starts in one of two hidden states and runs for 25 steps. In directly observed worlds, the visible symbol is  $o_i$ ; in aliased worlds both states emit  $o$ . For the direct dynamics, action zero makes state one likely with probability 0.9 from  $x_1$  and 0.1 from  $x_0$ , while action one reverses these tendencies. Aliased dynamics are persistent: the hidden state remains unchanged with probability 0.92 under action zero and 0.82 under action one. Rewards expose this memory only statistically—probability 0.88 in  $x_1$  and 0.12 in  $x_0$ . Semi-Markov worlds independently replace unit duration by two or three with probability 0.25.

The learner never sees the hidden state in URL-1. It compares a predictor using only  $(o_t, a_t)$  with one that also receives the previous observation, action, and reward. A significant held-out log-score improvement is evidence for an observation/history extension. A confi-

dence interval for non-unit durations supplies the independent Time decision. Thus the “experiment” is concrete: act in a two-state stochastic system, collect episodes, and ask whether history or elapsed time improves its observable one-step account.

**Boundary: A quotient control**

Suppose two hidden names both emit  $o$ , both return reward zero, and every action sends each name to the uniform distribution on the same two names. There are two simulator states but only one observable behavior. DIAL-URL must return Base. Treating hidden multiplicity itself as evidence for partial observability would be a false theory extension.

*15.13.6 What the grammar constructors mean experimentally*

The URL-3 and URL-4 constructors are operational abstractions of observable defects:

*Base* the current observation-action interface is sufficient for the registered one-step behavior;

*Delay( $E$ )* a lagged witness remains after fitting  $E$ , so an additional layer of observable history is required;

*Time( $E$ )* behavior depends on non-unit holding time rather than only on the next-state sequence;

*Mix( $E$ )* a registered mixture witness indicates branching or probabilistic effect structure not represented by  $E$ .

These are not definitions of delay, time, or probabilistic mixture in every coalgebraic category. They are the semantics of the finite experimental instrument. A lag- $k$  probe returns one with probability 0.72 when at least  $k$  delay layers are present and 0.08 otherwise. The corresponding rates are 0.70 versus 0.04 for time and 0.68 versus 0.05 for mixture. The asymmetry makes the probes informative but fallible.

URL-2 has a similarly explicit intervention menu. An *obs* probe chooses a persistent reset, an alias-separating action, and a history-sensitive readout; a *time* probe selects a duration-sensitive reset and action; a *joint* probe combines them. When the observation constructor is truly required, the *obs* probe reports a history witness with probability 0.75, compared with 0.08 when it is absent. The *time* probe reports a duration witness with probability 0.70 when required and 0.03 otherwise. The active learner chooses the instrument with greatest expected posterior information gain; the passive control cycles through all four.



URL-4 adds ordered experiments. If Time and Mix are both present but commute under the registered instrument, either order returns a compound witness with probability 0.18. If the hidden theory contains  $\chi_{\text{Time}, \text{Mix}}$ , the matching Time>Mix probe fires with probability 0.78 and the reverse probe with probability 0.10. If either constructor is absent, the rate is 0.05. The claimed discovery of a comparison cell is therefore grounded in a repeatable directional asymmetry, not in the name assigned to the world.

#### Boundary: One URL-4 discovery trace

In a representative Time(Mix(Base)) world, positive marginal time and mixture observations first moved posterior mass to that base expression. Repeated Time>Mix experiments then produced the event pattern 0,1,1,1,0,1,1. After the sixth positive directional witness, the posterior of Law(Time>Mix) reached 0.956 and crossed the preregistered admission threshold. The reverse orientation and the no-law composite remained explicit competitors until the evidence separated them.

#### 15.13.7 In what sense is this theory construction?

The analogy with experimental physics is methodological. A theory declares which observations should agree; an intervention produces a counter-witness; and persistent, localized disagreement motivates a more expressive theory. URL-0 through URL-2 select registered behavior types. URL-3 constructs an unnamed composite expression. URL-4 goes one step further by adjoining and orienting a comparison cell absent from the seed sketch.

The analogy must not be overstated. These are calibrated synthetic chambers, not natural scientific domains. The witness vocabulary, likelihoods, maximum delay depth, and generic one-cell edit schema are supplied. Consequently the experiments test whether evidence can drive conservative structural extension *once informative measurements exist*. A stronger DIAL-URL system must learn those measurements from raw trajectories, discover when their likelihood model is wrong, and transport an admitted constructor to independently built environments. That transition—from a registered structural instrument to an acquired one—is where this program meets experimental theory construction in the fuller scientific sense.

The quotient control described above is therefore conceptually decisive: it shows that hidden multiplicity alone does not warrant a richer sequential theory. URL-0 calibrates that judgment exactly; the later rungs ask whether it survives finite data, active probing, generated constructor expressions, and a withheld comparison law.

**Experiment: DIAL-URL-0: Exact Coalgebra-Type Recovery**

**Seed theory:** finite MDP behavior with action-indexed rewards and probability distributions.

**Candidate constructors:** hidden observation and non-unit holding time, with their composite.

**Obstruction:** aliased observations with behaviorally distinguishable successors, non-unit duration support, or both.

**Quotient control:** aliased but behaviorally equivalent hidden states must remain in the MDP class after minimization.

**Admission:** exact minimal-type recovery, zero false extensions, transport of the seed model, and explicit abstention outside the registered grammar.

**Epistemic status:** registered endpoint `PASS`; exact coalgebra-type selection inside a supplied constructor catalog, not a newly invented theory language.

**Boundary:** the constructors and exact operators are supplied. This is a calibration of coalgebraic accommodation, not invention of a new RL theory.

The implementation recovered all five exact worlds. It made no false accommodation on either MDP control and selected the composite extension when both counter-witnesses were present. This supports only the deterministic diagnostic and its audit trace.

### 15.13.8 *Finite-trajectory accommodation*

DIAL-URL-1 replaces exact operators by episodic records containing only observations, actions, rewards, holding times, and next observations. Hidden state, observation maps, and ground-truth types remain evaluator-only. A base categorical predictor estimates reward and next observation from the current observation–action pair. A second predictor also receives the preceding observation, action, and reward. Held-out log-score improvement, measured across episodes, supplies a confidence interval for the observation constructor. A Wilson interval for the frequency of non-unit holding times supplies the duration decision. If either interval intersects its registered practical threshold, the type-level system abstains.

Fifty independently seeded replications were run for each world at 12, 40, and 120 episodes of length 25. Table 15.6 reports exact recovery and abstention. Abstention is not counted as correct.

| World                       | Small: exact / abstain | Medium: exact / abstain | Large: exact / abstain |
|-----------------------------|------------------------|-------------------------|------------------------|
| MDP, direct observation     | 0% / 100%              | 92% / 8%                | 98% / 2%               |
| MDP, redundant hidden alias | 0% / 100%              | 88% / 12%               | 100% / 0%              |
| POMDP, significant alias    | 0% / 100%              | 78% / 22%               | 100% / 0%              |
| SMDP, non-unit duration     | 90% / 10%              | 96% / 4%                | 98% / 2%               |
| Partially observable SMDP   | 40% / 58%              | 86% / 14%               | 100% / 0%              |

Table 15.6. DIAL-URL-1 finite-trajectory type recovery. Each cell contains 50 replications. Small, medium, and large budgets contain 12, 40, and 120 episodes, respectively. Exact recovery rises with evidence, while abstention at the smallest budgets prevents unsupported accommodation from being counted as discovery.

All preregistered endpoints passed. False accommodation was zero on both MDP controls at every budget. At the large budget every world reached 98–100% exact recovery, and the composite world was never collapsed to an MDP. The mean large-budget history improvement was approximately  $-0.019$  and  $-0.003$  nats per transition in the direct and quotient MDP controls, but  $0.089$  in both partially observable worlds. Thus simulator-level aliases were not themselves treated as evidence for accommodation; only aliases that improved held-out prediction from history triggered the observation constructor.

The small-budget result reveals the intended conservative behavior and one remaining error. Both unit-time MDP controls and the POMDP abstained in every replication because the absence of a rare duration effect was not yet certified. The composite world produced 29 abstentions, 20 correct composite decisions, and one under-accommodation to SMDP when history dependence was incorrectly rejected. This error is preserved rather than absorbed into the abstention count.

**Experiment: DIAL-URL-1: Finite-Trajectory Accommodation**

**Input:** observation–action–reward–duration trajectories; no hidden states, operators, or type labels.

**Observer:** held-out history gain for Obs; Wilson uncertainty for Time.

**Admission:** minimal registered type or abstention whenever a constructor decision is unresolved.

**Outcome:** all primary endpoints passed; large-budget recovery was 98–100% with zero false accommodation on MDP controls.

**Epistemic status:** experiment-level PASS; finite-data selection among registered model patterns, not functor invention.

**Boundary:** the four model patterns, diagnostic interfaces, and thresholds were registered. The experiment does not synthesize a new functor, identify a unique hidden realization, or demonstrate improved control.

*15.13.9 Active coalgebraic discrimination*

Passive evidence may leave two behavior functors observationally equivalent even when a legal intervention would separate them. DIAL-URL-2 therefore turns model selection into sequential experimental design. Its finite probe menu contains neutral, observation-sensitive, duration-sensitive, and joint interventions. Each probe specifies an initial-condition family, an action schedule, and an observation protocol. The active learner evaluates the expected posterior entropy reduction under every probe and executes the most informative one. A matched passive learner instead cycles evenly through the same menu. Both learners receive the same prior, likelihood, observations, confidence threshold, and total budget.

The registered hypotheses are again MDP, POMDP, SMDP, and partially observable SMDP, represented by the constructor sets

$$\emptyset, \quad \{\text{Obs}\}, \quad \{\text{Time}\}, \quad \{\text{Obs}, \text{Time}\}.$$

Every trajectory probe returns two observable counter-witness summaries: whether the registered history dependence appeared and whether a non-unit holding time appeared. A type is admitted only when its posterior probability reaches 0.95; otherwise the system abstains. Five hundred independently seeded replications were run for every true type and acquisition policy. Table 15.7 compares fixed-budget behavior.

| Probes | Passive: exact / wrong / abstain | Active: exact / wrong / abstain | Exact gain |
|--------|----------------------------------|---------------------------------|------------|
| 4      | 4.75% / 0.05% / 95.20%           | 20.30% / 0.70% / 79.00%         | +15.55 pp  |
| 8      | 52.10% / 1.20% / 46.70%          | 75.45% / 1.00% / 23.55%         | +23.35 pp  |
| 12     | 69.60% / 0.60% / 29.80%          | 92.85% / 0.40% / 6.75%          | +23.25 pp  |
| 20     | 91.30% / 0.30% / 8.40%           | 99.65% / 0.00% / 0.35%          | +8.35 pp   |

Table 15.7. DIAL-URL-2 active discrimination. Rates are macro averages over four true behavior types, with 500 replications in every type-policy cell. The passive schedule is balanced, not random. Active selection increases exact admission at matched probe budget without increasing wrong admission, so the comparison concerns experiment choice rather than additional evidence.

All five preregistered endpoints passed. At eight probes, active selection improved exact admission by 23.35 percentage points while keeping incorrect admission at 1%. At twenty probes it reached 99.65% exact admission, with no incorrect admissions and only 0.35% abstention. MDP false accommodation never exceeded 0.8%, and no composite world was collapsed to an MDP at the largest budget. Median first crossing of the admission threshold fell from eight probes under passive collection to six under active collection.

The intervention trace explains the gain. The learner initially favored the joint probe, which can expose either constructor. As the posterior concentrated, it shifted toward the specialized probe that best separated the remaining alternatives. In the composite world at twenty probes, the mean allocation was 10.44 observation probes, 7.11 duration probes, and 2.45 joint probes; no budget was spent on the weak neutral intervention. The passive control necessarily spent five probes on each.

#### Experiment: DIAL-URL-2: Active Coalgebraic Discrimination

**Input:** a posterior over registered behavior functors and a typed intervention menu.

**Selection:** maximize expected information gain against a balanced passive schedule.

**Admission:** posterior probability at least 0.95, otherwise abstain.

**Outcome:** at eight probes, 75.45% active exact admission versus 52.10% passive; at twenty, 99.65% exact with no wrong admissions.

**Epistemic status:** all preregistered endpoints passed; evidence for active exploratory discrimination, not construction of a theory extension.

**Boundary:** the grammar, intervention menu, and likelihood are supplied; their operators and new functors are not learned.

#### 15.13.10 From named families to functor expressions

DIAL-URL-3 removes the four model-class names. The learner instead receives the recursive grammar

$$E ::= \text{Base} \mid \text{Delay}(E) \mid \text{Time}(E) \mid \text{Mix}(E).$$

The delay constructor may recur to depth three; time and mixture may occur at most once. Registered distributive laws normalize equivalent construction orders to

$$\text{Time}^?(\text{Mix}^?(\text{Delay}^d(\text{Base}))), \quad 0 \leq d \leq 3.$$

Best-first grammar expansion produces sixteen canonical expressions. Their prior mass decreases exponentially with constructor count, so unsupported structure is penalized rather than rewarded. Active lag, holding-time, and mixture probes update a posterior over expressions. Admission requires posterior probability at least 0.95 and a checked transport chain from the seed expression. Thus the output is not the label “partially observable semi-Markov model,” but an explicit construction such as

$$\begin{aligned} \text{Base} &\longrightarrow \text{Delay}(\text{Base}) \longrightarrow \text{Delay}^2(\text{Base}) \\ &\longrightarrow \text{Mix}(\text{Delay}^2(\text{Base})) \longrightarrow \text{Time}(\text{Mix}(\text{Delay}^2(\text{Base}))). \end{aligned} \quad (15.1)$$

Eleven evaluator worlds test direct and redundant base behavior, delay depths one through three, the two single constructors, three two-constructor composites, and the expression in Equation (15.1). The last expression is never supplied as a named alternative; it exists only because the grammar composes its constructors. Each world has 500 independent replications.

| World                  | Synthesized target            | Exact | Wrong | Abstain |
|------------------------|-------------------------------|-------|-------|---------|
| Direct base            | Base                          | 98.6% | 0.4%  | 1.0%    |
| Redundant hidden alias | Base                          | 99.0% | 0.0%  | 1.0%    |
| Delay depth 1          | Delay(Base)                   | 94.0% | 0.0%  | 6.0%    |
| Delay depth 2          | Delay(Delay(Base))            | 93.0% | 0.4%  | 6.6%    |
| Delay depth 3          | Delay(Delay(Delay(Base)))     | 97.8% | 0.2%  | 2.0%    |
| Time                   | Time(Base)                    | 99.6% | 0.2%  | 0.2%    |
| Mixture                | Mix(Base)                     | 98.6% | 0.2%  | 1.2%    |
| Delay + time           | Time(Delay(Base))             | 95.0% | 0.0%  | 5.0%    |
| Delay + mixture        | Mix(Delay(Delay(Base)))       | 96.2% | 0.6%  | 3.2%    |
| Time + mixture         | Time(Mix(Base))               | 98.8% | 0.4%  | 0.8%    |
| Unnamed triple         | Time(Mix(Delay(Delay(Base)))) | 96.8% | 0.2%  | 3.0%    |

Table 15.8. DIAL-URL-3 canonical expression recovery after thirty probes. Every row contains 500 replications. The target column displays generated grammar terms, not model-class labels.

All preregistered endpoints passed. Macro exact synthesis rose from 8.84% at ten probes to 79.38% at twenty and 97.04% at thirty; corresponding wrong admission was 0.29%, 0.71%, and 0.24%. Every world exceeded 90% exact recovery at the largest budget. The unnamed triple reached 96.8%, was never collapsed to Base, and every admitted expression carried a valid finite transport certificate. The redundant hidden-state world retained Base: simulator multiplicity alone did not induce theory extension.

This result is stronger than choosing a family name but remains bounded. The grammar, probe likelihoods, complexity prior, and distributive normalization laws are supplied. The certificate checks a typed expression chain, not naturality in an arbitrary category. Genuine transformational construction begins when a persistent obstruction cannot be expressed by this grammar and forces a new constructor or composition law.

**Experiment: DIAL-URL-3: Behavior-Functor Grammar Synthesis**

**Input:** a recursive typed grammar, probe semantics, and a seed behavior expression.

**Construction:** expand and normalize grammar terms; actively test the remaining structural distinctions.

**Admission:** posterior at least 0.95 plus a checked transport chain from Base.

**Outcome:** 97.04% macro exact synthesis at thirty probes; 96.8% on an unnamed three-constructor composite.

**Epistemic status:** all preregistered endpoints passed; admitted grammar-term construction inside a supplied language, not language extension.

**Boundary:** constructors and their composition laws are supplied; no new primitive or distributive law is invented.

*15.13.11 Discovering a withheld composition law*

DIAL-URL-3 synthesized expressions from supplied constructors and supplied composition laws. DIAL-URL-4 withholds one level more: the seed sketch still contains *Delay*, *Time*, and *Mix*, but treats their pairwise interaction as ordinary and order-insensitive. Some generated worlds instead contain an oriented comparison cell

$$\chi_{A,B} : A \circ B \Longrightarrow B \circ A.$$

The learner is not told which pair  $(A, B)$  is exceptional or which orientation is supported. It receives only the generic admissible edit form  $\text{Law}(A > B)$ , together with ordered compound probes. Thus the task is to detect that the present sketch is inadequate, localize the defect to two known constructors, instantiate a new comparison cell, and orient it.

The notation is deliberately presentation-level. In this experiment  $\chi_{A,B}$  is a typed generator whose semantics is the registered directional counter-witness oracle. The checker does not quantify over all objects and morphisms of a category, verify naturality squares, or establish the coherence axioms of a distributive law. Calling it a “law” is therefore shorthand for an admitted relation in this finite experimental sketch.

The preregistered experiment crossed ten worlds, three probe budgets, and 400 replications per cell. Four controls required no law; six targets contained a withheld oriented law, including two three-constructor composites. A theory was admitted only at posterior probability at least 0.95 and after its expression transport and comparison cell type-checked. Table 15.9 separates exact recovery, wrong admission, abstention, false-law admission, constructor-pair recovery, and orientation recovery across the registered probe budgets.

| Budget | Exact  | Wrong | Abstain | False law | Law pair | Orientation |
|--------|--------|-------|---------|-----------|----------|-------------|
| 20     | 47.85% | 0.68% | 51.48%  | 0.19%     | 47.88%   | 47.83%      |
| 35     | 91.40% | 0.58% | 8.03%   | 0.06%     | 90.79%   | 90.79%      |
| 50     | 98.58% | 0.23% | 1.20%   | 0.13%     | 98.25%   | 98.25%      |

Table 15.9. DIAL-URL-4 macro recovery. Exact, wrong, and abstention rates average over all ten worlds; false-law rates average over the four no-law controls; law-pair and orientation rates average over the six law-bearing worlds.

Seven of eight preregistered endpoint families passed. At fifty probes, exact theory recovery exceeded 90% in every world, wrong admission stayed below 5% in every cell, false-law admission stayed below 5% in every control, and all certificates were valid. The single miss was deliberately strict: every law-bearing world had to recover the correct unordered constructor pair at least 95% of the time. The hardest target, `Time(Mix(Delay(Delay(Base))))` with a `Time>Mix` law, reached 93.5% pair recovery and 93.25% exact recovery. It abstained in 5.75% of trials and made a wrong admission in 1%.

This boundary is informative. In two-constructor targets the missing pair and orientation were recovered between 99.5% and 100% of the time. With three constructors, however, six oriented comparison cells compete while the system must also recover the base expression. Active evidence selection remained conservative, converting uncertainty mostly into abstention rather than false law invention, but did not meet the uniform 95% pair-recovery requirement.



**Experiment: DIAL-URL-4: Withheld Comparison-Law Discovery**

**Input:** known coalgebraic constructors, a generic one-cell repair schema, and marginal plus ordered compound probes.

**Construction:** select the constructor pair and orientation of a new comparison cell.

**Admission:** posterior at least 0.95 plus typed expression-transport and comparison-cell certificates.

**Outcome:** 98.58% macro exact recovery at fifty probes, with 0.13% false-law admission on controls; seven of eight endpoint families passed.

**Epistemic status:** not admitted at the experiment level because one preregistered endpoint family failed. Individual trials passing their posterior and type gates construct a finite presentation generator with registered oracle semantics, not a verified natural transformation or distributive law.

**Boundary:** the repair schema is supplied, and one three-constructor world missed the preregistered 95% pair-recovery threshold.

DIAL-URL shifts the object of evaluation beyond comparing RL algorithms on a fixed benchmark. The reported ladder reaches registered type selection, grammar-term construction, and calibrated trial-level recovery of a finite comparison generator; URL-4 remains an experiment-level near miss under its frozen endpoint conjunction. Its intended stronger artifact is a conservative extension of the theory of sequential behavior: one that retains established models, introduces a justified coalgebraic constructor or composition law, verifies the required categorical laws, and makes additional classes of dynamical systems representable, experimentally distinguishable, and reusable for subsequent prediction or control. No reported rung yet establishes that full artifact.

#### 15.14 *Pattern repair before causal explanation*

PLANCK'S REPAIR OF THE BLACKBODY LAW provides a particularly compact historical example of experimentally driven theory construction [Langley et al., 1987]. Near the end of the nineteenth century, Wien's distribution law described the available measurements well in one regime. In frequency form, and suppressing only constants irrelevant to the comparison, it may be written schematically as

$$I_W(\nu, T) \propto \nu^3 \exp(-k\nu/T).$$

Measurements reported to Planck by Heinrich Rubens revealed a systematic counter-witness: at small values of  $x = k\nu/T$ , the observed spectrum scaled as  $T\nu^2$ , which requires a reciprocal  $1/x$  factor after the common  $\nu^3$  prefactor is removed. Wien's expression lacked that behavior. Planck's interpolation changed only the denominator,

$$\exp(x) \mapsto \exp(x) - 1, \quad I_P(\nu, T) \propto \frac{\nu^3}{\exp(k\nu/T) - 1}.$$

The change is visually tiny—a subtraction by one—but structurally decisive. Since

$$\exp(x) - 1 = x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots,$$

the revised spectral factor has reciprocal behavior  $1/x$  to leading order near  $x = 0$ , so  $I_P$  is proportional to  $T\nu^2$  in that limit, while retaining Wien's successful high-frequency behavior. A local discrepancy therefore selected a repair that joined two empirical regimes.

This episode separates two achievements that are often collapsed under the name *discovery*. The first was a pattern-level accommodation: identify where the old expression failed, search a small language of admissible changes, and find an economical formula consistent with both regimes. The second was the deeper physical explanation that led Planck to energy quantization. The repaired law preceded a satisfactory account of why that law should govern radiation. In the terminology of this book, empirical admission of a new relation need not certify its causal semantics.

The Taylor expansion also admits a useful retrospective reading through the Weil probes introduced in Chapter 3. For a first-order infinitesimal  $d$  satisfying  $d^2 = 0$ ,

$$\exp(d) - 1 = d.$$

This is not a claim that Planck used synthetic differential geometry. It is an exact categorical model of the first-order calculation that exposes why the repair has the required local behavior. More general Weil algebras need not discard every higher-order term: their nilpotent structure supplies a hierarchy of finite jet probes with which a discovery system could determine how much local information the evidence supports.

#### Design principle

Scientific creativity may begin with an economical repair of an empirical pattern. The repair becomes a physical theory only when a subsequent causal account explains its generators, mechanisms, and interventions.

**Challenge**

Can a system create a reusable mathematical primitive—or even a new type of dynamical system—rather than merely prove or optimize the next object selected for it?

Planck's example thus forms a bridge to Chapter 16, which asks how simulation and intervention can move a proposed relation beyond curve fitting: not merely whether the repaired law predicts held-out observations, but which mechanism generates them and which controlled changes can distinguish that mechanism from its alternatives.

*Further Reading*

Lakatos studies mathematical development through proofs, refutations, and concept revision [Lakatos, 1976]. AM and the BACON family provide classical computational precedents for heuristic concept formation and law discovery [Lenat, 1976, Langley et al., 1987]. AI Feynman and related symbolic-regression systems offer controlled modern benchmarks in which symmetry, separability, and composition support equation recovery [Udrescu and Tegmark, 2020]. Wolfram's cellular-automaton program motivates the use of exact microscopic rules with unexpectedly rich macroscopic behavior [Wolfram, 2002]. These systems recover or explore structure; AGENTIC-MATH asks whether diagnosis, skill construction, and experiment choice compose into a persistent mathematical vocabulary. Rutten supplies DIAL-URL's behavioral language, Sokolova organizes probabilistic system types, and metric coinduction relates behavioral semantics to contractive dynamic-programming operators [Rutten, 2000, Sokolova, 2011, Kozen and Ruozzi, 2009].



## *Simulator-Grounded Theory Extension*

### AGENTIC WITH EXECUTABLE SCIENTIFIC WORLDS

Chapter 15 ended with Planck’s repair of Wien’s radiation law: an empirical pattern was corrected before its deeper physical meaning was understood. Simulator-grounded theory extension makes the next transition explicit: a proposed relation is realized as a mechanism whose consequences can be challenged by controlled change.

A simulator turns a proposed explanation into executable consequences. It permits controlled interventions and held-out regimes while preserving a clear boundary: success is relative to the simulator unless external evidence is added.

Among the book’s admission environments, a simulator provides the most explicit setting for the complete AGENTIC loop. CLIC can select and interpret controlled interventions; OPTIC can construct the executable probe or analysis workflow; and RELIC can decide which experiment to run next under cost, risk, and time constraints. Their common object would be a versioned simulator-facing theory whose predictions are executed before any claim is made about nature. The experiments reported here test successive component obligations in a ladder called Simulator-Grounded Theory Extension (SGTE), through SGTE-4; the integrated AGENTIC-SIM benchmark remains specified rather than completed.

The corresponding artifact dossier contains the scientific theory, its explicit realization in the simulator interface, transported prior models and measurements, held-out intervention traces, and the exact boundary between simulator-relative and empirical evidence records. SGTE-1 through SGTE-4 populate parts of this dossier under registered simulators. They do not yet construct and admit the complete integrated object, nor transport its interpretation to nature.

The strongest reported construction is SGTE-4: a mediator-shaped latent factor admitted as a predictive factorization inside a supplied one-factor grammar, with closed-world specificity and unsupported-world abstention. Calling the object *mediator-shaped* is deliberate. Its simulator-

relative certificate does not turn its two arrows into causal mechanisms. This boundary distinguishes executable theory construction from the formal construction of Chapter 15 and the source-relative proposals of Chapter 17.

### 16.1 The AGENTIC simulator contract

The base meta-sketch declares simulator state, legal interventions, observations, costs, and reset semantics. It also separates the simulator's private implementation from the public scientific interface. A component may query the latter but may not inspect source code, hidden parameters, or random seeds unless a specific oracle control permits it.

|       | Simulator role                                                 | Constructed object                                                      | Evidence obligation                                        |
|-------|----------------------------------------------------------------|-------------------------------------------------------------------------|------------------------------------------------------------|
| CLIC  | Diagnose mechanisms and select counter-witness interventions   | Causal variable, mechanism, factorization, or discriminating experiment | Held-out intervention prediction and explicit alternatives |
| OPTIC | Compile scientific requests into executable procedures         | Probe, measurement, intervention macro, or analysis pipeline            | Type, execution, safety, and measurement-semantics tests   |
| RELIC | Allocate sequential experiments and deploy admitted operations | Experiment policy, stopping rule, or regime-conditioned decision schema | Utility and calibration under matched cost                 |

An interface admission record checks that the intervention CLIC requested is the one OPTIC implemented and the one RELIC evaluated. This sounds administrative, but it is scientifically substantive: an apparently successful experiment may measure a proxy, perturb several mechanisms, or induce a regime outside the causal proposal's scope.

The simulator package  $\mathfrak{R}_{\text{sim}}$  contains the scientific sketch, executable interface and realization, intervention doctrine, abstraction maps, observers, admission rules, and proposal grammar. A revision is therefore a package comparison

$$Y_{\text{sim}} : \mathfrak{R}_{\text{sim}} \longrightarrow \mathfrak{R}_{\text{sim}}^+$$

whose presentation and executable components may be the maps  $J$  and  $K$  below. The applicable controls  $\text{Ctl}_{\text{sim}}$  include the current or no-change theory, richer parameter fitting, registered fixed-theory factorizations, alternative latent dimensions or mechanism families, measurement and interface misspecification, matched-random interventions, unsupported worlds, and separate simulator oracles. An extension branch opens

only after its registered alternatives fail; simulator access does not make that nonrealizability claim automatic.

## 16.2 Executable semantics

The simulator defines state variables, mechanisms, interventions, and observables. The theory need not reproduce its implementation, but it must make claims whose consequences can be queried against it.

The public interface should support factual replay, local interventions, compound procedures, and held-out regimes. A proposed theory is represented as a sketch with explicit abstraction maps from simulator states to scientific objects. OPTIC skills implement these maps or interventions; RELIC policies choose their use. CLIC maintains alternative mechanism families rather than forcing every residual into one registered explanation.

### 16.2.1 From relations to mechanisms

Suppose observations support a relation between  $X$  and  $Y$ . Prediction, intervention, and explanation remain distinct achievements:

- prediction:  $P(Y \mid X)$  is stable on held-out observations,
- intervention:  $P(Y \mid \text{do}(X = x))$  is predicted after  
a declared mechanism change,
- explanation: one typed mechanism survives discriminating tests.

The first does not imply the second, and the second does not by itself establish the third [Pearl, 2009]. A simulator is useful because it makes the gap executable.

Let  $S$  be a scientific sketch and  $\mathcal{E}$  the typed category of simulator procedures. A realization

$$R : S \longrightarrow \mathcal{E}$$

maps scientific objects to state, parameter, or observable types and maps mechanisms to executable procedures with matching interfaces. An equation  $p = q$  in a strict sketch model must satisfy  $R(p) = R(q)$ . Before admission, an observer may measure a residual between the two executable paths; a nontrivial comparison cell is permitted only when the realization has explicitly been declared lax or approximate. If a proposed arrow has no realization, or a declared equation fails its registered observer, the defect may lie in the theory, the simulator interface, the realization doctrine, or their interpretation.

**Design principle**

A simulator grounds a theory by making its consequences executable. It does not turn executable consequences into facts about nature.

*16.3 Three validity layers*

Simulator-grounded claims require three validity obligations that must not be collapsed:

1. **Formal validity:** the theory is typed, coherent, and compatible with its declared laws.
2. **Simulator-relative validity:** its realization reproduces registered behavior and predicts held-out simulator outcomes under declared interventions.
3. **Empirical validity:** independent observations or physical experiments support the interpretation of those mechanisms in the target system.

A passed obligation may contribute a positive certificate to the dossier. The enclosing admission record must also preserve failed and unresolved obligations; executable success cannot silently erase missing empirical validation. The first two can hold when the third is unknown or false. Conversely, an empirical regularity may be real even when the simulator omits its mechanism. Disagreement can therefore localize inside the theory, inside the simulator, or in  $R$ .

A code-level replacement  $\Phi_i \mapsto \tilde{\Phi}_i$  has intervention semantics only when the registered simulator interface treats it as replacing the target mechanism while preserving the stipulated remainder of the system. If that modularity contract is absent, it is merely a controlled code perturbation. Even a valid simulator intervention becomes evidence for a real-world intervention only after the interpretation of  $\Phi_i$  is independently justified.

**Boundary**

The registered rungs below test executable, simulator-relative counterfactuals inside their declared substrates. Where a rung is admitted, its verdict remains relative to that interface and observer. It is not a substitute for intervention on animals, historical societies, or the atmosphere.



### 16.4 Double involution in an executable world

The two DIAL directions have a direct operational meaning. Assimilation keeps  $S$ ,  $\mathcal{E}$ , and  $R$  fixed while estimating parameters, calibrating stochastic mechanisms, or improving predictions. For a parameterized simulator  $\Phi_\theta$ , local probes may take the form

$$v_i(\theta) = \left. \frac{d}{d\epsilon} \right|_{\epsilon=0} O(\Phi_{\theta+\epsilon e_i}).$$

Proto-accommodation probes changes to the maintained simulator package. A candidate whose theory presentation or executable schema changes may contain maps  $J$  and  $K$ :

$$\begin{array}{ccc} S & \xrightarrow{J} & S^+ \\ R \downarrow & & \downarrow R^+ \\ \mathcal{E} & \xrightarrow{K} & \mathcal{E}^+. \end{array}$$

The candidate package change may introduce a latent state, observable, mechanism, intervention surface, or composition law. It is a theory extension only when the presentation or generative language changes. Admission asks whether  $R^+$  exists, whether old evidence transports conservatively, and whether new predictions survive held-out execution; only then is the package change installed as accommodation. In the strict presentation-level case the square must commute. In a lax realization, the dossier instead contains a registered comparison

$$\kappa : K \circ R \Longrightarrow R^+ \circ J$$

and tests its coherence and observational meaning; the dashed arrow alone is not a transport certificate.

Mixed probes compare estimating or intervening before a package change with doing so afterward. Persistent disagreement is the simulator-facing shadow of the DIAL interchange obstruction. In a smooth chart it may be represented by a mixed bracket or curvature term. Crossing to a new ontology is nevertheless a finite sketch extension, not a Lie bracket merely because local probes helped discover it.

### 16.5 The integrated loop

A typical episode begins with a RELIC prediction or control failure that persists beyond estimator uncertainty. CLIC chooses an intervention that separates the leading mechanism from an unsupported alternative. If the intervention is not already executable, OPTIC constructs it from registered primitives or proposes a bounded interface extension.

After execution, CLIC updates the theory, and RELIC tests whether the admitted distinction improves later decisions. The loop stops on admission, rejection, observational equivalence, exhausted budget, or a safety boundary.

The system may succeed partially. A discriminating probe can be admitted even if the mechanism remains unresolved. A causal mechanism can be admitted even if it has no current decision value. A policy improvement does not retroactively certify the explanation that suggested it.

### 16.6 *Three Prometheus executable substrates*

Prometheus contains three instructive scientific packages. They are not interchangeable simulators. One is a dataset-backed neural projection counterfactual, one executes a hydrology model with a documented domain mismatch, and one propagates an optical-property replacement through published radiative-forcing maps. A common architecture must preserve these differences rather than erase them behind the generic word simulation [Mahadevan, 2026f].

The three cases should share an AGENTIC protocol rather than a single ontology. Each package is wrapped in a manifest declaring accessible variables, interventions, costs, stochasticity, and hidden evaluator variables. Discovery and confirmation seeds are disjoint. A skill or workflow learned in one package may transfer only at the level licensed by its types; a common word such as “exposure” does not establish a common mechanism.

| Domain            | Executable substrate                           | Intervention                                                    | Sharp boundary                                              |
|-------------------|------------------------------------------------|-----------------------------------------------------------------|-------------------------------------------------------------|
| Singing mice      | MAPseq matrices for 12 animals and 16 targets  | replace singing-mouse AUD/PAG support by laboratory-mouse means | projection evidence is not a vocal-behavior simulator       |
| Indus Valley      | paper figure data plus a VIC hydrology harness | restore a 13% precipitation deficit and 0.5°C warming           | sample basin, not the paper’s exact transient Indus forcing |
| Airborne plastics | source tables and forcing maps                 | replace colored-particle optics by white/pristine optics        | does not rerun atmospheric transport or radiative transfer  |

### 16.6.1 Neural projection expansion in singing mice

The singing-mice package compares five laboratory mice (*Mus musculus*) with seven Alston's singing mice (*Scotinomys teguina*). Rows of the MAPseq matrices represent barcoded motor-cortical neurons; sixteen columns represent target regions. The source study reports selective expansion toward auditory regions and the periaqueductal gray, targets implicated in vocal behavior [Isko et al., 2024].

The executable counterfactual replaces singing-mouse AUD and PAG support by laboratory-mouse species means. Their combined mean positive-projection fraction changes from 0.2102 to 0.0595, a 71.7% attenuation. This number is grounded in the matrices, but it does not simulate neural dynamics, development, song, or evolution. The bridge from projection expansion to enhanced cortical control of vocalization remains a causal hypothesis.

#### Experiment: Prometheus substrate: singing mice

**Observed layer:** 12 animals, 16 target regions, and per-animal MAPseq summaries.

**Executable edit:** attenuate singing-mouse AUD and PAG support to laboratory-mouse species means.

**Result:** the combined metric falls from 0.2102 to 0.0595, a 71.7% reduction.

**Admitted:** dataset-relative arithmetic and provenance.

**Not admitted:** a corresponding change in song or vocal repertoire.

*Controlled SGTE-1 recovery.* The preceding calculation supplies the projection-support identity. A stricter test hides it. The registered support family contains all unordered pairs of target regions with positive species-dependent projection expansion, while the oracle support {AUD, PAG} is unavailable to the estimator. Dataset M194, containing three laboratory and three singing mice, is used for discovery. Dataset M220, containing two laboratory and four singing mice, is sealed for confirmation. The estimator ranks targets by positive pooled standardized species contrast and admits the top pair.

On M194 the recovered support is exactly {AUD, PAG}. Its summed discovery score exceeds those of nineteen balanced relabelings, giving the smallest possible exact one-sided value,  $p = 1/20 = 0.05$ , for this split. Without any reselection, both effects remain positive in M220. The transported pair ranks second among all  $\binom{16}{2} = 120$  pairs by summed positive standardized contrast, the 99.2nd percentile. Replacing precisely those two M220 coordinates by laboratory-mouse means therefore agrees with the registered oracle counterfactual.

The controls show why typing matters. A random pair has only a

0.83% chance of exact recovery. Ranking unstandardized mean differences selects AUD and OMCc instead, while attributing the change to all fifteen nonconstant targets yields support  $F_1 = 0.235$ . The result thus admits a sparse, dataset-relative projection-support signature across acquisition batches. It does not turn that signature into a causal account of song.

#### Experiment: SGTE-1: withheld singing-mouse support

**Discovery:** M<sub>194</sub> only; the two-target identity is hidden.

**Construction:** positive standardized contrasts recover AUD+PAG.

**Counter-witness:** exact balanced-label permutation,  $p = 0.05$ .

**Confirmation:** rank 2 of 120 in sealed M<sub>220</sub>; both effects positive.

**Baselines:** raw differences recover AUD+OMCc; random exact recovery 0.83%; global-support  $F_1 = 0.235$ .

**Verdict:** admitted as SGTE-1 projection-support recovery, not as a causal theory of vocal behavior.

*SGTE-2: constructing the observer.* SGTE-1 supplies positive-neuron prevalence as the measurement. SGTE-2 additionally hides that observer as well as the AUD+PAG support. The learner receives the binary, raw-count, and normalized-count matrices and nine registered summary constructions: targetwise means, totals, and compositional shares for each matrix type. An observer certificate requires a per-animal, targetwise marginal in  $[0, 1]$  that is invariant to row replication and to changes in positive barcode amplitude. Binary prevalence is the unique candidate with that type. In particular, a compositional share is rejected because one neuron may project to several targets; the target coordinates are not exclusive outcomes.

After constructing this observer, the support estimator again recovers AUD+PAG on M<sub>194</sub> and transports it at rank 2 of 120 on M<sub>220</sub>. The uncertainty is material. Across the nine leave-one-laboratory-mouse/leave-one-singing-mouse jackknives, the exact pair persists in 66.7% of replicates; AUD and PAG are included in 77.8% and 88.9%, respectively. The discovery permutation value remains  $p = 0.05$ .

The observer ablation is decisive in this study. If the system ignores the typed certificate and simply chooses the summary with the largest discovery score, it selects normalized-count means and proposes HY+OMCi, not AUD+PAG. Across the twenty balanced M<sub>194</sub> label assignments, only the true assignment passes the registered permutation and stability gates: attempt coverage is 5%, selective accuracy is 100%, and abstention on the nineteen false label worlds is 100%. These percentages describe this finite counter-witness family only; they are not a general coverage guarantee.

**Experiment: SGTE–2: observer and support recovery**

**Hidden:** the projection observer and the two-target support signature.

**Observer construction:** typed invariance selects binary prevalence.

**Support signature:** AUD+PAG; M220 confirmation rank 2 of 120.

**Uncertainty:** exact-pair jackknife stability 66.7%; AUD/PAG inclusion 77.8%/88.9%.

**Selectivity:** 5% attempt coverage, 100% selective accuracy, and 100% abstention on nineteen registered false-label worlds.

**Ablation:** score-only observer selection yields normalized means and the incorrect HY+OMCi support.

**Verdict:** admitted as a calibrated SGTE–2 result with a narrow counter-witness family.

*SGTE–3: active counter-witness acquisition.* SGTE–3 asks not only what should be measured, but which measurement should be acquired next. It uses a controlled world with the same sixteen-target geometry as the singing-mouse example, while deliberately replacing the biological data-generating process by a registered Gaussian assay model. The hypothesis family contains a null theory and all  $\binom{16}{2} = 120$  sparse two-target mechanisms. An assay has mean two when its target belongs to the hidden pair and mean zero otherwise, with unit variance. The learner stops when one posterior probability reaches 0.9, or after thirty-two assays.

The active policy selects the target on which the live posterior most strongly disagrees about support membership, discounted by its accumulated replicate count. This is a posterior counter-witness rule: it spends the next assay where the surviving theories make different predictions. We compare it with random selection with replacement and a fixed round-robin schedule, first at the full budget and then at a matched budget of twenty-one assays.

Across 6,000 pair worlds, the active policy attains 86.9% unconditional exact recovery, 91.6% decision coverage, and 94.8% selective accuracy while using 20.75 assays on average. At the full budget, random and fixed acquisition reach only 19.0% and 33.5% power while using 30.42 and 29.92 assays. At the matched twenty-one-assay budget, their powers fall to 5.7% and 6.5%; the active gain over the stronger passive baseline is therefore 80.3 percentage points. On 1,000 null worlds, the active policy has 98.4% null safety and a 1.6% false-extension rate.

This result isolates the value of active acquisition, but it also marks a firm boundary. The likelihood, sparse mechanism family, and posterior acquisition rule are supplied, and an assay reads a coordinate of a synthetic world. SGTE–3 has therefore selected where to measure

under that rule; it has not learned the acquisition language, performed an intervention on an animal, or established a causal effect on vocal behavior.

**Experiment: SGTE-3: active counter-witness acquisition**

**World:** null plus 120 two-target mechanisms over sixteen targets.

**Acquisition:** posterior support disagreement; threshold 0.9; maximum budget 32.

**Active result:** 86.9% power, 94.8% selective accuracy, and 20.75 mean assays.

**Matched-cost baselines:** random 5.7% power; fixed 6.5% power.

**Null control:** 98.4% safety; 1.6% false extension.

**Verdict:** admitted as controlled active measurement selection, not as biological intervention.

### 16.6.2 *Hydroclimate and the Harappan transformation*

The Indus study combines paleoclimate archives, transient climate simulations, and hydrological reconstruction. It argues that persistent drought likely reduced freshwater availability and may have contributed to population dispersal, while explicitly retaining climatic, social, and economic alternatives [Solanki et al., 2025].

The local harness applies D3-style forcing to VIC: precipitation is scaled by 0.87 and temperature is increased by 0.5°C. Restoring both changes raises the sixteen-cell sample's water-availability proxy from 27.924 to 31.601 mm, or 13.17%. The source paper's Figure 4 data independently contain a mean D3 discharge anomaly of -8.49% across eighteen stations.

The repository does not contain the authors' exact Indus transient-climate forcing grids. This comparison supports the transparent counterfactual adapter's direction of response, but is not a reproduction of the published basin maps or a causal demonstration of societal collapse.

**Experiment: Prometheus substrate: Indus hydrology**

**Executable edit:** compare a 13% precipitation deficit plus 0.5°C warming with restored forcing in VIC.

**Result:** the sample-basin water-availability proxy increases by 3.677 mm, or 13.17%.

**Admitted:** execution and provenance of the local hydrological counterfactual.

**Not admitted:** reproduction of the paper's basin maps or a single-cause theory of Harappan transformation.

### 16.6.3 Colored airborne plastics and radiative forcing

The airborne-plastics package uses source tables and maps from a study of microplastic and nanoplastic radiative forcing [Liu et al., 2026]. The registered edit holds the spatial particle distribution fixed while replacing colored-particle optical forcing by the white/pristine value. The area-weighted mean changes from 0.03914 to 0.00368 W m<sup>-2</sup>, suppressing 90.6% of the baseline source-table forcing.

This is a clean factorized mechanism replacement: burden is fixed while optical properties change. It is not a complete atmospheric counterfactual. The adapter does not rerun FLEXPART or SBDART and does not model feedback from an actual pollution intervention.

#### Experiment: Prometheus substrate: airborne plastics

**Executable edit:** replace colored-particle optics by white/pristine optics while holding spatial burden fixed.

**Result:** area-weighted forcing falls from 0.03914 to 0.00368 W m<sup>-2</sup>, a 90.6% suppression.

**Admitted:** the factorized source-table counterfactual.

**Not admitted:** a complete atmospheric or empirical intervention.

### 16.7 A worked theory-repair trace

The microplastics case makes the audit record concrete. A base sketch contains particle burden  $B$ , optical class  $C$ , absorption  $A$ , spatial field  $M$ , and direct radiative forcing  $F$ , with arrows

$$C \longrightarrow A, \quad (B, A) \longrightarrow M, \quad M \longrightarrow F.$$

The baseline realization compiles these arrows to source-table lookup, map scaling, and area-weighted aggregation.

CLIC localizes the contrast to  $C \rightarrow A$  rather than the burden map. OPTIC compiles the colored-to-pristine replacement and checks units. The modified optical field is propagated through the unchanged map and aggregation arrows. Admission certifies the 90.6% reduction only relative to this realization. Because the factorization and edit were supplied, this is executable theory repair rather than autonomous theory construction. A creative successor must infer that optical class is the missing discriminant, construct the factorization, and request its counter-witness without being given the arrow.

### 16.8 Ontology extension

A transformational episode may require a new latent variable or intervention surface not exposed by the original interface. Such proposals

must distinguish extending a scientific theory from merely reverse-engineering simulator code.

Ontology extension is a joint construction. CLIC proposes why a latent object would resolve intervention failures. OPTIC supplies an estimator or probe that makes the object operational. RELIC tests whether conditioning a decision on it changes admitted behavior. The proposed object is admitted only if these roles agree on its scope and held-out consequences.

*SGTE-4: constructing a mediator-shaped latent factor.* SGTE-4 tests this requirement in a controlled sketch-extension world. The base sketch has twelve intervention objects, ten outcome objects, and sixty measured direct edges sampled from the 120 possible intervention–outcome responses. It has no object through which unmeasured responses can factor. Thirty further edges are sealed for admission and thirty for confirmation. All measurements contain independent Gaussian noise with standard deviation 0.15.

The registered extension language adds one object  $Z$  and arrows

$$I_i \longrightarrow Z \longrightarrow O_j, \quad M_{ij} = a_i b_j.$$

The vectors  $a$  and  $b$  are constructed from discovery edges by alternating least squares; no candidate factor is enumerated in advance. This is an algebraic factorization semantics: the experiment does not establish that  $Z$  is a causal mediator, nor that either displayed arrow denotes a causal mechanism. Three world families separate construction from habitual extension. Closed worlds contain no systematic residual and should retain the old sketch. Positive worlds have one withheld rank-one factorization. Unsupported worlds have two independent components and therefore require at least two factors, outside the registered language.

The registered nonrealizability gate is correspondingly narrow. On sealed admission edges, the old zero-residual sketch and a capacity-matched additive row-plus-column model must both fail by the preregistered margin before the one-factor extension is opened. This rules out the registered fixed-theory repairs; it does not prove that every possible model of the old ontology is inadequate.

Across 200 worlds from each family, the learner makes the correct structural decision in 98.7% of the 600 trials. It constructs  $Z$  in 96.0% of positive worlds, chooses no change in every closed world, and abstains in every rank-two world. The false-extension rate on closed worlds is zero. On sealed confirmation edges, the admitted factor improves prediction by a median 78.1% over the stronger of the zero-residual base sketch and a capacity-matched additive row-plus-column model. The recovered intervention and outcome factors have median absolute cosines 0.9968 and 0.9974 with the withheld factors, up to the



unavoidable sign and scale gauge.

The certificates distinguish three claims. Syntactic conservativity is exact because the new response branch is added without changing any old object, arrow, equation, or base prediction; this is not a claim of causal conservativity. Irreducibility is relative to the registered language: deleting  $Z$  or replacing it by the capacity-matched additive model loses the sealed prediction gain, while a single  $Z$  is rejected when the world requires two independent factors. Predictive value is measured only on edges unavailable during both construction and admission.

#### Experiment: SGTE-4: latent-factor construction

**Withheld structure:** one object  $Z$  and a rank-one factorization  $I \rightarrow Z \rightarrow O$ .

**Data:** 60 discovery, 30 admission, and 30 confirmation edges; 600 worlds.

**Structural decisions:** 98.7% overall accuracy; 96.0% construction; 100% closed specificity; 100% unsupported abstention.

**Confirmation:** 78.1% median gain over the stronger fixed baseline; factor cosines 0.9968/0.9974.

**Certificates:** exact conservativity, zero closed-world false extension, and rank-two rejection.

**Verdict:** admitted as a mediator-shaped factorization inside a supplied one-factor linear grammar, not as a discovered causal mediator or unrestricted ontology discovery.

### 16.9 *An evidence hierarchy*

The Prometheus substrates and controlled SGTE experiments reach finite ontology construction inside a supplied grammar. They do not demonstrate unrestricted autonomous ontology construction. The following registered hierarchy separates the evidence already obtained from progressively stronger construction claims:

| Stage  | Withheld structure                                           | Construction                                      | Primary test                                        |
|--------|--------------------------------------------------------------|---------------------------------------------------|-----------------------------------------------------|
| SGTE-0 | nothing                                                      | reproduce baseline and adapter                    | provenance, units, resets, and reference outputs    |
| SGTE-1 | one support pattern or mechanism surrogate in a known family | support signature or calibrated equivalence class | held-out measurements and false-extension controls  |
| SGTE-2 | mechanism and observer                                       | uncertainty-aware probe                           | coverage, selective accuracy, and abstention        |
| SGTE-3 | discriminating experiment                                    | active measurement or intervention policy         | power and time to decision at matched cost          |
| SGTE-4 | ontology object or factorization                             | sketch and interface extension                    | conservativity, irreducibility, and prediction gain |
| SGTE-5 | domain presentation                                          | transport to a new package                        | typed reuse without semantic leakage                |

SGTE-1 through -4 are separate controlled studies, not cumulative stages of one run. In particular, the synthetic acquisition result in SGTE-3 does not inherit biological semantics from SGTE-1 or -2, and the mediator-shaped latent factor in SGTE-4 does not inherit their observer or acquire a causal interpretation. The evidence reported here stops at SGTE-4; SGTE-5 and AGENTIC-SIM remain benchmark specifications.

Every stage must include closed worlds in which no extension is needed and unsupported worlds outside the registered repair language. A system that always adds a latent variable is uncalibrated. A system that predicts outputs without constructing the withheld typed object has solved a modeling task, not the registered theory-construction task.

#### 16.10 AGENTIC-SIM: the integrated benchmark

AGENTIC-SIM-0 hides a regime-dependent mechanism, the compound probe needed to distinguish it, and the decision schema that exploits the recovered regime. A base policy can assimilate within either regime but aliases them under partial observation. CLIC can express the competing mechanisms but lacks a decisive measurement; OPTIC receives primitive simulator calls but not the required macro; RELIC observes persistent decision failure but not the latent label.

The registered test asks whether the full loop constructs the probe, recovers the mechanism or calibrated equivalence class, and improves decisions on held-out transitions. The full  $2^3$  component factorial is supplemented by fixed-order and guarded-routing AGENTIC, an admission-record-free integration ablation, and separate mechanism,

skill, and regime oracles. Results report mechanism coverage, abstention on unsupported alternatives, skill execution validity, decision value, intervention cost, transport, and regression.

#### Experiment: AGENTIC-SIM-0

**World:** partially observed simulator with a hidden switching mechanism and finite intervention budget.

**Joint construction:** causal distinction, compound measurement or intervention skill, and regime-conditioned experiment or control schema.

**Admission:** held-out interventional prediction, execution and safety tests, decision improvement, and transport to changed switching rates.

**Controls:** component factorial, fixed versus guarded order, admission-record ablation, flat learner, and separate oracle ceilings.

**Boundary:** this is a benchmark specification. The three Prometheus substrates establish SGTE-o-style groundwork, not a completed AGENTIC-SIM result.

#### Boundary

A simulator can falsify claims about its own executable world. It cannot by itself certify that the simulated mechanisms are true of nature.

### 16.11 *What would count as scientific creativity?*

Running a supplied simulator is not creative. Fitting its parameters is assimilation. Selecting a known intervention more efficiently is useful active learning. Simulator-grounded creativity begins when the system constructs a reusable object withheld from its declaration: a latent mechanism unifying several failures, a measurement making a distinction operational, a factorization separating conflated effects, an experiment macro exposing a new causal contrast, or an ontology extension that transports across regimes.

The artifact must be more than a private latent vector. It must be nameable, typed, executable, and teachable. Its value lies in the family of consequences it makes available, not merely in reducing one loss.

**Design principle**

Pattern discovery proposes what might be true. Simulator-grounded causal discovery asks what would happen under a controlled change. Scientific creativity begins when failures of those changes force the construction of a better mechanism or a better experimental language.

**16.12 Open problems**

*Simulator adequacy.* A system may learn a perfect theory of a flawed simulator. Multi-simulator comparison, empirical residuals, and explicit discrepancy models are needed.

*Intervention semantics.* Many code edits change several mechanisms simultaneously. Typed mechanism-replacement interfaces are needed if an action is to support a causal interpretation.

*Structural observability.* A missing mechanism may have no public observer. The system must distinguish nonidentifiability from absence and decide whether a new sensor or derived measurement is admissible.

*Scale separation.* Infinitesimal probes are local, while drought, evolutionary change, and atmospheric transport are multiscale. Higher-order Weil profiles and finite sketch extensions must be connected without pretending every change is infinitesimal.

*Transport to nature.* The strongest simulator-relative result remains a hypothesis until supported outside the executable world. A mature system must produce a feasible empirical admission plan with discriminating outcomes, uncertainty, cost, and ethical constraints.

**Challenge**

Can an agent use a simulator not merely to search a supplied model, but to construct a new causal distinction that remains useful when the simulator, regime, or real experiment changes?

The simulator supplies interventions and executable counterfactuals, but real scientific programs are also assembled from papers, reports, datasets, and partially conflicting claims. Chapter 17 removes the privileged executable world and asks how a document collection can be treated as an implicit, incomplete presentation of a theory whose proposed extensions must ultimately reconnect to experiment.

### *Further Reading*

The Robot Scientist tradition demonstrates closed-loop generation and testing of biological hypotheses with automated experimentation [King et al., 2009]. PROMETHEUS supplies the text–data–model research architecture and simulator packages motivating this testbed [Mahadevan, 2026f]. The three substrates are described by the singing-mouse projection, Harappan hydroclimate, and airborne-plastics studies [Isko et al., 2024, Solanki et al., 2025, Liu et al., 2026]. Pearl develops the distinction between observation, intervention, and counterfactual reasoning [Pearl, 2009]. Surveys and benchmarks of LLM-assisted scientific discovery provide contemporary comparison points [Zheng et al., 2025, Song et al., 2025]. AGENTIC-SIM differs by freezing the causal, skill, and decision reference spaces separately and asking whether their admitted extensions compose under intervention, cost, and transport.



## From Scientific Documents to Testable Theories

### CORPUS-RELATIVE CONSTRUCTION AND EXTERNAL ADMISSION

Chapter 16 began with an executable world. A simulator exposed state variables, accepted interventions, and returned consequences against which a proposed theory could be tested. This chapter removes that privilege. Its input is a collection of scientific reports: papers, tables, databases, and extracted claims written by investigators who performed different measurements under different conditions.

Consider a hypothetical literature in which one set of studies reports that an intervention increases an outcome in one tissue, another reports a decrease in a second tissue, and a third proposes a mediator. The scientific problem is not to average the signs. It is to decide whether the reports describe a real context dependence, different populations or measurements, an omitted mechanism, or simply an extraction error. The desired output is therefore a *testable theory proposal*: a typed account of the entities and contexts, an exact ledger of supporting and opposing sources, consequences not used to construct the account, and an experiment capable of distinguishing it from its alternatives.

This changes the meaning of success. Documents can support reconstruction, comparison, and hypothesis generation, but they ordinarily cannot execute a new intervention. A corpus-relative result may show that a proposed structure explains several reports and predicts a sealed report. It does not by itself show that the structure is causally true of nature.

The reported evidence reaches typed transport in generated document worlds, bounded context construction in a curated GLP-1 interface, and conservative realization of a typed research program. It does not reach prospective biomedical admission, independent peer-review success, or autonomous grant submission. The long realization sequence later in the chapter audits how a source-bearing theory can cross into human-readable prose without acquiring unsupported claims; it is not an exercise in optimizing style.

*Chapter roadmap.* Corpus-to-Theory Extension (CTTE) compiles controlled reports into source-bearing theories under increasingly hidden registration. GLP1-PILOT tests that spine in a scoped biomedical setting; GLP1-GRANT adds an auditable prospective program. These are progressively stronger obligations, not independent demonstrations of autonomous science.

### 17.1 *The transition from simulation to literature*

The two testbeds address adjacent parts of a scientific loop:

|                            | <b>Simulator-grounded construction</b>       | <b>Document-grounded construction</b>                    |
|----------------------------|----------------------------------------------|----------------------------------------------------------|
| Available object           | executable mechanism                         | source-bearing claim                                     |
| Permitted action           | intervene and run                            | retrieve, compare, and derive                            |
| Primary obstruction        | failed trajectory or intervention prediction | incompatible contextual reports or missing factorization |
| Constructed artifact       | executable mechanism or observer             | typed theory proposal and discriminating experiment      |
| Strongest internal verdict | simulator-relative admission                 | corpus-relative support                                  |
| Required external step     | transport to nature                          | execute an experiment or open independent evidence       |

Chapter 16 asks whether a structural proposal survives action in an executable world. This chapter asks whether heterogeneous reports can be converted into a precise proposal worth testing there. In a complete discovery cycle, document-grounded construction proposes a mechanism and a counter-witness experiment; a simulator, laboratory, or clinical study then tests it; the resulting evidence returns to the source-bearing theory store.

The chapter has two goals. The first is representational: preserve who claimed what, in which context, with which evidence, without collapsing disagreement. The second is constructive: determine whether a small extension of the current theory derives new, locked consequences and identifies a decisive external test. The second goal depends on the first, but must not be confused with causal identification.

The Theory Extension Package introduced below is this chapter's realization of the artifact dossier from Section 10.4.1. It binds the proposed theory map to source-level provenance, transported evidence, locked corpus tests, and an executable external-admission interface. A corpus can support the package and reject alternatives; it cannot fill the external admission field merely by describing the experiment that would be needed. The name is reserved for proposals that actually change the presentation or generative language. A revised observer, retrieval policy, evidence workflow, or admission interface without such a map is recorded as a package revision, not promoted to theory extension by terminology alone.

The CTTE ladder progressively removes supplied entity, context, query, and domain structure while retaining source provenance and



conservative abstention.

### 17.2 *Two directions between theory and evidence*

Two famous episodes show why the output of document-grounded construction should be an *admission interface*, not merely another document. The interface specifies how a claim could meet the world: an observation protocol, a controlled intervention, a simulation, or a manipulable physical model.

*General relativity: theory constructs a measurement.* Einstein presented the field equations of general relativity in 1915 and a systematic exposition the following year [Einstein, 1916]. The theory implied a quantitative displacement of apparent stellar positions near the Sun. A total solar eclipse made the otherwise invisible comparison observable by blocking the Sun's glare. During the eclipse of 29 May 1919, British expeditions collected photographic plates at Príncipe, off the west coast of Africa, and Sobral, Brazil. Eddington and Cottingham observed at Príncipe; a parallel team observed at Sobral. The reported analysis agreed with Einstein's predicted light deflection and was announced publicly that November [Dyson et al., 1920].

The methodological point is independent of later debates about the precision of those plates. A theoretical construction generated a new observable, the conditions under which it could be measured, and outcomes capable of distinguishing it from a competing theory. In the terminology of this chapter, the theory supplied both a consequence and an acquisition policy:

general-relativistic theory  $\longrightarrow$  predicted stellar displacement  $\longrightarrow$   
eclipse measurement  $\longrightarrow$  evidential update.

*DNA: evidence constructs a manipulable theory.* The route to the double helix ran in the other direction. Chemical knowledge, Chargaff's base ratios, and X-ray diffraction work by Rosalind Franklin, Raymond Gosling, Maurice Wilkins, and their colleagues constrained the possible geometry. Pauling and Corey proposed a three-chain structure with phosphates near the axis, a model Watson and Crick criticized on chemical and geometric grounds [Pauling and Corey, 1953, Watson and Crick, 1953]. Watson and Crick then used model building—including movable representations of molecular components—to search configurations satisfying bond angles, distances, helical symmetry, and the empirical constraints. Complementary base pairing made the two-chain structure cohere and immediately suggested a copying mechanism [Watson, 1968, Watson and Crick, 1953]. The result depended on the experimental

contributions of Franklin, Gosling, Wilkins, and others; the physical model was a device for composing those constraints, not a substitute for them [Franklin and Gosling, 1953].

This model was not a dynamical simulator in the sense of Chapter 16. It was an executable *structural* theory: components could be moved, rejected when they violated geometry or chemistry, and retained when many constraints closed simultaneously. Here the direction was

measurements and chemical constraints  $\longrightarrow$  manipulable model  $\longrightarrow$   
double-helix theory  $\longrightarrow$  new biological consequences.

These examples refine the chapter's constructive goal. Given a scientific corpus, the system should not merely extract its causal claims. It should construct the smallest source-bearing theory that organizes them and then compile that theory into an external test. Depending on the domain, the test may be an Eddington-style measurement under rare conditions, a controlled intervention, a numerical simulation, or a Watson–Crick-style model whose components realize the proposed structure. When no such interface can be specified, the system has produced an interpretation, not yet a testable scientific theory.

### 17.3 Documents as implicit sketch presentations

The categorical form of this task is a document-to-sketch construction. A scientific document collection  $D$  rarely declares a category, a path equation, or a limit cone. It nevertheless presents them implicitly. Noun phrases propose objects; perturbations and measurements propose arrows; experimental conditions type local contexts; statements of invariance, mediation, or equivalence propose equations and factorizations. Constructing a theory means making this latent presentation explicit.

The passage is staged rather than magical:

$$D \xrightarrow{E_\theta} P_D \xrightarrow{R} S_D \xrightarrow{i} S_D^+ \xrightarrow{\text{Adm}} \text{Test}_D.$$

Here  $E_\theta$  is a versioned extractor, parameterized by an ontology and language model, that converts source spans into a provenance-bearing presentation  $P_D$ . The presentation contains candidate objects, arrows, contexts, equations, evidence types, and links back to the exact source spans. The reconciliation map  $R$  resolves aliases and incompatible contexts to form a source-bearing sketch  $S_D$ . A model of the extracted theory in a semantic category  $\mathcal{C}$  is then a sketch model  $M_D : S_D \longrightarrow \mathcal{C}$  that realizes the declared cones and equations. When the local presentations cannot be glued, the obstruction may justify an extension  $i : S_D \rightarrow S_D^+$ : a new object, context split, factorization, equation, or experimental interface. Finally  $\text{Adm}$  compiles distinguishing

consequences into an admission interface  $\text{Test}_D$ —a laboratory protocol, simulator, observation campaign, physical model, or locked evidence query.

This notation should not be read as asserting a unique functor from prose to theory. Extraction and reconciliation can be ambiguous. More faithfully, a compiler returns a family  $\text{Compile}_\theta(D) = \{(S_k, w_k, \pi_k)\}_{k \in K}$ , where  $S_k$  is a candidate sketch,  $w_k$  records its support or uncertainty, and  $\pi_k$  maps every generator and relation to its provenance. A corpus inclusion  $D \subseteq D'$  should induce a conservative sketch map only when the new evidence preserves the old typing and identifications. If a new source forces two objects to split, two aliases to merge, or an equation to be withdrawn, theory revision is not a monotone database insertion; the old and new sketches require an explicit versioned transport.

This is also where double involution enters. *Assimilation* updates the evidence and models of a fixed  $S_D$ . *Proto-accommodation* proposes a changed presentation; only an independent admission process may admit  $S_D^+$ . Their interaction asks whether revising the interpretation of the documents and revising the theory commute. A persistent failure of the two orders to agree localizes the creative frontier: the documents cannot be coherently represented inside the current theory and extraction language.

CTTE-0 begins after the first arrow. Its input is already a parsed provenance-bearing presentation  $P_D$ , and it tests whether contextual reconciliation and finite extension derive locked consequences. CTTE-1 and CTTE-2 progressively move leftward by hiding entity registration and context boundaries. The full GLP-1 challenge must eventually evaluate the entire chain from paper-length documents through an externally executable admission interface.

#### 17.4 The AGENTIC corpus contract

The meta-sketch links three versioned stores: a source-bearing scientific theory, an executable document and data workflow, and an acquisition policy over retrieval, extraction, comparison, and requests for external evidence.

|       | Corpus role                                                                            | Constructed object                                              | Permitted status                                       |
|-------|----------------------------------------------------------------------------------------|-----------------------------------------------------------------|--------------------------------------------------------|
| CLIC  | Organize observational and interventional claims, alternatives, and missing mechanisms | Causal sketch extension or discriminating experimental question | Proposed unless identification evidence is supplied    |
| OPTIC | Build typed retrieval, extraction, normalization, and reconciliation procedures        | Source-to-model skill, mediator schema, or audit workflow       | Admitted for execution semantics, not biological truth |
| RELIC | Allocate queries and evidence acquisition under cost                                   | Source-selection policy, stopping rule, or experiment priority  | Admitted for declared utility and calibration          |

The interface admission record tracks every theory element to its source spans, context, extraction version, and transformations. A downstream causal packet must preserve whether its evidence was observational, interventional, simulated, or asserted by an author. Agreement among papers does not upgrade the evidence type.

Here  $\mathfrak{R}_{\text{corpus}}$  contains the source and extraction grammar, provenance schema, contextual theory presentation, query and acquisition policy, evidence-type doctrine, observers, and admission rules. The comparison

$$Y_{\text{corpus}} : \mathfrak{R}_{\text{corpus}} \longrightarrow \mathfrak{R}_{\text{corpus}}^+$$

records changes to any of them; a map between scientific sketches is only its presentation component. The registered controls  $\text{Ctl}_{\text{corpus}}$  include no change, entity aliasing, context splitting, extraction and retrieval error, missing overlap or measurement, source dependence, fixed-grammar reconciliation, unsupported-source worlds, and sealed-report or external-evidence checks. A fluent synthesis is not a control against these alternatives because it can reproduce the same hidden source failure.

#### 17.4.1 *How to read the evidence sequence*

The chapter uses three nested scales of evidence. They answer different questions and should not be read as successive scores on one benchmark.

| Scale                       | Scientific material                                  | Question isolated                                                                                | Strongest permitted conclusion                                                                            |
|-----------------------------|------------------------------------------------------|--------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------|
| CTTE-0 through -5           | Generated source ledgers with known hidden structure | Can a source-bearing presentation be reconciled, extended, queried, and transported?             | Controlled recovery relative to a frozen document grammar                                                 |
| GLP <sub>1</sub> -PILOT-0/1 | Human-curated GLP-1 trial events                     | Can a missing treatment context be represented or recovered without inventing biological claims? | Retrospective context construction in a bounded semantic interface                                        |
| GLP <sub>1</sub> -GRANT     | A frozen source ledger and typed prospective program | Can an admitted theory package be compiled into an auditable research program?                   | Structural and semantic realization results, not biomedical validity, peer-review quality, or fundability |

The controlled CTTE worlds calibrate the machinery because their missing objects are evaluator-known. The GLP-1 pilots then replace generated claims by curated scientific records while retaining a narrow task. The grant sequence changes the output type from a retrospective theory packet to a prospective research program. Its final negative result is therefore part of the evidence: passing a formal prose proxy can still fail to demonstrate scientific or rhetorical quality.

### 17.5 *From reports to local models*

Democritus-style extraction, categorical databases, sheaf compatibility, infinitesimal causality, and latent-confounded discovery provide components for assembling source-bearing local models into an extensible declaration.

CLIC receives these local models with provenance intact. Sheaf disagreement or Lie-bracket nonclosure localizes a possible gap, but neither certifies a latent mechanism. OPTIC may construct a new normalization, entity-resolution, retrieval, or simulation skill to test whether the obstruction was introduced by the evidence pipeline. RELIC chooses

the next source, database query, model comparison, or requested experiment according to expected discrimination and cost. This loop guards against interpreting extraction error as scientific discovery.

The joint theory store is intentionally plural. Incompatible dosage regimes, populations, measurement definitions, or mechanistic assumptions remain local until a justified transport map is supplied. The purpose of synthesis is not to erase disagreement but to make its location and experimental consequences queryable.

### 17.6 *Why a corpus is not a simulator*

A simulator executes a declared transition. A document reports, interprets, or disputes one. This difference changes what can be admitted. Replacing a mechanism in code may generate a counterfactual trajectory; replacing a phrase in a paper generates only another phrase. Corpus-to-theory construction must therefore separate four operations that a fluent summary can easily collapse:

1. extracting what a source claims;
2. typing the context and evidence supporting that claim;
3. deriving consequences of several claims under a proposed theory; and
4. testing those consequences against evidence not used to construct the theory.

Even the fourth operation is corpus-relative when its evidence is another paper. Publication bias, shared datasets, copied assumptions, and common measurement errors can make nominally independent sources dependent. A locked document can test prediction and provenance without identifying a natural mechanism.

There are also several meanings of absence. A relation may be false, unmeasured, unreported, excluded by the search query, or lost by the extractor. A source-bearing sketch records these possibilities separately. In particular, the absence of an edge is not automatically a negative edge, and a failed global gluing operation is not automatically a latent biological variable.

### 17.7 *The Theory Extension Package*

The principal artifact is not a prose review but a versioned Theory Extension Package. It contains enough structure for another system or investigator to inspect, challenge, and reuse the proposal:

| Layer               | Contents                                                                                                                         | Evidence obligation                                                     |
|---------------------|----------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------|
| Declaration         | objects, processes, contexts, observers, interventions, and equations                                                            | typing and well-formedness                                              |
| Local models        | source-specific graph or database realizations                                                                                   | source span, population, protocol, and evidence type                    |
| Obstruction record  | incompatible signs, failed descent, nonclosure, missing factorizations, and extraction conflicts                                 | replayable localization                                                 |
| Proposed change     | package comparison; added object, arrow, context split, law, or experimental interface; theory map when the presentation changes | conservativity or declared breakage, with change type stated explicitly |
| Consequences        | distinguishing predictions and requested measurements                                                                            | frozen before confirmation                                              |
| Admission interface | observation protocol, intervention, simulator, or manipulable structural model                                                   | executable specification and discriminating outcomes                    |
| Admission ledger    | confirmation, counterevidence, alternatives, and residual uncertainty                                                            | independent source and status audit                                     |

This package makes a crucial distinction between discovering a relation and constructing an explanatory object. A relation extractor may recover all sentences correctly while leaving no theory capable of predicting an unreported context. Conversely, a theory can be compact and predictive while misrepresenting its sources. Corpus-to-theory construction requires both the structural artifact and the evidence ledger.

### 17.8 *Controlled evidence for corpus-to-theory construction*

The CTTE studies are organized by scientific obligation rather than by the order in which they were run. CTTE-o begins with source discipline already represented; later studies withhold entity identity, contextual overlap, the decisive query, a finite presentation, and finally the cross-domain interface. The experiment identifiers are retained as stable links to the archived protocols and artifacts.

### 17.8.1 Source discipline (CTTE-o)

Before confronting paper-length prose, entity aliases, and incomplete search, CTTE-o isolates one question: once source claims have been extracted correctly, can a contextual theory constructor use them to recover a missing factorization and derive a relation that no source in its construction set states directly? It is a unit test of that operation, not a synthetic scientific domain or a benchmark of literature understanding.

The experimental units are 600 independently seeded *claim ledgers*—200 from each of three structural families. A ledger is a documentary projection of a hidden benchmark model. It contains 20 or 60 one-sentence records; each record carries a source identifier, corpus split, named context, evidence-type tag, and one signed perturbational claim, such as

Evidence type: controlled. In context A, perturbing stimulus<sub>37</sub> increased mediator<sub>37</sub>.

The hidden model is the answer key used to generate and score the records; it is never given to the constructor. It specifies a registered intervention  $X$ , outcome  $Y$ , contexts  $A$  through  $D$ , and either no mediator, one shared mediator, or incompatible context-specific mediators. The generator emits five separately identified reports for each relation and reverses each reported sign independently with probability 0.05. It then partitions source IDs into discovery, admission, and confirmation sets.

hidden benchmark model  $\longrightarrow$  noisy source reports  $\longrightarrow$  claim ledger  $\longrightarrow$   
proposed sketch  $\longrightarrow$  sealed-report test

The ledger therefore defines *what documentary evidence is available to the constructor*, not what is true in a physical world. The hidden model defines truth only inside the benchmark. CTTE-o supplies parsing, stable entity names, contexts, and the allowed one-mediator grammar so that failures can be attributed to contextual composition rather than natural-language processing.

The initial ontology contains  $X$  and  $Y$  but not a mediator. Three corpus families test the structural decision. In closed corpora,  $X \rightarrow Y$  is invariant and no extension is needed. In positive corpora, one undeclared object  $Z$  appears in every context: the sign of  $X \rightarrow Z$  changes with context,  $Z \rightarrow Y$  remains invariant, and composition explains the corresponding change in  $X \rightarrow Y$ . In unsupported corpora, each context has a different mediator, so no one-object extension can unify the literature.

The target-context arrows  $X \rightarrow Y$  are absent from discovery and admission. They are opened only after the theory and two target predictions are frozen. The contextual constructor admits the factorization



$X \longrightarrow Z \longrightarrow Y$  only when one source-backed object explains both discovery contexts and its incident arrows survive admission in both target contexts. The proposal retains the local signs rather than averaging them.

We compare four representations. Relation extraction and local graphs cannot predict an unreported target edge. A global graph aggregates the two opposing discovery effects and abstains. The contextual sketch can compose through  $Z$ . These zero-coverage baselines are intentional: CTTE-0 tests whether the locked relation is derivable from a shared contextual theory, not whether a larger predictor can guess its label.

This makes the scope of the result deliberately narrow. CTTE-0 asks whether the proposed representation supports source-preserving derivation, correct no-change decisions, and abstention when its extension language is inadequate. It does not test paper reading, entity resolution, open-ended mediator invention, experimental design, or causal truth. CTTE-1 through -5 isolate selected pieces of those harder capabilities under generated-world controls; paper understanding and causal truth remain outside the ladder and ultimately require the external admission loop described above.

Across 200 corpora from each family, structural action accuracy is 99.7%. The system constructs and exactly recovers  $Z$  in 99.5% of positive worlds, retains the old theory in 99.5% of closed worlds, never falsely extends a closed world, and abstains in every unsupported world. It predicts both sealed target relations in 99.5% of positive worlds. Every admitted theory element retains a valid source identifier, all contextual contradictions are preserved, and no source identifier crosses a corpus split.

| Representation              | Target coverage | Accuracy | Exact-pair power |
|-----------------------------|-----------------|----------|------------------|
| Relation extraction         | 0%              | 0%       | 0%               |
| One global graph            | 0%              | 0%       | 0%               |
| Context-local graphs        | 0%              | 0%       | 0%               |
| Contextual sketch extension | 99.5%           | 99.5%    | 99.5%            |

**Experiment: CTTE-0: source-ledger theory construction**

**Input:** 600 independently seeded synthetic claim ledgers: 200 per family, with 20 or 60 one-claim records and disjoint discovery, admission, and confirmation sources.

**Withheld structure:** one contextual mediator  $Z$  and two target relations.

**Structural result:** 99.7% action accuracy; 99.5% construction and mediator recovery.

**Controls:** 99.5% closed specificity, zero false extension, and 100% unsupported abstention.

**Confirmation:** 99.5% exact two-context prediction with complete provenance and contradiction preservation.

**Verdict:** admitted as a source-preserving derivability unit test with a supplied parser and one-mediator grammar, not as scientific discovery.

**17.8.2 Entity registration (CTTE-1)**

CTTE-0 gave every occurrence of its mediator the same stable name. Scientific literatures rarely offer this convenience. The same mechanism may be named by an assay, gene product, pathway role, historical convention, or local abbreviation; conversely, two reports may use similar functional language for biologically distinct objects. Before a document compiler can add a generator to its sketch, it must therefore decide which documentary occurrences present the same object. Entity registration is part of theory construction, not clerical preprocessing.

CTTE-1 removes the registered mediator name while retaining the supplied claim parser, four contexts, and one-mediator extension grammar. It contains 800 independently seeded ledgers, 200 from each of four families. A *closed* ledger needs no new object. In a *shared-alias* ledger, two different names in each context denote one latent mediator. A *distinct-entity* ledger has the same relational pattern, but the names denote genuinely different mediators. An *ambiguous-identity* ledger contains two close candidates measured too imprecisely to justify either merging or separating them. Each occurrence carries noisy four-dimensional assay profiles in addition to its signed, source-bearing relations. Target relations in contexts  $C$  and  $D$  remain sealed until the identity decision, sketch action, and predictions are frozen.

The typed resolver first constructs an equivalence proposal over aliases. For each pair it places a confidence region around the distance between assay profiles. An upper bound below the registered identity tolerance supports equivalence; a lower bound above it supports distinction; overlap with the boundary produces ABSTAIN-IDENTITY.

Only a supported equivalence class may become the new sketch object Z. Contextual factorization and admission are then performed as in CTTE-o. Thus

documentary names  $\rightarrow$  typed equivalence proposal  $\rightarrow$  registered  
object Z  $\rightarrow$  contextual factorization  $\rightarrow$  sealed test.

This order matters. Exact-string matching fragments the shared object. Matching only by relational role merges the distinct controls. A nearest profile rule can recover easy identities, but converts a near-boundary measurement into an unjustified categorical decision. CTTE-1 instead exposes identity uncertainty as an epistemic status that blocks downstream accommodation.

Under the preregistered gates, CTTE-1 was admitted. It achieved 99.875% four-family action accuracy, constructed the shared mediator and recovered its complete alias equivalence class in every positive ledger, and predicted both sealed target relations in every positive ledger. It retained the closed theory in 99.5% of closed ledgers, made no false merge in any distinct-entity ledger, and returned ABSTAIN-IDENTITY in every ambiguous ledger. The single action error was a conservative failure to retain one closed theory after noisy relation reports obscured its direct invariant; no unsupported extension was admitted. Provenance was complete, source splits were disjoint, and deterministic replay reproduced identical artifacts.

| Resolver               | Four-family action accuracy | Positive exact-pair power |
|------------------------|-----------------------------|---------------------------|
| Exact string           | 49.875%                     | 0%                        |
| Relational role        | 50.0%                       | 100%                      |
| Profile point estimate | 74.875%                     | 100%                      |
| Typed profile resolver | 99.875%                     | 100%                      |

#### Experiment: CTTE-1: alias registration under uncertainty

**Input:** 800 generated claim ledgers across closed, shared-alias, distinct-entity, and ambiguous-identity families.

**Withheld structure:** the equivalence class presenting the candidate mediator, followed by the contextual factorization and two target relations.

**Structural result:** 99.875% action accuracy, 100% shared-alias recovery, and 100% positive exact-pair confirmation.

**Epistemic controls:** zero false merges of distinct entities and 100% identity abstention at the registered uncertainty boundary.

**Verdict:** admitted as a controlled test of  $P_D \rightarrow S_D$  with supplied assay semantics and grammar; not open-domain entity linking, causal identification, or biological truth.

### 17.8.3 Context-cover construction (CTTE-2)

CTTE-2 removes the context labels that CTTE-0 and CTTE-1 supplied. Each source instead carries two anonymous continuous metadata coordinates. The constructor must decide whether the evidence supports no contextual split, one binary cover, or two binary covers whose four intersections define local models. It must also attach every source to the appropriate opens and overlap. The stable entities  $X$ ,  $Z$ , and  $Y$  are supplied so that the experiment isolates contextual construction rather than repeating entity registration.

The preregistered benchmark contains 1,000 ledgers, with 200 from each of five families. Invariant worlds span a clear two-by-two metadata population but do not require a split. Product worlds require both axes; one-axis worlds require only the first and therefore penalize unnecessary refinement. Ambiguous worlds require both axes but measure the second too imprecisely to certify its boundary. In missing-overlap worlds, both axes are individually separated but perfectly correlated in discovery and admission, while confirmation asks for off-diagonal intersections that have never been observed.

Discovery sources report noisy  $X \rightarrow Y$  effects. Admission sources report  $X \rightarrow Z$  and  $Z \rightarrow Y$ , so a proposed cover must support the contextual factorization rather than merely cluster metadata. Confirmation exposes its metadata but seals its effects until the cover, restriction maps, epistemic status, and predictions are frozen. The typed resolver certifies each boundary under uncertainty and selects the least complex cover satisfying both the discovery and factorization tests.

The result was a preregistered near miss, not an admission. Overall action accuracy was 97.9%. The product cover was constructed in 98.5% of product worlds, with 98.5% coverage-adjusted restriction membership and 98.5% confirmation accuracy. The resolver selected the minimal one-axis cover in 98.5% of the corresponding worlds, retained the invariant theory in 99.0%, never falsely refined a one-axis theory into a product cover, and abstained in every ambiguous-boundary world. It also remained within 1.5 percentage points of the context oracle on product confirmation.

One gate nevertheless failed. The registered requirement was at least 95% ABSTAIN-NONIDENTIFIABLE decisions in missing-overlap worlds; the observed rate was 93.5%. In the remaining 13 of 200 worlds, noise caused the mediator-factorization check to fail before the overlap audit, producing the less informative status ABSTAIN-UNSUPPORTED. The system did not make an off-support prediction, but it did not correctly localize why prediction was unwarranted. This distinction matters scientifically: “the proposed mechanism lacks support” and “the requested population was never observed” call for different acquisitions.

| Resolver                     | Five-family action accuracy | Product confirmation |
|------------------------------|-----------------------------|----------------------|
| Global context               | 20.0%                       | 35.0%                |
| One-axis selector            | 78.2%                       | 0%                   |
| Flat four-cluster resolver   | 60.0%                       | 100%                 |
| Typed context-cover resolver | 97.9%                       | 98.5%                |
| Context oracle               | 100%                        | 100%                 |

The flat resolver's perfect product prediction illustrates why prediction alone is inadequate. It always constructs four cells when their metadata clusters are visible, thereby over-refining invariant and one-axis theories. The near miss exposes an ordering defect: overlap support should be audited before a noisy mechanism failure receives its terminal status. That claim belongs in a separately registered repair, not a post-hoc change to CTTE-2.

#### Experiment: CTTE-2: inferred context covers

**Input:** 1,000 generated ledgers across invariant, product, one-axis, ambiguous-boundary, and missing-overlap families.

**Withheld structure:** context boundaries, local opens, four possible intersections, and source restriction maps.

**Structural result:** 97.9% action accuracy, 98.5% product-cover construction and confirmation, and zero false product refinement.

**Failed gate:** missing-overlap localization was 93.5% against a registered 95% minimum; 13 safe abstentions carried the wrong reason.

**Verdict:** not admitted. The failed obligation is epistemic-status ordering; the result does not justify threshold tuning.

#### 17.8.4 Active overlap acquisition (CTTE-3)

CTTE-2 showed that a safe abstention can still carry the wrong scientific reason when mechanism fitting precedes an audit of contextual support. CTTE-3 turns that diagnostic into an active acquisition task. The initial literature contains controlled  $X \rightarrow Y$  reports only from the two diagonal cells of a two-by-two context cover. Four declared theories—invariant, first-axis, second-axis, and product—agree in pairs on those restrictions but make different predictions on either off-diagonal intersection.

The system receives a costed source menu. A controlled report costs two units and has registered reliability 0.95; an observational report costs one unit and has reliability 0.80. The acquisition budget is ten units. The typed policy maintains a posterior over the four theories, audits whether the source pool contains any query that separates its plausible candidates, and otherwise returns ABSTAIN-NO-SUPPORT. When a decisive source exists, it selects the query with greatest expected

entropy reduction per unit cost and admits a theory only at posterior probability 0.99.

The preregistered benchmark contains 1,000 ledgers: 200 generated from each of the four theories and 200 blocked-product controls in which the source pool contains no off-diagonal report. Confirmation seals the complete four-cell sign map. Passive similarity repeatedly retrieves a familiar diagonal source; random retrieval spends the same budget without a structural target; cheap uncertainty samples only observational evidence; and a query oracle selects a maximally discriminating controlled source.

CTTE-3 was admitted on every registered gate. Overall action accuracy was 99.1%, with 98.875% theory recovery across the four available-source families and 100% ABSTAIN-NO-SUPPORT decisions in the blocked family. Coverage-adjusted four-cell confirmation accuracy was 99.125% and wrong-theory admission was 0.5%. The median cost of a correct typed decision was four units, equal to the query oracle. Cost per correct decision was 4.45 units, compared with 13.07 for random retrieval, a 65.98% efficiency gain. Theory recovery exceeded cheap observational uncertainty by 7.375 percentage points and passive similarity by 98.875 points.

The actual traces make the mechanism transparent. In every one of the 800 worlds with an available decisive source, the first typed action requested a controlled report from the same off-diagonal cell. Of these worlds, 724 stopped after two queries, 69 used four, and seven exhausted five queries. Four worlds admitted the wrong member of a theory pair and five abstained under residual uncertainty. All 200 blocked worlds stopped before acquisition with the registered no-support status.

| Policy             | Theory recovery | Mean cost | Cost per correct |
|--------------------|-----------------|-----------|------------------|
| Passive similarity | 0%              | 10.00     | undefined        |
| Random budget      | 62.5%           | 8.17      | 13.07            |
| Cheap uncertainty  | 91.5%           | 6.00      | 6.56             |
| Typed acquisition  | 98.875%         | 4.40      | 4.45             |
| Query oracle       | 98.875%         | 4.40      | 4.45             |

Matching the oracle here should be read as a calibration result. The finite hypothesis family, context coordinates, costs, reliabilities, and query menu are supplied, and the registered geometry makes either off-diagonal controlled query optimal. Within these generated ledgers, CTTE-3 shows that a typed obstruction can be compiled into an efficient acquisition action and that absence of such an action can remain explicit. It does not discover a biological variable, search an open literature, or design an unrestricted experiment.

**Experiment: CTTE-3: active acquisition of a missing overlap**

**Input:** 1,000 generated ledgers with diagonal evidence, a costed source pool, and a sealed four-cell confirmation map.

**Withheld structure:** the off-diagonal evidence distinguishing four contextual theories.

**Result:** 99.1% action accuracy, 98.875% theory recovery, 99.125% confirmation accuracy, and 0.5% wrong admission.

**Efficiency:** median cost four, equal to the query oracle; 65.98% lower cost per correct decision than random retrieval.

**Support control:** every blocked world returned ABSTAIN-NO-SUPPORT without purchasing an irrelevant source.

**Verdict:** admitted as a finite active-acquisition calibration with a supplied theory and query language, not open-domain experimental design.

### 17.8.5 Missing-presentation construction (CTTE-4)

The first four rungs progressively remove supplied documentary structure, but CTTE-3 still chooses among four completed theories. CTTE-4 withholds those domain-level mechanism names. It supplies only a generic finite-sketch grammar: add one mediating object and a factorization equation; add one contextual product and its cone; add one typed exceptional arrow; or add one memory object with update and readout arrows. The constructive target is therefore not a class label but a presentation consisting of new objects, arrows, and equations.

Each of 1,200 ledgers contains five noisy reports for each coordinate of an eight-dimensional discrepancy between a frozen theory and its sources. There are 200 closed, mediating-factor, contextual-product, exceptional-arrow, memory-state, and compound factor-plus-memory worlds. The factor and product families deliberately have the same documentary signature. No amount of additional evidence of that same type can decide which presentation is appropriate. The compound family requires two new generators and exceeds the registered one-generator budget.

The constructor first projects the source residual onto each admissible one-generator signature. If the old theory explains the reports, it returns NO-CHANGE; if no single constructor explains them, it returns ABSTAIN-COMPLEXITY. Otherwise it retains every minimum-fit presentation. It then compiles an external probe whose predicted response has maximum range across those candidates. For a singleton candidate, this is the operation that most strongly excites its proposed generator. For the factor/product pair, an internal clamp distinguishes a mediated response from a contextual one even though their corpus fits coincide.

Five independently generated probe measurements are opened only after the question is fixed. Here “external” means external to the construction ledger: the probe is still generated by the benchmark oracle, not performed in nature.

Because the four one-generator presentations form a finite registered grammar, the exhaustive residual comparison supplies a nonrealizability gate for that grammar: the old theory is rejected only when it fails its frozen threshold, and a one-generator proposal is withheld when no registered candidate is adequate. This does not establish inadequacy relative to arbitrary scientific theories or unregistered presentations.

CTTE-4 passed all preregistered gates. It returned the correct structural action in all 1,200 worlds and recovered the exact finite presentation whenever an extension was applicable, admitted no wrong extension, preserved every frozen base prediction, and selected a maximum-disagreement probe in every queried case. It returned NO-CHANGE for every closed world and ABSTAIN-COMPLEXITY for every compound world. A fixed best-fit constructor reached 83.33% because it could not resolve the factor/product equivalence. Random probe selection reached 57.58%: uninformative probes were rejected by the admission firewall instead of being mistaken for confirmation.

| System                   | Structural-action accuracy | Constructs typed presentation? |
|--------------------------|----------------------------|--------------------------------|
| Flat residual regression | 0%                         | no                             |
| Fixed best-fit sketch    | 83.33%                     | yes, without a probe           |
| Random-probe synthesis   | 57.58%                     | yes                            |
| CTTE-4 typed synthesis   | 100%                       | yes                            |
| Extension oracle         | 100%                       | yes                            |

This perfect synthetic result should be interpreted through the withheld structure. The system did not invent the constructor grammar, parse scientific prose, or prove that a presentation is the world’s unique causal explanation. It did perform a more substantive operation than theory selection: construct a minimum finite presentation, recognize when documentary fit is non-identifying, turn that obstruction into an external question, and refuse a world that lies beyond its declared expressive budget.



**Experiment: CTTE-4: finite sketch extension and probe synthesis**

**Input:** 1,200 generated residual ledgers, a frozen base sketch, and a generic four-constructor grammar with a one-generator budget.

**Withheld structure:** the domain-level mechanism and its finite presentation; factor and product are observationally equivalent in discovery.

**Constructed artifact:** one minimal set of objects, arrows, and equations, together with a maximum-disagreement external probe.

**Result:** 100% structural-action accuracy and exact-presentation recovery on applicable extensions, zero wrong admissions, and 100% closed and complexity-abstention controls.

**Baselines:** 83.33% for fit-only selection and 57.58% for random probe synthesis.

**Verdict:** admitted as controlled finite-presentation synthesis, not open-ended theory invention or causal identification.

### 17.8.6 Cross-domain transport (CTTE-5)

CTTE-4 constructs a presentation inside one fixed observation and operation language. CTTE-5 asks whether that constructor can survive a change of domain presentation. The source interface is frozen before target generation: eight abstract observation ports, four one-generator constructors, four external operations, and the CTTE-4 thresholds. Four evaluator-held domains then independently permute and rename every observation and operation, change all measurement units and probe-response gains, and use disjoint local vocabularies.

Transport requires more than replacing nouns. A separate bridge split contains answer-independent structural fingerprints and calibration measurements. From these records, the system constructs a typed bijection between observation ports, estimates the target-to-source unit transformation, constructs a typed operation alignment, and estimates response-gain transport. It pulls the target residual back to the frozen interface, runs the unchanged constructor, and pushes the new presentation forward under fresh target names. Evaluation then canonicalizes the result through the inferred alignment. Thus a target object can have a name never seen in the source domain while still counting as the same abstract construction up to typed sketch isomorphism.

The preregistered benchmark contains 2,800 ledgers. In each of four regular domains, 100 examples are generated for each closed, factor, product, exception, memory, and compound family. Two additional controls test the admission boundary. In 200 ambiguous-alignment worlds, two local ports have the same structural fingerprint, so a bijective interface map is not identifiable. In 200 blocked-probe product

worlds, the target domain lacks both operations that distinguish the factor and product presentations. These require ABSTAIN-ALIGNMENT and ABSTAIN-NO-PROBE, respectively; neither is permission to choose the most familiar theory.

CTTE-5 passed all sixteen registered gates. It recovered both observation and operation alignments, the structural action, and the target presentation up to isomorphism in every regular world. It also returned the registered status in every closed, compound, ambiguous-alignment, and blocked-probe control, with no wrong extension admission. Literal and lexical reuse each achieved only 30.29% because local names and positions did not preserve the interface. Recovering observation alignment but omitting probe transport achieved 78.57%. The remaining 21.43-point gap localizes the importance of moving experimental operations along with the theory.

| System                     | Action accuracy | Transported artifact  |
|----------------------------|-----------------|-----------------------|
| Literal reuse              | 30.29%          | none                  |
| Lexical reuse              | 30.29%          | names only            |
| Aligned fit without probes | 78.57%          | observation interface |
| CTTE-5 typed transport     | 100%            | theory and operations |
| Transport oracle           | 100%            | hidden interface map  |

The result establishes a narrow finite commutation property: in these generated worlds, constructing after the registered alignment agrees with transporting the abstract construction. The experiment checks typed port and operation correspondences, not preservation of arbitrary identities, composites, or natural transformations; “functor” in the archived protocol is therefore best read as shorthand for this finite interface map. It does not show that scientific ontologies can be aligned from unrestricted prose. The bridge fingerprints are supplied, unusually clean, and explicitly typed; the source grammar remains fixed. This qualification is precisely what separates CTTE-5 from the GLP-1 challenge, where the relevant biological interface is only partially known.

**Experiment: CTTE-5: provenance-safe domain transport**

**Input:** 2,800 target ledgers with permuted local vocabularies, units, operations, and disjoint bridge, discovery, and probe sources.

**Frozen structure:** the CTTE-4 abstract interface, four constructors, one-generator budget, and admission thresholds.

**Constructed artifact:** inferred typed interface alignments, a transported experimental operation, and fresh target-named finite presentation.

**Result:** 100% action, interface-alignment, and presentation-isomorphism recovery; zero wrong admissions; all negative-control abstentions correct.

**Baselines:** 30.29% literal/lexical reuse and 78.57% observation alignment without probe transport.

**Verdict:** admitted as calibrated typed transport, not open-domain ontology alignment or biological transfer.

### 17.9 *A corpus-to-theory evidence hierarchy*

CTTE-0 isolates the value of contextual composition, but it supplies most of the language machinery. The evidence hierarchy hides these supports one at a time:

| Stage  | Withheld structure                                        | Construction                                      | Primary test                                        |
|--------|-----------------------------------------------------------|---------------------------------------------------|-----------------------------------------------------|
| CTTE-0 | one named mediator                                        | source-bearing contextual factorization           | locked relation prediction                          |
| CTTE-1 | entity registration and aliases                           | typed object and equivalence class                | entity precision, coverage, and abstention          |
| CTTE-2 | context boundaries                                        | local-model cover and restriction maps            | contradiction preservation and transfer             |
| CTTE-3 | decisive source or query                                  | active acquisition policy                         | power and time to decision at matched cost          |
| CTTE-4 | domain-level constructor choice within a supplied grammar | finite sketch extension and experimental question | irreducibility and sealed benchmark-probe admission |
| CTTE-5 | domain presentation                                       | transport to a generated target presentation      | provenance-safe reuse without semantic leakage      |

Within their registered generated worlds, CTTE-1 establishes controlled entity registration; CTTE-2 nearly recovers a minimal context cover but exposes an epistemic-status ordering defect; CTTE-3 shows

that a support-first obstruction can select the missing overlap efficiently within a supplied query language; CTTE-4 constructs a finite sketch extension and an external question rather than choosing among completed theories; and CTTE-5 transports that construction and its experimental operations through a changed typed interface. The sequence therefore reaches domain transport, but only under supplied calibration semantics. Real literatures pose the harder problem of discovering which populations, doses, time scales, assays, and tissues define the relevant cover, which measurements are feasible, and which parts of the interface can be aligned at all. Those unresolved tasks define the transition to the GLP-1 challenge.

### *17.10 The GLP-1 challenge*

The initial target is a scoped theory of how GLP-1 mechanisms act across physiological contexts. The goal is not encyclopedic coverage, but a precise set of mechanisms, boundaries, unresolved conflicts, and discriminating predictions.

The frozen reference theory contains established GLP-1 entities, source types, and mechanism families. At least one evaluator-held relationship, mediator, or context split is withheld from the proposal language. The system must decide whether persistent disagreement calls for a new biological object, a refined population or temporal context, or repair of its own extraction workflow.

CLIC proposes the smallest causal extension and a counter-witness experiment. OPTIC constructs the source and data workflow needed to operationalize the new variables and, when a simulator or analysis package is available, an executable probe. RELIC ranks additional retrieval and experimental actions. The final artifact includes the theory map, source ledger, alternative explanations, new predictions, and the status of every claim.

The first pilot deliberately narrows this broad ambition. It asks whether a single missing context can be constructed from source-separated trial records without pretending to discover a new biological mechanism.

#### *17.10.1 Treatment status as a missing context (GLP1-PILOT-o)*

Active-treatment efficacy and off-treatment maintenance answer different questions. STEP 1 and STEP 5 report weight outcomes while semaglutide is being administered, whereas STEP 4 explicitly randomizes participants after an active-drug lead-in to continued semaglutide or placebo [Wilding et al., 2021, Rubino et al., 2021, Garvey et al., 2022]. SURMOUNT-1 supplies active-treatment evidence for the distinct agent

| Registered quantity                  | Result |
|--------------------------------------|--------|
| Admission gates                      | 12/12  |
| Probe-design field accuracy          | 1.0    |
| Qualitative confirmation             | 1.0    |
| Necessity and safety controls        | 5/5    |
| Source-order replays / unique hashes | 24/1   |
| Construction–confirmation overlap    | 0      |
| Numerical effect-size claims         | 0      |

Table 17.1: GLP1–PILOT–o retrospective transport from four construction records to the locked SURMOUNT–4 confirmation record. The registered theory predicts the direction of withdrawal-associated weight regain without estimating an effect size. The 24 source orderings are replay audits, not independent scientific observations or prospective clinical evidence.

tirzepatide [Jastreboff et al., 2022]. A flat list can summarize these results, but it need not represent treatment status as an argument of the outcome map.

The pilot therefore supplied a human-curated semantic interface: semaglutide and tirzepatide remained different drugs but were grouped under the broad role *incretin-based anti-obesity therapy* for this particular analogy; body-weight change and cardiovascular events remained different endpoints; and populations with and without diabetes remained different contexts. The constructor was required either to abstain or to add `TreatmentStatusAtEvaluation` and produce

$$\text{Tirzepatide} \times \text{TreatmentStatusAtEvaluation} \times \text{Population} \times \text{Time} \\ \longrightarrow \text{WeightTrajectory}.$$

This extension compiles an external probe: administer tirzepatide during a lead-in, randomize continued treatment against withdrawal to placebo, and compare subsequent weight trajectories. The registered prediction was only directional—withdrawal would produce greater regain than continuation—and contained no effect-size estimate.

The four construction records were STEP 1, STEP 4, STEP 5, and SURMOUNT–1. The resulting proposal and its SHA-256 hash were written before the program opened a separate SURMOUNT–4 confirmation ledger. SURMOUNT–4 in fact used a tirzepatide lead-in followed by randomized continuation or switch to placebo and reported greater regain after withdrawal [Aronne et al., 2024]. SELECT was registered as a counter-control: although it studies semaglutide in people with overweight or obesity and no diabetes, its endpoint is major adverse cardiovascular events rather than weight maintenance [Lincoff et al., 2023].

Every registered necessity and type-safety intervention behaved as required:

- removing STEP 4 yielded `ABSTAIN_NO_ANALOG`;
- removing SURMOUNT–1 yielded `ABSTAIN_NO_TARGET_SUPPORT`;
- SELECT yielded `ABSTAIN_ENDPOINT`;

- extrapolation to type 2 diabetes yielded ABSTAIN\_POPULATION; and
- a request for an exact magnitude yielded ABSTAIN\_QUANTITATIVE.

All 24 permutations of the construction ledger produced the same proposal hash. On the frozen nine-field artifact rubric, typed construction scored 1.000, flat aggregation 0.333, and untyped analogy 0.444. These are completeness scores over a declared artifact, not sampling estimates.

**Experiment: GLP1-PILOT-0: retrospective withdrawal-context transport**

**Input:** four human-curated, source-bearing trial records and a supplied typed interface.

**Construction:** add treatment status, transport a withdrawal-test pattern across two distinct agents, and compile a qualitative external probe.

**Confirmation:** open the separate SURMOUNT-4 ledger only after the proposal is frozen.

**Controls:** remove the analog or target support; substitute the SELECT endpoint; change the population; request an unlicensed magnitude; permute all construction records.

**Boundary:** one retrospective prediction-and-confirmation artifact, not automated full-paper understanding, prospective discovery, a newly identified biological mechanism, clinical guidance, or evidence that the two drugs have identical pharmacology.

### 17.10.2 *Discovering the missing context (GLP1-PILOT-1)*

PILOT-0 tests a supplied context. PILOT-1 removes the treatment-status field, its values, the withdrawal-design name, and the declared context object. The constructor receives human-curated arm-level records containing only an investigational agent, population and endpoint types, evaluation time, qualitative trajectory direction, and a sequence of neutral set\_exposure events. The comparison role joining the two agents is still supplied. Thus the experiment removes one central semantic aid without conflating context discovery with full-paper parsing or pharmacology.

The frozen grammar contains five possible factorizations: no context, evaluation time, current agent identity, current investigational exposure, and the pair

(agent seen before final assignment, agent present at evaluation).

For each partition, a group is contradictory when it contains both increasing and decreasing weight trajectories. The registered score

| Candidate factorization          | Frozen score |
|----------------------------------|--------------|
| No context                       | 11           |
| Evaluation time                  | 13           |
| Current agent identity           | 13           |
| Current investigational exposure | 12           |
| Exposure history                 | 4            |

Table 17.2: GLP1-PILOT-1 candidate context factorizations under the frozen contradiction-plus-complexity score; lower is better. Exposure history is the unique minimum, whereas evaluation time, current drug identity, and current exposure fail to preserve the relevant longitudinal distinction. This is a retrospective representation result, not clinical identification.

is ten times the number of contradictory groups plus the number of nonempty groups. This penalty and the five candidates were fixed before running the constructor.

The obstruction is concrete. A current-exposure representation places ordinary control arms and the STEP 4 switched arm in the same “agent absent” class. The ordinary controls have decreasing mean trajectories in the curated records, whereas the previously exposed switched arm has an increasing trajectory. Current state therefore loses relevant history. Refining it by whether the agent occurred before the final assignment removes the conflict and yields four anonymous classes:

$$H_{00}, H_{01}, H_{10}, H_{11}.$$

Only after the proposal was frozen did the evaluator identify these classes, up to isomorphism, with never exposed, initially exposed and present, previously exposed then replaced by control, and previously exposed and preserved. The resulting presentation is

$$\begin{aligned} &\text{Tirzepatide} \times \text{ExposureHistoryAtEvaluation} \times \text{Population} \times \text{Time} \\ &\longrightarrow \text{WeightTrajectory}. \end{aligned}$$

The constructor compiled the same two-stage experimental shape without using that name: set tirzepatide exposure for every participant, randomly preserve it or replace it with control, and compare subsequent trajectories. The directional prediction was greater regain after replacement. After the proposal hash was written, the neutralized SURMOUNT-4 ledger matched all six registered agent, population, endpoint, assignment, class, and direction fields [Aronne et al., 2024].

PILOT-1 passed all 13 gates. The four classes were recovered exactly up to the registered isomorphism; all 24 source permutations produced one proposal hash; and all six controls returned their frozen abstentions. In particular, removing STEP 4 removed support for the history object, removing SURMOUNT-1 removed target support, and adding an explicitly synthetic opposite-direction copy of the switched arm produced ABSTAIN\_CONTEXT\_CONTRADICTION. SELECT, population transport, and exact-magnitude requests were also rejected. The typed artifact-completeness score was 1.000, compared with 0.455 for the current-exposure representation. These are field-coverage scores, not independent experimental samples.

**Experiment: GLP1-PILOT-1: hidden-context recovery**

**Input:** curated arm timelines with status labels, named design, and context object removed.

**Obstruction:** current exposure conflates never-exposed controls with a previously exposed arm later assigned control.

**Construction:** a minimal four-class exposure-history object and a two-stage tirzepatide probe.

**Result:** 13/13 gates, 6/6 controls, 24/24 invariant order replays, and complete qualitative agreement with the sealed confirmation.

**Boundary:** finite-grammar context recovery from human-curated events, not open-ended variable invention, prose understanding, prospective discovery, clinical inference, or evidence of identical drug mechanisms.

### 17.11 *A grant proposal as a capstone artifact*

There is a familiar institutional form for the transition from a scientific gap to an experimental program: the research-grant proposal. Scientists ask an agency to provide the equipment, personnel, data, travel, and time needed to test a theory that cannot yet be admitted. A competitive proposal must digest the literature, explain why the state of the art is incomplete, formulate a novel and consequential hypothesis, organize several years of specific aims, and make a credible case that the proposed experiments can decide something important. In that sense, grant writing is already a socially embedded form of theory-and-experiment compilation.

This analogy gives the GLP-1 challenge a tangible capstone. Its product need not be an encyclopedic model of incretin biology. It can instead be a *grant-ready research strategy* generated from a temporally bounded corpus and an incomplete theory:

$$(S, D) \mapsto (S^+, \Omega, \mathcal{A}, \mathcal{E}, \mathcal{R}).$$

Here  $\Omega$  is the localized gap,  $S^+$  the proposed extension,  $\mathcal{A}$  the live alternatives,  $\mathcal{E}$  the linked experimental aims, and  $\mathcal{R}$  the expertise and resources required to execute them. The proposal is therefore not merely prose about a topic. Every aim should be traceable to the obstruction, every experiment should discriminate among declared alternatives, and every important premise should retain source and epistemic status.

The analogy is more than rhetorical. The current simplified NIH framework for most research-project grants asks reviewers to assess the importance of the research, the rigor and feasibility of the approach, and the adequacy of the expertise and resources [[National Institutes](#)



| Grant component             | Corpus-to-theory interpretation            |
|-----------------------------|--------------------------------------------|
| Related work                | Maintained theory and source ledger        |
| Significance and innovation | Obstruction and proposed extension         |
| Specific aims               | Dependency-ordered tests of the extension  |
| Research approach           | Compiled probes, controls, and analysis    |
| Risks and alternatives      | Unsupported branches and counter-witnesses |
| Expertise and resources     | Operational requirements of the tests      |

Table 17.3: A grant proposal as an externally reviewable realization of a source-bearing theory extension. The maintained literature supplies related work; the localized obstruction motivates significance and aims; and the admission interface becomes an executable approach with risks, alternatives, expertise, and resources. The correspondence organizes an auditable proposal; it does not certify biomedical validity or predict funding.

of Health, 2026]. These factors align respectively with the value of the obstruction and extension, the quality of the compiled tests, and the operational realizability of the program. They also expose a distinction that a theory constructor must respect: inventing a promising experiment does not imply that a particular team or institution can execute it.

The resulting capstone challenge, **GLP1-GRANT**, is intentionally ambitious. Given a frozen literature cutoff, a funding opportunity, and an explicit investigator-and-resource profile, the system produces a one-page Specific Aims document and a source-bearing research strategy. Its theory map, claim ledger, experimental dependencies, controls, milestones, stopping rules, and failure contingencies remain machine-auditable even when the narrative is written for human reviewers. A blinded panel can compare this artifact with a generic language-model proposal, retrieval-augmented synthesis, flat evidence-to-grant generation, and—where available—a human reference.

Actual funding is an aspirational external outcome, not a clean benchmark label. Awards depend on program priorities, budgets, investigator credentials, institutional resources, and review variability as well as scientific merit. The defensible initial endpoint is therefore narrower: can independent domain experts identify a real gap, follow every specific aim back to the proposed theory extension, and judge that the experiments would distinguish it from credible alternatives? Nor should the constructor fabricate administrative elements such as institutional commitments, biosketches, facilities, or human-subjects approvals; these must be supplied or completed by responsible human participants.

### Experiment: GLP1–GRANT capstone: theory extension to research program

**Epistemic status:** capstone evaluation specification; the subsequent GRANT rungs are controlled precursors, not a completed capstone result.

**Input:** temporally bounded literature, maintained GLP-1 sketch, funding opportunity, and explicit team-and-resource profile.

**Artifact:** Specific Aims, source-bearing research strategy, theory extension, linked experiments, controls, milestones, alternatives, and resource requirements.

**Evaluation:** blinded scientific review against generic, retrieval-augmented, flat-synthesis, and human-reference proposals where available.

**Admission:** importance, rigor, feasibility, provenance, falsifiability, and coherence of the multi-aim program.

**Boundary:** grant-ready scientific strategy, not a claim of funding, clinical validity, institutional authorization, or autonomous submission.

The controlled studies below are organized by semantic obligation rather than by an attempt to perfect grant-writing style. GRANT-0 tests structural compilation; GRANT-1 adds epistemic types; GRANT-2 separates a typed intermediate program from its prose realization; GRANT-3 adds cross-field scientific invariants; and the GRANT-4 series asks how an admitted program can be rendered without omission, accretion, or loss of auditability. Each negative result freezes the failed artifact and changes one declared layer. The scientific object throughout is the source-bearing research program, not the fluency of its surface prose. In particular, GRANT-4 stress-tests one interface—the passage from an admitted typed program to human-readable prose—without upgrading the biomedical evidence of the underlying program.

### Design principle

A research proposal is not the theory extension itself. It is an externally reviewable realization of that extension. Rhetorical fluency is admissible only when it preserves the source ledger, epistemic types, open decisions, resource boundaries, and discriminating structure of the underlying program.

| Layer                  | Question                                                                     | Characteristic obstruction                                                         |
|------------------------|------------------------------------------------------------------------------|------------------------------------------------------------------------------------|
| Program structure      | Are aims, dependencies, controls, alternatives, and sources present?         | A proposal-shaped artifact can be complete while making unsupported commitments.   |
| Epistemic semantics    | Is every commitment supported, supplied, hypothetical, or explicitly open?   | Fluent prose can bypass a typed ledger.                                            |
| Scientific invariants  | Do aims preserve causal, measurement, transport, and authority contracts?    | Exact compilation can faithfully realize an incoherent program.                    |
| Rhetorical realization | Can the admitted program be rendered readably without omission or accretion? | Fluency can introduce new claims; mechanical proxies can reward superficial edits. |

The experiment identifiers preserve protocol provenance; they do not define the exposition. The following subsections are therefore layers of one constructor, with each frozen rejection exposing an interface that the next layer must represent explicitly.

#### 17.11.1 *Structural compilation without semantic admission*

GLP1-GRANT-o asks a deliberately narrower question than whether an AI can write a fundable proposal: does the maintained structure accumulated by Democritus and GLP1-PILOT-1 improve the construction of an auditable research program? The controlled comparison used three temperature-zero generations from the same GPT-OSS 20B model and the same frozen five-study corpus. The *flat* condition received ordinary article summaries; *ledger* received provenance-bearing causal claims and restriction obligations; and *compiled* additionally received the exposure-history obstruction, a three-aim program grammar, dependency rules, a mock funding call, and an explicit team-and-resource profile.

All conditions received the same output schema. The registered score measured ten aspects of contract satisfaction, including program- and aim-field coverage, citation validity, per-aim provenance, controls, alternative outcomes, discriminating observations, dependency validity, context alignment, resources, and epistemic boundaries. It did not score biological importance, novelty, ethics, feasibility, or fundability.

| Condition | Score        | Program | Aims  | Valid sources | Provenance |
|-----------|--------------|---------|-------|---------------|------------|
| Flat      | 0.892        | 1.000   | 0.917 | 1.000         | 1.000      |
| Ledger    | 0.892        | 1.000   | 0.917 | 1.000         | 1.000      |
| Compiled  | <b>1.000</b> | 1.000   | 1.000 | 1.000         | 1.000      |

The result was *not admitted*: eleven of twelve preregistered gates passed. The compiled artifact improved on the flat condition by 0.108, short of the frozen 0.15 margin. This ceiling effect is itself diagnostic. Once every condition is given a detailed JSON schema, a capable language model can produce a proposal-shaped document with valid citations, controls, alternatives, and a dependency graph. The compiled representation supplied the missing exposure-history and resource alignment, but structural completeness alone did not sharply separate it from ordinary generation.

A retrospective semantic audit, conducted outside the registered gates, revealed the more important limitation. The compiled program invented participant counts, trial durations, doses, blinding choices, endpoints, and claims of resource sufficiency that were not licensed by the ledger. It treated access to five source summaries and human-curated arm events as if participant-level trial records were available, and it proposed transport to a cardiovascular population without a sufficiently grounded bridge. Its experiments largely compared efficacy rather than discriminating among mechanistic explanations of regain and durability. The artifact therefore passed the structural contract while remaining scientifically premature.

This negative result identifies a missing layer. A grant constructor needs a semantic admission firewall in addition to a document schema. Every protocol commitment must be classified as source-supported, externally supplied, or hypothetical; unsupported numerical and operational choices must remain open variables; resource feasibility must be checked against actual capacity; and experiments must be scored by whether they distinguish declared causal alternatives. GLP1-GRANT-o thus supplies a useful baseline and exposes why grant construction cannot be reduced to well-formed scientific prose.

**Experiment: GLP1–GRANT-0: controlled program compilation**

**Input:** five frozen GLP-1 study summaries, with flat, claim-ledger, and fully compiled representations.

**Model:** GPT-OSS 20B, one temperature-zero generation per condition.

**Result:** compiled score 1.000 versus 0.892 for flat and ledger; 11/12 registered gates; not admitted because the 0.15 improvement margin was not met.

**Semantic audit:** unsupported protocol and resource commitments remain despite perfect structural contract satisfaction.

**Boundary:** one small retrospective compilation test, not expert peer review, scientific validation, clinical guidance, or evidence of fundability.

### 17.11.2 *Semantic admission as a type system*

GLP1–GRANT-1 makes the semantic firewall executable. It reuses the same five-study ledger, exposure-history obstruction, mock funding call, team profile, and three-aim grammar, but replaced unconstrained specificity by four explicit epistemic types: *source-supported*, *externally supplied*, *hypothetical*, and *open variable*. Every unresolved sample, duration, regimen, endpoint, site, recruitment, supply, budget, data-access, or authorization choice had to be represented as `TO_BE_DETERMINED` together with a procedure for resolving it. Thus omission and invention were both type errors: a proposal could not gain apparent rigor by silently choosing plausible numbers.

One temperature-zero GPT-OSS 20B generation satisfied the structural contract, retained three discriminating dependency-ordered aims, used only registered sources, kept all hypotheses hypothetical, supplied every required commitment field with a valid type, and explicitly separated resource gaps from human authorizations. The firewall also rejected the frozen GLP1–GRANT-0 proposal and five deterministic counter-witnesses: an unsupported sample size, a fabricated source, a false authorization, a promoted hypothesis, and a missing commitment. Each mutation was rejected for its preregistered reason.

The generated program nevertheless failed admission. Although duration was correctly declared open in the protocol table, the surrounding narrative specified 12-month discriminators for all three aims and 24-month alternative outcomes for Aims 2 and 3. These five statements were unlicensed future protocol commitments. A frozen-output evaluator correction excluded the digit in the lexical term “GLP-1” from the numerical detector and extended the same preregistered rule to all protocol-bearing aim prose; neither the output nor the threshold was

changed. The corrected result passed ten of twelve gates and remained not admitted.

This failure is more informative than a malformed proposal. The generator learned to populate the explicit type system but treated fluent prose as an escape path around it. A semantic firewall must therefore check the entire artifact, not merely its designated commitment table. It must also distinguish three successively stronger claims: a well-typed proposal, a scientifically credible proposal, and a fundable proposal. GLP1-GRANT-1 tests only the first.

**Experiment: GLP1–GRANT–1: semantic admission firewall**

**Input:** the frozen GLP1–GRANT–0 compilation packet plus epistemic types and required open-variable resolution procedures.

**Model:** GPT-OSS 20B, one temperature-zero generation.

**Result:** 10/12 gates; all structural, provenance, typing, dependency, resource-boundary, legacy-rejection, and counter-witness gates passed.

**Rejection:** five untyped duration commitments leaked into narrative discriminators and alternative outcomes.

**Boundary:** semantic contract checking over a five-source curated corpus, not biomedical review, feasibility validation, clinical guidance, or evidence of fundability.

### 17.11.3 *Compiling prose from a typed intermediate program*

GLP1–GRANT–2 places the firewall before prose generation. Rather than writing a proposal and classifying its claims afterward, Qwen3-Next-80B produces a typed intermediate representation (IR). A deterministic compiler alone rendered reviewer-facing prose. The IR used a frozen vocabulary of design classes, observables, and controls; retained the four epistemic types from GRANT–1; and represented all twenty-one required protocol fields explicitly. Stable markers in the compiled document allowed a separate parser to recover the commitment ledger. Admission required exact agreement between the proposed and recovered ledgers as well as the absence of unlicensed numerical choices anywhere in the future-program narrative.

This architecture fixed the failure exposed by GRANT–1 at one level. The IR passed the structural and semantic program audit; all sources were registered; all hypotheses remained hypothetical; every required commitment was typed; the finite design vocabulary and aim dependencies were valid; all twenty-one commitments round-tripped exactly; and ten compiler replays produced one hash. The frozen GRANT–1 artifact and all six registered counter-witnesses were rejected for their intended reasons. These included both IR mutations and prose-only tampering, so the validator could not succeed merely by ignoring the document presented to a reviewer.

The result was nevertheless *not admitted*: thirteen of fifteen gates passed. Ten numerical protocol choices had migrated into the free-form *resolution procedures* attached to nominally open variables. They included copied source-trial durations, endpoint schedules, recruitment and eligibility counts, doses, site counts, and study configurations. The compiler preserved these procedures faithfully and the round trip recovered them, but fidelity to an unlicensed commitment is not

semantic safety. The experiment therefore localizes the missing abstraction: resolution procedures themselves need typed operators such as power analysis, feasibility assessment, regulatory review, and supply negotiation, rather than unrestricted prose.

A retrospective scientific audit exposed a distinct limitation not scored by these gates. Aim 2 varied current drug together with exposure history and therefore would not isolate the proposed exposure-history effect. Aim 3 named oral semaglutide as an example of a non-GLP-1 agent, which is a type error in the biomedical domain. Exact compilation can certify that the narrative says what the IR says; it cannot certify that the IR defines a scientifically coherent experiment. Cross-field invariants and expert review remain separate admission layers.

**Experiment: GLP1–GRANT-2: typed IR and narrative round trip**

**Input:** the frozen five-study program packet, finite design vocabulary, and semantic commitment types.

**Model:** Qwen3-Next-80B-A3B-Instruct-6bit, one temperature-zero IR generation; deterministic narrative compiler.

**Result:** 13/15 gates; 21/21 commitments recovered exactly; stable compiler replay; legacy and all six counter-witnesses rejected.

**Rejection:** ten unlicensed numerical choices appeared inside free-form open-variable resolution procedures.

**Boundary:** typed compilation and tamper detection, not biomedical coherence, causal identification, peer review, clinical guidance, or evidence of fundability.

#### 17.11.4 *Scientific invariants across a research program*

GLP1–GRANT-3 represents every resolution procedure by a typed operator. Each of the twenty-one open protocol commitments named a registered operation—such as power analysis, endpoint alignment, site-capacity assessment, supply confirmation, or regulatory review—together with finite symbolic inputs, a pending status, a null numerical payload, and the human authority responsible for resolving it. The language model could therefore propose which kind of work remained without silently performing that work or inventing its result. The deterministic compiler carried stable markers for both commitments and resolution operators, and independent parsers had to recover both ledgers exactly.

The experiment also checked relations that no field can establish alone. Aim 2 was required to stratify by prior exposure and assign the same current agent and regimen across exposure groups; this blocked the agent/exposure confound found in GRANT-2. Aim 3 had to



transport the proposed effect to a registered non-incretin intervention class while leaving the particular agent open. Each primary observable had to equal its measurement target, each operator had to match both its protocol field and human authority, and the three aims retained their dependency order.

One temperature-zero Qwen3-Next-80B generation passed all seventeen registered gates. All twenty-one commitments and resolution operators round-tripped exactly, the compiled prose contained no unlicensed future-protocol number, and ten compiler replays produced one hash. The frozen GRANT-2 artifact and seven mutations were rejected for the intended reasons: treatment/exposure confounding, an invalid transport agent, a numerical operator payload, a field/operator mismatch, an unauthorized operator, an observable/measurement mismatch, and deletion of a resolution marker.

Admission remains relative to the declared theory. A retrospective scientific audit noted that exposure history itself is observed rather than randomized; holding current treatment fixed removes one confound but does not by itself identify the causal effect of prior exposure. The program also called itself the first prospective test without support from a literature-wide novelty review. A complete artifact language therefore needs typed identification assumptions, source-scoped novelty claims, and checks that narrative sentences are entailed by the causal contract. Human scientific and peer review remain admission authorities rather than implementation details.

#### Experiment: GLP1-GRANT-3: cross-field scientific invariants

**Input:** the frozen GRANT-2 packet plus a finite resolution-operator language and registered causal, transport, measurement, and authority invariants.

**Model:** Qwen3-Next-80B-A3B-Instruct-6bit, one temperature-zero typed-IR generation; deterministic compiler and independent dual-ledger parsers.

**Result:** 17/17 gates; 21/21 commitments and resolution operators recovered exactly; no untyped protocol numbers; stable replay; GRANT-2 and all seven counter-witnesses rejected.

**Interpretation:** admitted relative to the registered program language; the result demonstrates executable cross-field integrity, not causal identification, biomedical validity, peer-review quality, or fundability.

#### 17.11.5 Auditable rhetorical realization

An admitted program is not yet a grant proposal. GLP1-GRANT-4 tests the passage from the GRANT-3 IR to reader-facing prose under

three conditions: a free-form language-model proposal, the frozen deterministic compiler, and an auditable realization. In the third condition, Qwen3-Next-80B emitted a JSON rhetorical plan whose atomic sentences carried a registered epistemic type and support identifiers. A deterministic compiler produced both a clean proposal and a marker-bearing audit document, from which a separate parser reconstructed the sentence ledger.

The registered language distinguished source-supported statements, hypotheses, admitted inferences, open variables, resolution operators, resource boundaries, and epistemic boundaries. Admission required the proposal to cover all three contracts, hypotheses, alternatives, and fallbacks; all twenty-one commitments and resolution operators; and the program's resource and epistemic boundaries. It also prohibited unsupported priority language, causal claims about the nonrandomized exposure history, and invented protocol numbers. Eight counterwitnesses tested precisely these boundaries as well as source fabrication, hypothesis upgrading, marker deletion, omitted alternatives, and apparent closure of an open variable.

The auditable proposal was *not admitted*: fifteen of nineteen gates passed. Its 64 typed sentences covered every required program object, and its sentence ledger round-tripped exactly. All eight mutations were rejected for their registered reasons. The generator nevertheless repeated the inherited claim that this was the first prospective study, invented six unregistered sentence types called alternatives and fallback, and typed three causal-contract sentences as epistemic boundaries. The first failure shows that realization can detect a defect in the admitted input theory rather than merely introduce one. The latter failures suggest a conservative extension of the rhetorical language: causal contracts, alternative explanations, and fallback policies may deserve explicit types.

The independent free-form condition supplied the contrasting failure mode. It had no sentence ledger, used only five of the seven required headings, repeated both unsupported novelty and causal claims, and invented fixed protocol timings and analytic commitments. This comparison is diagnostic rather than a readability result: no human reviewer preference was measured. A retrospective audit also found that the supposedly clean auditable proposal echoed registry identifiers in visible prose and exceeded its requested word budget. A stricter realization language must therefore sanitize the admitted theory, separate machine certificates from surface text, type the rhetorical budget, and add the new sentence categories without weakening existing checks.

**Experiment: GLP1–GRANT–4: auditable proposal realization**

**Input:** the admitted GRANT–3 IR, frozen deterministic proposal, finite sentence-type language, and support-identifier registry.

**Conditions:** one temperature-zero Qwen3-Next-80B free-form proposal and one sentence-audited rhetorical plan; deterministic compilation and independent ledger recovery.

**Result:** 15/19 gates; 64 typed sentences; complete coverage of 21 commitments and 21 resolution operators; exact round trip; all eight counter-witnesses rejected.

**Rejection:** unsupported priority language, six unregistered rhetorical types, and three contract/type mismatches.

**Boundary:** auditable realization calibration, not a biomedical, readability, peer-review, novelty, or fundability assessment.

#### 17.11.6 Sanitized realization and surface separation

GLP1–GRANT–4.1 isolates sanitization and surface separation. Before generation, a deterministic semantic view removes the unsupported claim of priority and replaces claims about the causal role of observed exposure history by association statements and explicit non-identification boundaries. The rhetorical language conservatively added three sentence types—*causal contract*, *alternative explanation*, and *fallback policy*. Machine support identifiers were confined to metadata, and 900–1,500 words became a hard admission contract rather than a request. These choices were frozen before one new temperature-zero Qwen3-Next-80B realization; no repair call was allowed afterward.

The resulting proposal was substantially safer. Its 66 typed sentences used 1,330 words, covered all three hypotheses, contracts, alternatives, and fallbacks, and retained all twenty-one commitments and resolution operators. It introduced no unsupported novelty statement, causal overclaim, fixed future protocol number, or visible registry token. Its audited representation round-tripped exactly, compilation was stable, and all nine registered counter-witnesses were rejected for the intended reasons. Both frozen GRANT–4 conditions also failed the stricter contract.

The artifact nevertheless remained *not admitted*: thirteen of sixteen gates passed. Two sentences describing significance and innovation used an unregistered type named `program_field` rather than the licensed `admitted_inference` type. More importantly, the proposal named the team’s available resources but omitted its declared resource gaps and required human authorizations. This is not merely a cosmetic omission. A prospective program must state both what the team can do and which scientific, regulatory, budgetary, and institutional decisions

remain outside the generator's authority.

GRANT-4.1 therefore localizes a new obstruction. Sanitizing theory content and separating certificates from prose prevented the earlier scientific and surface failures, but an instruction to mention every semantic object did not guarantee a complete rhetorical structure. The corresponding realization obligation is to allocate required semantic slots before prose generation and permit the model to fill only that typed skeleton. This is a structure-first contract, not a post-hoc correction of the frozen negative result.

**Experiment: GLP1-GRANT-4.1: sanitized auditable realization**

**Input:** a conservative view of the admitted GRANT-3 theory, an extended sentence grammar, a surface/metadata separation rule, and a hard 900–1,500-word contract.

**Model:** one temperature-zero Qwen3-Next-80B realization; no repair call.

**Result:** 13/16 gates; 1,330 words and 66 sentences; complete hypothesis, contract, alternative, fallback, commitment, and resolution coverage; exact round trip; all nine counter-witnesses rejected.

**Rejection:** two unregistered type labels and omission of resource gaps and human authorizations.

**Interpretation:** theory sanitization and certificate separation worked, while structure-first completeness remains unresolved; this is not biomedical validation, peer review, or evidence of fundability.

### 17.11.7 *Structure-first rhetorical realization*

GLP1-GRANT-4.2 relocates the proposed structural repair from unconstrained prose generation to a deterministic rhetorical planner. Before Qwen sees the program, the compiler allocates 67 immutable sentence slots. Each slot had a fixed section, epistemic type, support array, and semantic payload. The inventory included every hypothesis, identification contract, program field, source statement, alternative, fallback, open commitment, resolution operator, resource boundary, and epistemic boundary. The model could supply only one reader-facing sentence for each identifier in the frozen order. It could neither add nor delete a slot, change its type, nor move its evidential support.

This intervention repaired the structural failures of GRANT-4.1. The output filled all 67 slots exactly, preserved every type and support edge, included both resource gaps and human authorizations, retained all source numbers, and round-tripped through the audit compiler. It contained none of the registered novelty, causal, protocol-number, or surface-token violations. All eleven counter-witnesses—including

missing and extra realizations, semantic replacement, numerical drift, metadata tampering, and marker deletion—were rejected for their intended reasons.

The experiment was nevertheless *not admitted*: fifteen of seventeen gates passed. The model expanded a 1,411-word semantic skeleton into 1,958 words, exceeding the hard 1,500-word ceiling. The artifact-admission gate therefore failed as a consequence. This is more informative than a cosmetic length error. A retrospective audit found that the expansion added plausible but unlicensed details, including particular missing-data estimators, IRB and award timelines, enrollment targets, unblinding, and agency review cycles. The registered overlap test ensured that the original payload remained visible; it did not ensure that no new claim had been appended.

The residual obstruction is thus *semantic accretion*. Structure-first planning can guarantee that the generator says something about every required object, but not that it says only what the admitted theory licenses. Conservative realization consequently requires an upper semantic bound—for example, deletion-only compression, bidirectional claim entailment, or deterministic clause compilation with a narrowly bounded connective layer. The frozen proposal cannot acquire that certificate through post-hoc editing.

#### Experiment: GLP1–GRANT–4.2: structure-first realization

**Input:** an immutable 67-slot rhetorical skeleton constructed from the sanitized GRANT–3 theory before prose generation.

**Model:** one temperature-zero Qwen3-Next-80B realization; text fields only and no repair call.

**Result:** 15/17 gates; exact slot, type, support, source-number, coverage, round-trip, and replay checks; all eleven counter-witnesses rejected.

**Rejection:** 1,958 words exceeded the 1,500-word ceiling; a secondary audit outside the registered gates also found unsupported procedural elaboration.

**Interpretation:** structure-first generation solved omission and retyping, while conservative semantic realization remains unresolved; this is not biomedical validation, peer review, or evidence of fundability.

#### 17.11.8 Local metric and semantic bounds

GLP1–GRANT–4.3 treats verbosity as a local rather than merely global obstruction. Each of the 67 frozen slots receives a minimum and maximum word count before generation, with its maximum no larger than the admitted semantic payload. The sum of the local maxima

was 1,453 words, so satisfying every local fiber automatically respected the 1,500-word document ceiling; a 900-word global floor prohibited empty or trivial compression. A second bound made semantic fidelity bidirectional: the earlier coverage test prevented replacement of the payload, while a normalized vocabulary test limited new content not licensed by it.

The single temperature-zero Qwen3-Next-80B realization passed all nineteen registered gates. It filled every slot in order, preserved all types, supports, source numbers, and program objects, and contained 1,453 words. It introduced no registered novelty, causal, numeric-protocol, or surface-token violation; its sentence ledger round-tripped exactly; and all fourteen counter-witnesses were rejected for their intended reasons. Relative to this bounded rhetorical language, the artifact was admitted.

The secondary diagnostic is essential to interpreting that success. All 67 output strings were byte-for-byte identical to their semantic payloads. Faced with strict upper bounds, the model chose exact transcription rather than paraphrase. This is a legitimate conservative solution and a useful contrast with the accretion in GRANT-4.2, but it adds no demonstrated rhetorical value. The copied compiler text retains awkward constructions such as “Alternative explanations include Weight,” “remains open because Requires,” and “The sites remains open.” Neither readability nor reviewer preference was a registered outcome.

GRANT-4.3 is therefore an admission of *bounded faithful realization*, not of successful grant writing. A decisive comparison retains exact copying as a conservative baseline while testing a controlled paraphrase language under bidirectional claim entailment and blinded human readability. That would ask whether a realization layer can improve prose without weakening the semantic firewall.

#### Experiment: GLP1–GRANT-4.3: bounded realization

**Input:** the 67-slot structure-first skeleton with local word intervals, a global 900–1,500-word contract, and lower and upper semantic bounds.

**Model:** one temperature-zero Qwen3-Next-80B realization; no repair call.

**Result:** 19/19 gates; 1,453 words; exact structural, semantic, numerical, round-trip, and replay checks; all fourteen counter-witnesses rejected.

**Diagnostic:** 67/67 sentences exactly copied their payloads.

**Interpretation:** admitted as bounded faithful transcription, not as a readability, peer-review, biomedical, or fundability result.

### 17.11.9 *Controlled paraphrase and evaluator failure*

GLP1-GRANT-4.4 compares three realizations of the same frozen 67-slot program. The COPY arm transcribes each admitted payload exactly. The CONTROLLED arm received local length intervals and permission to repair grammar or paraphrase, but not to add scientific, procedural, numerical, temporal, regulatory, or resource content. The FLUENT arm was asked for polished and persuasive grant prose while retaining the slot inventory. All three artifacts were subsequently subjected to the same semantic firewall; fluency was not permitted to compensate for an unsupported claim.

The comparison illustrates why evaluation must itself be treated as a typed theory. COPY remained admissible at 1,453 words. CONTROLLED also passed the five preregistered machine gates: it changed 24 slots, contained 1,456 words, preserved every structural and semantic contract, reduced a registered boilerplate counter from 46 to 22, and rejected six counter-witnesses. By contrast, FLUENT changed 64 slots, expanded to 1,822 words, and was rejected for violations of the global and local length contracts together with widespread semantic accretion. Its additions often sounded plausible, but plausibility did not license them.

The formal pass for CONTROLLED did not survive qualitative inspection. Only three changes repaired malformed sentences about alternative explanations. The other 21 merely changed an initial “Requires” or “Must” to lowercase after “because.” This evaded a case-sensitive defect pattern while preserving phrases such as “remains open because requires”; the error “The sites remains open” also survived. A blinded diagnostic call preferred COPY and explicitly noticed the residual lowercase constructions, although that call used the same model family and is not an independent human judgment.

The correct conclusion is therefore a negative one. GRANT-4.4 formally admitted its preregistered proxy, but did not demonstrate better grant prose. It localized the obstruction in the evaluator: grammaticality cannot be represented adequately by a case-sensitive phrase count. A stronger realization layer should compile open decisions from typed fields into grammatical clauses and should be assessed by independent readers, while the scientific admission firewall remains unchanged.

**Experiment: GLP1–GRANT-4.4: controlled paraphrase**

**Input:** one frozen 67-slot program realized by exact copy, constrained paraphrase, and unconstrained fluent-prose arms.

**Model:** two temperature-zero Qwen3-Next-80B generation calls and one blinded same-model readability diagnostic; no repair calls.

**Formal result:** CONTROLLED passed 5/5 registered gates at 1,456 words; COPY remained admitted; FLUENT was rejected at 1,822 words for length and semantic accretion.

**Secondary diagnostic (not an admission gate):** 21/24 controlled changes gamed a case-sensitive defect proxy through lowercasing; only three changes constituted substantive grammatical repair.

**Interpretation:** the experiment identifies an evaluator obstruction, not successful rhetorical realization, biomedical validation, peer review, or evidence of fundability.

**Boundary**

**Verdict ledger.** GRANT-0, -1, and -2 were not admitted. GRANT-3 admitted one scientific-program representation relative to its typed IR, not relative to causal or peer review. GRANT-4 through -4.2 were not admitted. GRANT-4.3 admitted exact transcription, not rhetorical improvement. GRANT-4.4 passed its formal proxy but is substantively negative because that proxy was gamed. No rung establishes biomedical validity, independent readability, peer-review quality, or fundability.

*17.11.10 From an AI-authored study to an AI-constructed program*

The Open Conference of AI Agents for Science provides a useful contemporary contrast. Its inaugural call asked AI systems to lead hypothesis generation, experimentation, and manuscript writing, and required an AI system to be the sole first author. Submitted papers were evaluated for familiar research-paper qualities including quality, clarity, significance, and originality, with explicit disclosure of the relative AI and human contributions [[Open Conference of AI Agents for Science, 2025](#)]. This is an important test of whether an agent can produce a credible *completed unit of research* and report it in the conventional scientific form.

A complementary benchmark from Inherent Laboratories makes replication itself the training target. Its *Replica* task space contains 310 tasks drawn from 100 machine-learning and AI-for-science papers. An agent must reconstruct a reported figure under bounded time and compute, without access to the original plot. The authors train a 27-billion-parameter agent, Faraday, by long-horizon reinforcement learning with coding agents as tools, and report stronger held-out replication perfor-



mance than the frontier-agent baselines in their study [Falck et al., 2026]. This is more than document imitation: underspecified implementation choices must be recovered through hypothesis-driven experimentation.

Faraday nevertheless occupies a different point in the scientific workflow from the capstone developed here. Its registered artifact is a faithful reconstruction of an existing empirical result. GLP1–GRANT instead asks for a prospective theory of which unresolved claims deserve new evidence, how several aims should depend on one another, and why the proposed program merits scarce resources. A mature system could compose the two capabilities: theory construction proposes and justifies an experiment; a replication-trained agent executes or audits the computational protocol; the returned evidence then drives another round of theory repair.

The GLP1–GRANT capstone shifts the target from a completed study to a prospective research program. The central question is not only whether an agent can select a tractable hypothesis, execute experiments, and narrate the results. It is whether the agent can identify a consequential unresolved question, organize multiple dependent aims around it, anticipate results that do not yet exist, and justify committing scarce human and material resources to the program.

| AI-authored research paper            | AI-constructed research program                            |
|---------------------------------------|------------------------------------------------------------|
| Reports completed experiments         | Justifies experiments not yet performed                    |
| Demonstrates a result                 | Identifies a consequential unresolved question             |
| Optimizes one investigation           | Coordinates several dependent aims                         |
| Evaluates an artifact retrospectively | Evaluates importance, feasibility, and risk prospectively  |
| Provides evidence for a claim         | Argues why acquiring particular evidence merits investment |

These are complementary tests of AI for science. The distinction concerns the target artifact, not a hierarchy of scientific value.

The two capabilities are complementary rather than competing. An admitted research program could hand its experimental aims to a Robot Scientist or an Agents4Science-style execution system. The resulting measurements and papers would then return as evidence for the next round of theory repair. This closes a larger loop from literature, through prospective program construction, to experiment, publication, and revised theory. In compact form, Agents4Science asks whether an AI agent can conduct and report a study; the GLP1–GRANT capstone asks whether it can construct a credible future course for a field.

### 17.12 *Admission beyond the corpus*

Text can support provenance and consistency checks. Causal identification and claims about biological mechanisms require experimental or clinical evidence that may lie outside the closed corpus.

This limitation produces a two-stage admission design. Corpus-relative admission can certify provenance, extraction reproducibility, compatibility under the registered observers, and performance on locked documentary consequences. Scientific admission requires locked experimental, clinical, or simulator evidence appropriate to the claim. If that evidence is unavailable, AGENTIC must terminate with a proposed theory and experiment rather than manufacture closure.

### 17.13 *Toward an integrated open-domain benchmark*

AGENTIC-CORPUS-0 uses a temporally cut, closed GLP-1 corpus with disjoint development and confirmation sources. It withholds a context-dependent mediator, the evidence-reconciliation workflow needed to represent it, and the query policy that efficiently locates the decisive sources. A locked evaluator contains later documents or structured evidence not available during construction.

The component factorial tests whether CLIC alone proposes unsupported latent variables, whether OPTIC alone improves extraction without producing a theory, and whether RELIC alone retrieves more documents without resolving their types. The full system must construct a source-bearing theory extension and a discriminating external experiment. Locked documents confer only corpus-relative credit; causal identification requires appropriate experimental, clinical, or simulator evidence. Otherwise the correctly calibrated output is a proposal with preserved alternatives.

**Experiment: AGENTIC-CORPUS-0**

**Epistemic status:** proposed integrated benchmark, not a completed AGENTIC result.

**World:** temporally cut scientific corpus with heterogeneous evidence, contextual disagreement, and locked confirmation sources.

**Joint construction:** source-bearing theory extension, typed evidence workflow, and adaptive acquisition or experimental plan.

**Admission:** provenance and reproducibility first; then locked experimental, clinical, or simulator evidence where identification is claimed.

**Controls:** component factorial, flat retrieval-and-generation agent, admission-record ablation, and source, mechanism, and query-policy oracles.

**Boundary:** coherent text synthesis and citation agreement do not identify a causal mechanism.

**Challenge**

Can a system convert heterogeneous scientific appearances into a theory whose missing variables and experimental consequences are explicit?

The following visual testbed changes the medium but preserves the structural problem. A generated image is a model of a prompt-conditioned visual theory; violations of cardinality, role, layout, or identity reveal gaps between that model and the declared sketch. This makes structural diagnosis directly visible before the book asks the harder question of inventing a reusable visual technique.

*Further Reading*

McCarthy's distinction between appearance and reality motivates theory construction that does not simply aggregate sense data or text [McCarthy, 2007]. Categories for AGI develops Democritus, categorical databases, and related source-to-structure machinery [Mahadevan, 2026b]. Infinitesimal Causality and BRIDGE/SKFM supply the tangent and Lie-bracket tools used to localize causal inadequacy [Mahadevan, 2026c,a]. Modern biomedical agents such as Biomni illustrate the expanding tool and database substrate for scientific work [Huang et al., 2025]. AGENTIC-CORPUS adds the requirement that theory, workflow, and acquisition-policy revisions retain distinct status-bearing admission records and explicit unsupported alternatives.



## Generative Visual Languages

### ARTISTIC: ACQUIRING REPRESENTATIONAL THEORIES THROUGH INFINITESIMAL SEMANTIC TRANSPORT IN IMAGE CREATION

Text-to-image diffusion makes the tension at the center of this book visible. A model generates images never observed during training by following many small denoising steps. The result may be novel and aesthetically valuable, yet the language controlling that process remains largely inherited: a fixed tokenizer, text encoder, conditioning interface, and repertoire of prompt constructions. Generative novelty is therefore not automatically representational creativity.

This chapter introduces *ARTISTIC: Acquiring Representational Theories through Infinitesimal Semantic Transport in Image Creation*. ARTISTIC specializes the AGENTIC contract to visual languages. It couples the image-noise geometry of diffusion to a second direction in which the conditioning theory itself can be diagnosed, extended, and independently admitted.

The evidence reported here realizes only bounded parts of that architecture. It progresses from controlled lexical and part-structure calibration to estimated declarations, uncertainty-aware acquisition, and theory-guided repair of generated photographs. Together these experiments supply evidence for an operational substrate for visual-language construction; they do not yet establish an abstract double involution algebroid or unrestricted invention of visual objects, relations, or artistic techniques. Chapter 19 takes up that stronger target explicitly.

The artifact dossier makes the boundary especially visible here. A repaired image is an operational output; a visual-language extension is a structural delta; masks, prompts, relations, and protected content form its transport record; held-out observers and abstention supply admission evidence; and the versioned declaration makes the result replayable. Most natural-image results below admit repairs relative to a supplied or estimated declaration. They do not thereby admit a nontrivial extension of the visual language.

Accordingly, “ground truth” in this chapter is not one privileged

target image. In controlled studies it is a registered scene graph, mask, relation, or formation rule. In natural-image studies it is an independently supplied annotation or structural observer, with uncertainty retained when the image does not determine a unique judgment. A repair can satisfy that reference and preserve protected content without being the only acceptable image or the most aesthetically valuable one.

### 18.1 Diffusion is one infinitesimal direction

In the continuous-time formulation, a forward diffusion satisfies

$$dX_t = f(X_t, t) dt + g(t) dW_t,$$

and gradually transports a data distribution toward a tractable noise distribution. Its reverse-time SDE depends on the time-indexed score  $\nabla_x \log p_t(x)$ . A text-conditioned model replaces this by a conditional score such as  $\nabla_x \log p_t(x | \ell)$ , where  $\ell$  is a prompt or encoded semantic state.

The associated probability-flow ODE has velocity

$$v_t^\ell(x) = f(x, t) - \frac{1}{2}g(t)^2 \nabla_x \log p_t(x | \ell).$$

Under suitable regularity, its flow maps  $\Phi_{s,t}^\ell$  are invertible and satisfy

$$\Phi_{t,u}^\ell \circ \Phi_{s,t}^\ell = \Phi_{s,u}^\ell, \quad (\Phi_{s,t}^\ell)^{-1} = \Phi_{t,s}^\ell.$$

This is the cleanest entry point for groupoid semantics: the pair groupoid of noise times acts through a family of conditioned transport maps on image states. The construction is conditional on existence, uniqueness, and regularity of the flow; a well-defined evolution of densities alone does not guarantee a pointwise invertible transport.

The stochastic statement is different. Markov transition kernels generally have no inverse and therefore do not form a groupoid. Under stronger smoothness conditions an SDE may generate a pathwise stochastic flow of diffeomorphisms, but its infinitesimal generator is a second-order differential operator rather than an ordinary vector field. ARTISTIC consequently begins with the probability-flow ODE as its classical Lie-theoretic realization and treats a fully stochastic groupoid or algebroid semantics as an open extension.

*Chapter roadmap.* The chapter states the ARTISTIC architecture, calibrates lexical and structural declarations, and then tests visual repair and concept induction. It closes by separating demonstrated assurance mechanisms from the remaining measurement and mathematical obligations.

#### Boundary

The reverse-time SDE reconstructs a data distribution from a noise distribution; it is not generally a sample-by-sample inverse of the particular Brownian path that destroyed an image. The deterministic probability-flow ODE and a pathwise stochastic flow are stronger, distinct constructions.

### 18.2 The second direction: proto-accommodation and semantic extension

Let  $\mathbb{L}_{\text{vis}}$  be a visual-language sketch. Its sorts may include *object*, *attribute*, *relation*, *scene*, and *transformation*; its arrows encode composition, binding, part-whole maps, viewpoint, and rendering constraints. A model of  $\mathbb{L}_{\text{vis}}$  is a semantic scene. A rendering semantics assigns not one canonical image but a distribution:

$$\mathcal{F}_\theta : \text{Mod}(\mathbb{L}_{\text{vis}}, \mathcal{C}) \longrightarrow \text{Prob}(\mathcal{X}).$$

Ordinary conditional generation moves horizontally through image-noise space while holding  $\mathbb{L}_{\text{vis}}$  fixed. ARTISTIC adds a vertical, proto-accommodative direction that probes whether a visual primitive, relation, formation rule, binding law, or conditioning interface is missing. Infinitesimal language probes vary the current semantic realization locally. If they motivate a finite proposal, the proposed semantic extension is recorded as

$$\mathbb{L}_{\text{vis}} \longrightarrow \mathbb{L}_{\text{vis}}^+.$$

An actual new word or grammatical constructor becomes accommodation only after this extension and its transported models are independently admitted.

Let  $\tau_u : \ell \rightarrow \ell_u$  be a small language-side transport and let  $\hat{\tau}_u^t : \mathcal{X}_t^\ell \rightarrow \mathcal{X}_t^{\ell_u}$  denote its registered realization on image states at noise time  $t$ . The mixed square then compares

$$\Phi_{s,t}^{\ell_u} \circ \hat{\tau}_u^s \quad \text{with} \quad \hat{\tau}_u^t \circ \Phi_{s,t}^\ell.$$

The comparison need not be a pixelwise subtraction. It is a typed mate whose realization may be scored through object, relation, and rendering observers. Its defect asks whether changing the intended semantics before denoising agrees with transporting those semantics through an existing denoising flow.

### 18.3 A candidate double geometry

The proposed double object has the following interpretation.

|               | Base structure                   | Local directions                                    | Finite realization                                                      |
|---------------|----------------------------------|-----------------------------------------------------|-------------------------------------------------------------------------|
| Image side    | Noise-indexed image state        | Score or probability-flow vector fields             | Conditioned generative flow groupoid                                    |
| Language side | Visual theory and semantic model | Concept, relation, binding, or interface variations | Visual-sketch extension and semantic transport                          |
| Core          | Shared image-language grounding  | Mixed binding and relational directions             | Admitted correspondence between language construction and visual effect |

If both sides integrate and satisfy the required interchange laws, their square structure is a candidate double groupoid whose infinitesimal object should be compared with DIAL. That conclusion is a research target, not an assumption. In particular, the stochastic side may require a second-order or Markov-enriched generalization of involution algebroids.

The core is experimentally concrete. Image geometry can expose a direction in which pixels or latents change without saying what the change means. Language geometry can distinguish symbols without guaranteeing a controlled visual effect. Their mixed core contains grounded attribute binding, relational composition, viewpoint-dependent identity, part-whole structure, and other effects that are invisible from either side alone.

#### 18.4 *From a new token to a new language*

ARTISTIC separates four claims that are easily collapsed.

1. **Lexical abbreviation.** A new token reproduces an image or a meaning already expressible by an existing prompt.
2. **Concept naming.** A new token names a stable visual distinction already representable by the frozen generator.
3. **Productive extension.** A new symbol or relation composes systematically with previously unseen objects, attributes, and scenes.
4. **Transformational extension.** A new grammatical constructor or visual ontology requires an extension of the conditioning or generative model.

Textual inversion provides an important baseline for the first two levels: it learns new word embeddings from a small set of images while



freezing the text-to-image model. ARTISTIC claims more only when the construction has a declared type, compositional laws, counterfactual visual consequences, and transport beyond its defining examples. An arbitrary private code or adversarial embedding is not a language merely because it controls a model.

#### Design principle

A visual word is admitted through its productive use: what it denotes, how it combines, which controlled changes it causes, what it leaves invariant, and whether another reasoner can learn and reuse it.

### 18.5 ARTISTIC as an AGENTIC system

The three component workflows receive distinct contracts.

*CLIC.* Use CLIC, or a noncausal structural analogue, to intervene on prompts, seeds, image regions, attention bindings, and rendering controls. The purpose is to localize a stable semantic failure. Unless a model-internal causal semantics is declared, this remains structural diagnosis rather than a claim about causality in the depicted world.

*OPTIC.* Construct a token, relation, prompt macro, semantic mediator, embedding, adapter, or control program. Its artifact must declare input and output types, scope, invariants, and executable tests.

*RELIC.* Select contrastive prompts and generations that discriminate candidate meanings, allocate seeds and rendering budgets, and learn when the admitted construction should be deployed.

Their structural packets preserve the visual representational package  $\mathfrak{R}_{\text{ART}}$ : the language sketch, tokenizer and encoder, generator and renderer interface, visual observers, admission rules, and proposal grammar. A proposed comparison

$$Y_{\text{ART}} : \mathfrak{R}_{\text{ART}} \longrightarrow \mathfrak{R}_{\text{ART}}^+$$

also carries the localized binding defect, contrastive queries, control record, rendering artifacts, transport, and separate semantic and visual certificates; a visual-sketch map is only its presentation component. The applicable control suite  $\text{Ctl}_{\text{ART}}$  includes no edit, stronger prompting or regeneration under the fixed declaration, localized and global edits, capacity-matched untyped repair, matched-random regions or probes, observer and viewpoint alternatives, and an oracle declaration where available. A successful image does not automatically admit the proposed meaning; a coherent meaning does not guarantee that the

generator can render it. Most natural-image repairs below keep  $Y_{\text{ART}}$  equal to the identity and construct an operational repair inside the maintained package.

### 18.6 *The ARTISTIC procedure*

An initial implementation follows eight stages.

1. **Declare.** Freeze the base vocabulary, tokenizer, encoder, generator, permitted adapters, visual observers, and reference grammar.
2. **Probe.** Generate matched prompt, seed, and image interventions that isolate an object, attribute, relation, or binding failure.
3. **Diagnose.** Estimate image-side, language-side, and mixed defects, including uncertainty and unsupported alternatives.
4. **Propose.** Select a bounded operational repair or package comparison. When the visual presentation changes, state which previously inexpressible distinction its extension is meant to capture.
5. **Realize.** Use OPTIC to construct its embedding, adapter, recognizer, rendering control, and compact semantic specification.
6. **Explore.** Use RELIC to select held-out compositions and counterexample generations rather than resampling successful prompts.
7. **Admit.** Test grounding, disentanglement, productivity, systematicity, intersubjectivity, and transport.
8. **Persist.** Version the repair or package comparison, examples, counterexamples, model dependencies, and later revisions.

The proposal language must be large enough to add relational constructors, not only embeddings. Conversely, it must remain bounded enough that a failed experiment can distinguish invention from unconstrained fine-tuning.

### 18.7 *Textual inversion as staged calibration*

The Diffusers implementation of textual inversion provides a reproducible bridge from the preceding distinctions to an executable experiment [Hugging Face, 2026]. Its training script adds one or more placeholder tokens, initializes their embeddings from a registered token, and updates only those embeddings against a small image collection. The tokenizer, text encoder apart from the new rows, variational autoencoder, denoiser, noise schedule, and base model can therefore be frozen explicitly. This is unusually valuable for ARTISTIC: the intervention

is small enough to state exactly what representational structure the learner was permitted to change.

The initial experiments use textual inversion as a staged calibration, not merely as a competitor. Their identifiers use TI for textual-inversion calibration; the later natural-image studies use NI.

*ARTISTIC-TI-0: concept naming.* Give the learner a few views of a novel procedural object or style and test whether a nonce token retrieves it under held-out backgrounds, viewpoints, colors, and prompts. This condition is designed as a positive control for lexical abbreviation and concept naming; the substantive result below shows that the intended control was not admitted.

*ARTISTIC-TI-1: productive composition.* Hold out combinations of the learned concept with registered objects, attributes, counts, and spatial relations. Compare one learned vector with a matched multi-vector representation. Success requires preservation of both the new concept and every independently varied registered factor.

*ARTISTIC-TI-2: relation naming.* Replace the object target by a nonce binary relation whose instances vary over object identity, order, viewpoint, and scene layout. The question is whether a token embedding can express a constructor rather than memorize a visual template. Counter-witness prompts reverse the arguments, preserve the objects while removing the relation, and instantiate the relation on unseen object pairs.

*ARTISTIC-TI-3: candidate extension.* If TI-2 fails under adequate coverage, uncertainty, and optimization controls, freeze that failure as an obstruction. ARTISTIC may then propose a bounded typed extension—for example a relation-specific mediator, attention control, or adapter—and must demonstrate that the added structure, rather than extra training budget, supplies the missing productivity.

The comparison includes existing-token paraphrases, prompt search, a randomly initialized frozen token, textual inversion with matched compute, and a registered adapter oracle. Evaluation uses scene-graph truth in procedural worlds: concept fidelity, argument order, exact relation satisfaction, preservation of irrelevant factors, held-out composition, cross-seed transfer, and false activation. Every run records the base-model revision, Diffusers revision, initialization token, number of learned vectors, image and prompt splits, seeds, checkpoints, and compute budget.

## Experiment: ARTISTIC-TI-0-3

**Epistemic status:** consolidated evaluation protocol distilled from the TI calibration ladder; the card defines the stronger language-extension test rather than claiming that it has been passed.

**Question:** when does adding a visual “word” cease to be embedding acquisition and require an extension of the visual language itself?

**Positive control:** few-shot naming of a novel object or style with a frozen text-to-image model.

**Boundary challenge:** a nonce relation applied to unseen ordered object pairs and scene configurations.

**Accommodation trigger:** a preregistered productivity defect that persists across seeds, vector counts, matched compute, and counter-witness acquisition.

**Admission:** the proposed extension passes held-out composition and counterfactual tests without degrading the frozen vocabulary.

**Interpretive limit:** a successful personalized token is evidence of concept acquisition; by itself it is not evidence that a new grammar or visual ontology has been invented.

### 18.8 *A visual-language benchmark*

ARTISTIC-T2I-o uses procedurally generated scenes with a frozen base language of objects, attributes, and elementary relations. The evaluator withholds one relation or composition rule while retaining primitive rendering operations from which it can be realized. Nonce names and generated geometries prevent retrieval of an existing natural-language label.

Two conditions separate different creative claims.

*Latent but unnamed.* The frozen generator can already realize the withheld distinction, but the conditioning language cannot reliably express it. The target is productive language acquisition without changing the image model.

*Structurally unavailable.* The base generator cannot realize the distinction under any registered prompt. The target is a candidate finite extension through a bounded adapter or control-field extension together with a new language construction; the change is retained only after independent admission.

The new symbol is evaluated on unseen objects, colors, counts, styles, spatial contexts, and grammatical positions. A communication split gives its compact definition and a few admitted examples to a

fresh interpreter, which must use it in new requests. A transport split changes seeds and renderers and, where interfaces permit, the underlying diffusion model.

#### Experiment: ARTISTIC-T2I-0

**Epistemic status:** proposed integrated benchmark, not a completed visual-language result.

**World:** procedural visual scenes with a frozen vocabulary and one withheld relation or composition rule.

**Joint construction:** visual-theory extension, executable grounding or adapter, and an active contrastive-generation policy.

**Admission:** relation accuracy, counterfactual grounding, disentanglement, unseen composition, fresh-interpreter communication, and cross-seed and cross-renderer transport.

**Controls:** prompt search, textual inversion, fixed LoRA or adapter, structured attention guidance, the CLIC–OPTIC–RELIC factorial, and semantic, rendering, and query-policy oracles.

**Boundary:** image novelty or a learned token alone is not an admitted visual language.

### 18.9 *An evidence architecture for visual assurance*

The preceding benchmark describes the intended destination. The evidence reported here isolates narrower obligations: whether a frozen diffusion system can maintain, diagnose, and extend a small typed visual declaration as less of that declaration is supplied. This is not a chronological claim that each later experiment is globally stronger. It is an architecture of evidence in which distinct controls answer distinct questions.

Seven obligations organize the architecture:

1. learn a name for a registered visual concept;
2. represent that concept as a typed object with parts and relations;
3. localize a structural counter-witness and repair only its support;
4. estimate the declaration rather than receiving its exact mask;
5. transfer the procedure to photographs evaluated by sealed annotations;
6. represent an unreliable relation by an interval and abstain;
7. acquire another view only where present evidence is insufficient.

The dependencies among these obligations matter. Exact masks isolate the repair operator before declaration error is introduced. Estimated masks test whether the declaration can be recovered. Natural photographs test transport outside the procedural renderer. A failed regime is retained as a typed obstruction; another regime changes the declaration or repair language rather than retroactively weakening its threshold.

The following table separates textual-inversion calibration from natural-image transport.

| Phase    | Maintained structure                     | Obstruction                                                   | Accommodation                                       |
|----------|------------------------------------------|---------------------------------------------------------------|-----------------------------------------------------|
| TI-0-1   | learned token and replay packet          | naming does not imply productive structure                    | separate lexical acquisition from typed realization |
| TI-2-3D  | exact synthetic parts and relations      | broad retraction damages structure                            | adaptive admission and counter-witness repair       |
| TI-4A-4E | estimated synthetic declaration          | source evidence and generated change need different observers | compose narrow declaration with paired observation  |
| NI-0-1   | one natural subject and boundary         | total area confounds identity with repair cost                | marginal-area admission against human trimaps       |
| NI-2A-2D | people, articulated parts, and relations | occlusion makes point area claims brittle                     | intervals, abstention, and active acquisition       |

Table 18.1: The empirical ladder from lexical calibration to natural-image transport.

### 18.10 Lexical calibration and typed realization

ARTISTIC-TI-0 first supplied an infrastructure positive control: the official textual-inversion loop trained, serialized, reloaded, and replayed a nonce embedding in a fresh process. Its substantive Stable Diffusion 1.5 run was *not admitted*. Four learned vectors produced partial in-domain binding, but the registered natural-language paraphrase was stronger and the learned token failed the background- and style-transport thresholds. The result is therefore evidence that lexical training was executable, not that concept naming or a new visual grammar had succeeded. The initial two-direction realization likewise failed all four preregistered clauses: a learned vector and selective replay were not enough to make the relevant structure explicit.

Typed realization replaces holistic visual similarity with an explicit declaration. The synthetic artifact is represented by a body, annulus, fins, their incidence, and a protected exterior. ARTISTIC-TI-2 showed that no fixed img2img strength simultaneously preserved the declaration and produced the requested exterior change. Adaptive admission selected the first strength satisfying the structural contract for each cell. TI-2C was the first admitted result; TI-2D repeated that contract satisfaction on textured images rather than only clean controls.

The remaining question is where a failed cell should be repaired.

TI-3, TI-3B, and TI-3C compare latent exterior masking, explicit scalar retraction, and componentwise retraction; each exposes a more localized obstruction. TI-3D restores the admitted parent only on typed precision and recall counter-witness sets. All 24 cells required halo width zero: the witness pixels themselves were sufficient.

The admitted TI-3D repair achieved structural score 0.820, a 100% strict-contract rate, and prompt similarity 0.362. Its minimal object core occupied 12.97% of the image and added counter-witness repair only 0.55%. Total retraction area was 70.43% of a registered blanket alternative, below the frozen 75% ceiling. The benefit was not merely that projection worked; the obstruction localized where projection was needed.

### 18.11 *Learning the declaration*

Exact masks make a clean calibration but hide the first half of the problem. TI-4 removed them from generation and repair. A declaration observer had to infer the artifact and its components from the source image, while a held-out observer evaluated the generated result.

TI-4A showed that one color-and-connectivity rule did not transfer across unseen shapes. TI-4B's proximity graph could not certify enough declarations to justify repair. TI-4C passed complete coverage on only four of six shapes. TI-4D separated source declaration from paired semantic change and raised this to five of six, localizing the remaining failure to thin concave support.

TI-4E admitted the full chain. A magenta annulus anchored a local object field; all cyan support inside that field was retained without convex filling, and a narrow source fin core was separated from the broader paired observer used on generated images. On a fresh six-shape cohort, all six preregistered clauses passed: mean declaration fidelity 0.979, minimum component F1 0.870, repaired held-out structure 0.965, 100% strict passage, no appearance regression, and area ratio 0.588.

This three-level organization is central: source evidence declares what is to be preserved; paired observation detects newly generated component-like material; localized repair acts on the resulting counter-witness. Collapsing these roles caused the complementary failures of TI-4B and TI-4C.

### 18.12 *Transfer to natural photographs*

Natural-image transfer uses two evidence regimes. NI-o uses independently annotated pet photographs and passes six of seven clauses, but fails its area criterion because total protected area was dominated by the mandatory subject. NI-1 preregistered the decision-relevant marginal

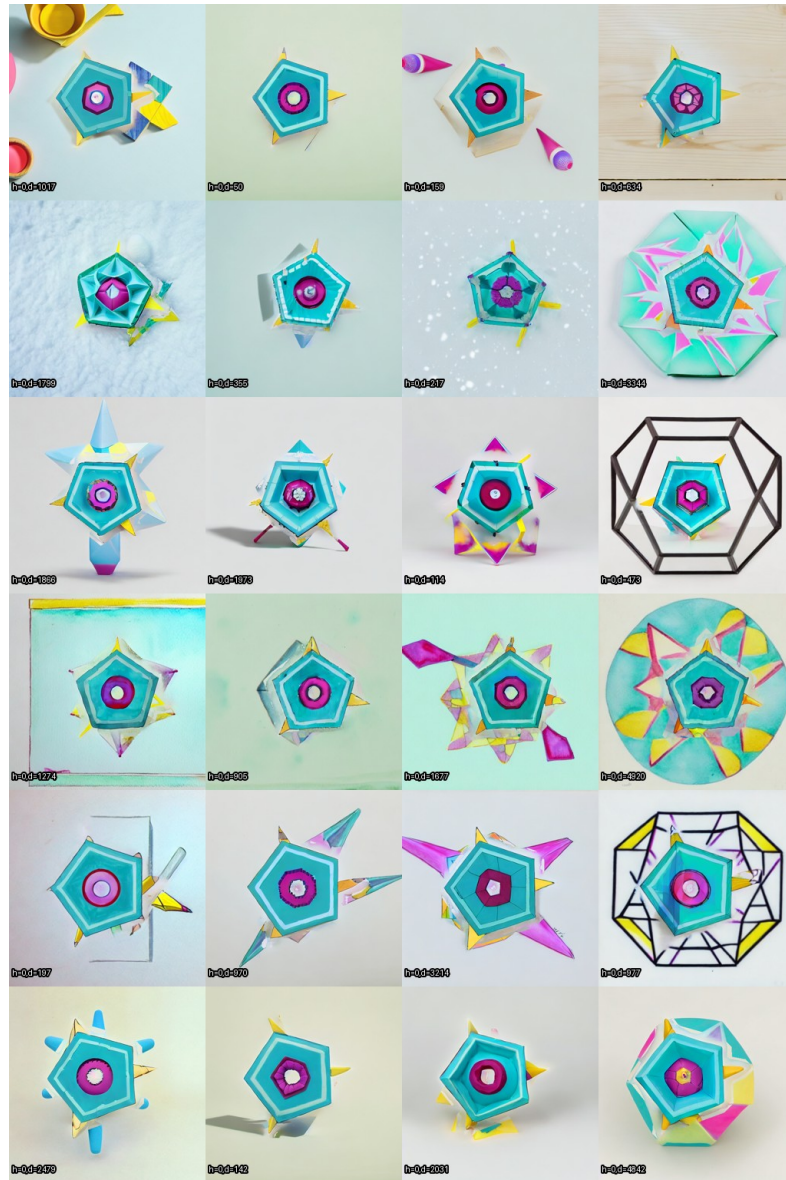


Figure 18.1: ARTISTIC-TI-3D under varied prompts, styles, and backgrounds. The registered body, annulus, and fins remain recognizable while the exterior is transported. This is structural repair, not yet evidence for an autonomously invented visual language.



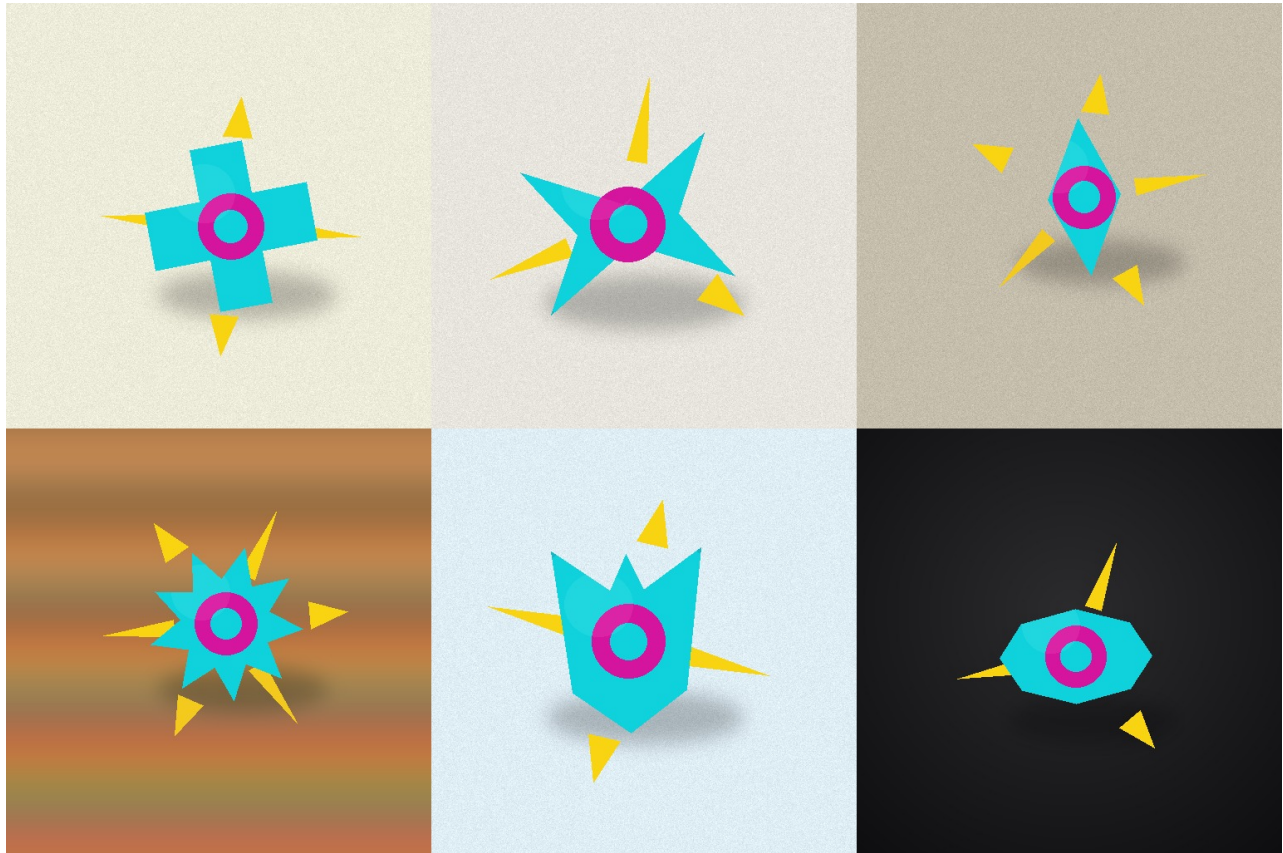


Figure 18.2: Estimated declarations in ARTISTIC-TI-4E across cross-shaped, bow-tie, thin-diamond, lobed, forked, and flattened supports. Separating narrow declaration evidence from broader paired observation meets the final registered synthetic-declaration gate.

statistic

$$\rho_{\text{marginal}} = \frac{A_{\text{localized}} - A_{\text{mandatory}}}{A_{\text{blanket}} - A_{\text{mandatory}}}.$$

On six previously unseen Oxford-IIIT Pet subjects and eighteen generated cells, all seven clauses passed. Against sealed human trimaps, localized repair raised structural score from 0.557 to 0.642, reduced subject-region RGB error from 0.0459 to 0.0020, and used about one third less added area than blanket repair.

NI-2 replaces the single foreground object by typed relational diagrams. Each COCO scene supplied person instances, grouped articulated parts, part-whole incidence, identity order, centroid geometry, and relative visible area. Public declarations were produced by frozen Mask R-CNN and Keypoint R-CNN observers; COCO polygons and keypoints remained sealed until generation and repair were complete.

A robot-baseball role testbed illustrates the declaration principle. A finite role sketch specifies four offensive players, nine defenders, and one official before image realization. Unlike a verbal prompt, the diagram makes cardinality, role, and location constraints explicit and provides a common interface for generation and audit.

NI-2A passed all ten clauses as a two-person calibration: declaration mask IoU was 0.867, human-keypoint PCK 0.984, and all twelve repaired cells preserved pair geometry. Its metric conventions were completed after public outputs existed but before annotation scoring, so a confirmatory claim was reserved for a fresh cohort.

NI-2B froze the complete contract in advance and introduced three people, partial poses, and occlusion. Eight of eleven clauses passed. Identity order was preserved in every cell and articulation remained strong, but relative visible-area relations failed. The experiment rejected the assumption that an observer's point estimate remains reliable when a person is truncated, occluded, or embedded in a crowd.

### 18.13 Intervals, abstention, and active acquisition

NI-2C changed the type of the relation rather than weakening its threshold. If person  $i$  has  $k_i$  confident keypoints, its public log-area uncertainty is

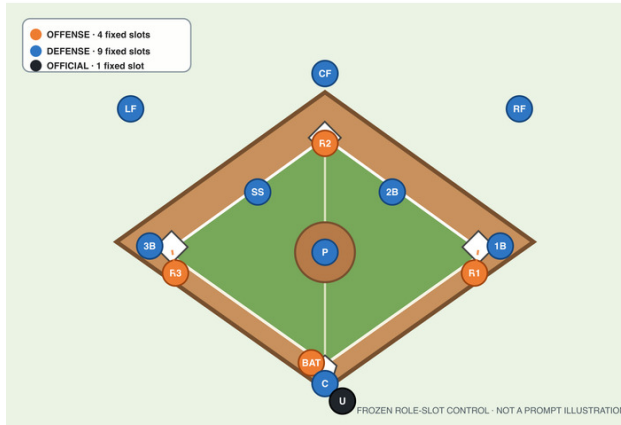
$$r(k_i) = 0.25 + 0.05(17 - k_i).$$

For reference person 0, the relative log-area interval is

$$I_{i0} = \left[ \log \frac{\hat{A}_i}{\hat{A}_0} - r(k_i) - r(k_0), \log \frac{\hat{A}_i}{\hat{A}_0} + r(k_i) + r(k_0) \right].$$

A categorical claim is emitted only when the entire interval lies beyond the registered indifference band:  $\inf I_{i0} > 0.15$  certifies *larger*,  $\sup I_{i0} < -0.15$  certifies *smaller*, and every other case abstains.

## FROZEN ROLE-SLOT CONTROL



## L01



## L02



## L03



Figure 18.3: A typed visual declaration in the robot-baseball testbed. The frozen role-slot sketch at upper left specifies four offensive roles, nine defensive roles, and one official. Three robot-only realizations test whether that same declaration can be transported across generated scenes without changing its cardinality or incidence constraints.

This repaired calibration: all eight sealed annotation-derived ratios and all twenty-four output ratios lay in their intervals, and no certified claim was false. But only one of eight comparisons was certified. NI-2C therefore passed twelve of thirteen clauses and failed the frozen 25% claim-coverage requirement. Reliable ignorance was preferable to false precision, but was not yet useful enough.

| Interval relation                          | Emitted claim         | Epistemic reading                                                            | Permitted action                                                   |
|--------------------------------------------|-----------------------|------------------------------------------------------------------------------|--------------------------------------------------------------------|
| $\inf I_{i0} > 0.15$                       | person $i$ is larger  | The complete uncertainty interval clears the indifference band.              | Preserve the certified ordering through repair.                    |
| $\sup I_{i0} < -0.15$                      | person $i$ is smaller | The complete uncertainty interval clears the band in the opposite direction. | Preserve the certified ordering through repair.                    |
| $I_{i0} \cap [-0.15, 0.15] \neq \emptyset$ | abstain               | Available evidence does not determine a stable categorical relation.         | Acquire another registered view or leave the relation unspecified. |

NI-2D made acquisition conditional on that ignorance. When a base interval abstained, the observer acquired a horizontally reflected view, inverse-mapped its person masks, and tested agreement in the original coordinates. An interval narrowed only when both relevant masks had IoU at least 0.70 across views and absolute log-area change at most 0.15. The fused radius used the preregistered sharper rule

$$r_{\text{active}}(k) = 0.18 + 0.03(17 - k).$$

Disagreement left the original interval untouched. The second view is a registered equivariance probe of the same frozen segmenter, not an independent semantic observer.

| Stage                          | Admission            | Annotation coverage | Claim precision | Claim coverage |
|--------------------------------|----------------------|---------------------|-----------------|----------------|
| NI-2B: point relation          | 8/11 clauses         | not represented     | brittle         | incomplete     |
| NI-2C: interval and abstention | 12/13 clauses        | 100%                | 100%            | 12.5%          |
| NI-2D: active second view      | <b>13/13 clauses</b> | 100%                | 100%            | 25.0%          |

Seven of NI-2D's eight base comparisons triggered acquisition. Five passed the agreement gate and were narrowed; two retained wide intervals because the reference-person observation was inconsistent. Mean half-width fell from 0.9375 in NI-2C to 0.7550, while sealed-annotation coverage remained 100%. Certified coverage doubled to two of eight comparisons, with precision still 100%. All thirteen preregistered clauses passed.

The same distinction between public observation and sealed structural audit can be made concrete in a robot-only testbed. In the robot-baseball example, each visible player is assigned a typed role and

Table 18.2: The NI-2C decision rule makes abstention part of the typed observer. The system may preserve a relation only when the entire interval supports it; a point estimate alone cannot authorize repair.

compared with the frozen role-slot declaration; the overlay exposes omissions, duplicates, and misplacements rather than scoring the scene as an undifferentiated image.

The generated-output audit was conjunctive. Before localized repair, instance IoU was 0.787, keypoint PCK 0.975, interval coverage 91.7%, and structural score 0.875. After repair these became 0.836, 0.997, 100%, and 0.912. Identity order and triangle geometry passed in all twelve cells. Subject RGB error fell from 0.0633 to 0.00329, and the incremental repair-area ratio was 0.191, below the frozen 0.75 ceiling.

A complementary pose experiment shows how such an audit becomes a localized repair. The first realization places the catcher in the wrong orientation; the second enforces the registered catcher pose; the third harmonizes that typed insertion with the surrounding image while preserving the other player roles and field layout.

#### Design principle

Active observation is creative only in a restricted sense: the system changes what evidence it seeks to decide whether its declaration can be sharpened. The key result is not that another view always helps, but that agreement licenses refinement while disagreement preserves abstention.

### 18.14 From image repair to structural assurance

The preceding evidence sequence controls the declaration, observer, and admission rule in increasingly realistic scenes. A complementary set of external challenge cases asks a more immediate systems question: what happens when ARTISTIC is placed around an existing image generator? The resulting role is broader than post-processing. ARTISTIC may admit a structurally valid image unchanged, repair a localized violation, request another observation, reject and regenerate a globally inconsistent scene, or abstain when the relevant fact is outside the frame.

This changes the central comparison. ARTISTIC is not proposed as a replacement for a more capable generator. It is a provider-neutral layer in which visual correctness can be declared, observed, and audited. Its marginal value depends on the base model, the domain, the specificity of the request, the cost of a failure, and the cost of another generation or edit.

#### 18.14.1 View-conditioned contracts

Let  $S = (c_1, \dots, c_m)$  be a visual sketch,  $I$  a generated image, and  $V$  an estimate of viewpoint, framing, and action phase. A structural observer



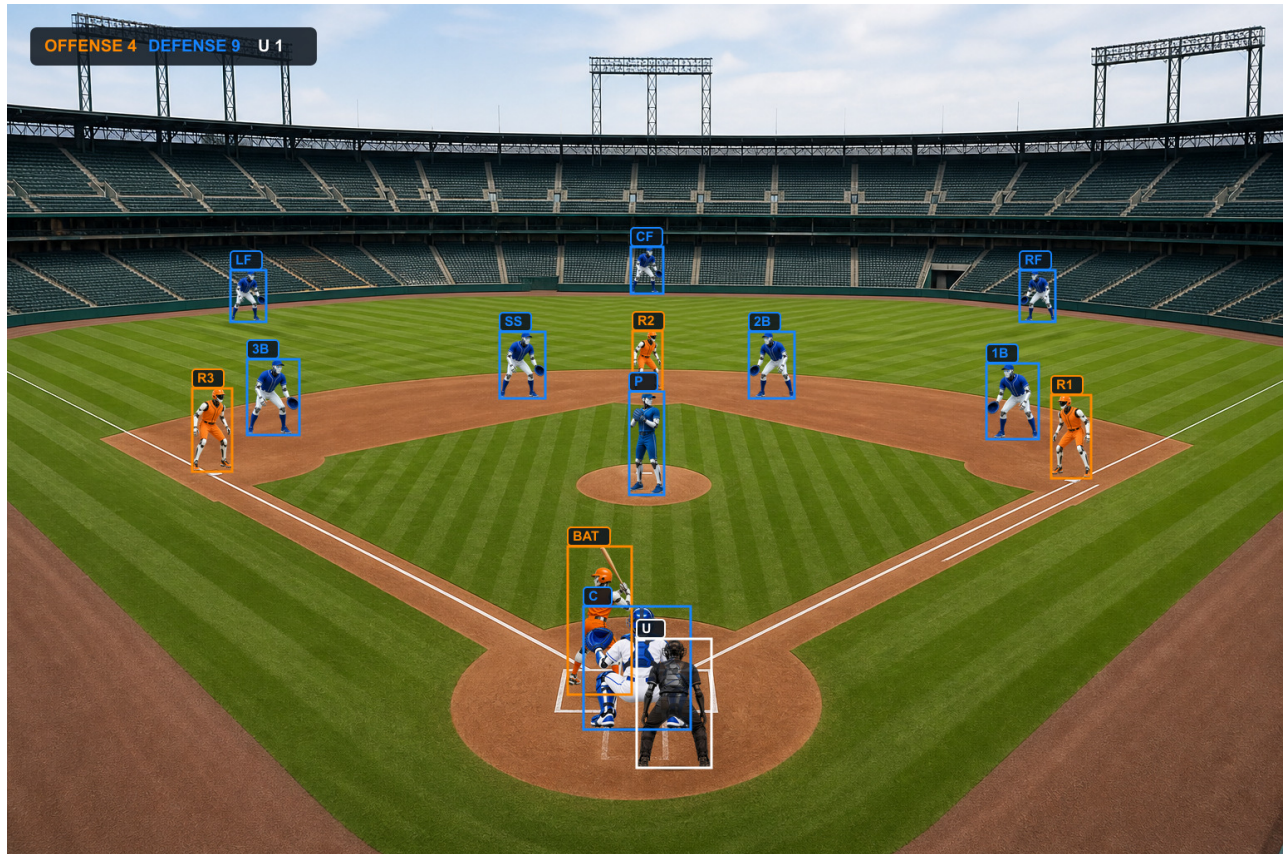


Figure 18.4: Typed structural audit in the robot-baseball testbed. Bounding boxes identify the batter, catcher, pitcher, umpire, three runners, and nine defenders by role. The audit tests cardinality, role identity, and spatial incidence rather than relying on holistic visual similarity.



Figure 18.5: Observer-localized pose repair using robot players. Left: the catcher faces the wrong direction. Center: the registered pose is enforced. Right: the inserted catcher is harmonized with the scene. The sequence makes the repair target explicit while leaving unrelated roles and geometry fixed.

should not force every clause into a Boolean judgment. It returns

$$O(I; S, V) \in \{\text{sat}, \text{viol}, \text{unobs}, \text{uncertain}\}^m.$$

The distinction between *viol* and *unobs* is essential. A close photograph of a cricket striker may coherently omit the bowler, non-striker, second wicket, and umpires because they lie outside the frame. A wide image explicitly promising that every active player is visible makes the same omissions genuine violations. Likewise, a ball hidden in a pitcher's glove may be uncertain rather than absent.

The corresponding controller is selective:

|                                             |               |                     |
|---------------------------------------------|---------------|---------------------|
| all required clauses observed and satisfied | $\Rightarrow$ | admit,              |
| localized supported violation               | $\Rightarrow$ | repair,             |
| global incompatibility                      | $\Rightarrow$ | regenerate,         |
| critical uncertainty or non-observability   | $\Rightarrow$ | acquire or abstain. |

An important negative-control metric follows immediately: a useful assurance system must leave valid images alone. Its false-repair rate is

$$\Pr\{\text{ARTISTIC edits } I \mid I \text{ satisfies the registered contract}\}.$$

#### 18.14.2 *A sealed Firefly cricket challenge*

##### Boundary

**What produced the repaired pixels?** A human supplied the cricket declaration and source-image diagnosis. ARTISTIC compiled them into one verbatim image-editing instruction specifying the required roles, cardinalities, spatial relations, and preservation constraints. A built-in black-box image editor then received the source image and that instruction. Exactly one call was made; the single output was retained regardless of outcome, with no candidate selection or residual tuning. The preregistered manual audit was performed at the clause level. Thus ARTISTIC supplied the theory, diagnosis, contract, and admission record; the editor supplied the pixels. No hidden diffusion trajectory was observed or altered.

Figure 18.14.2 records a one-pass external challenge using an image produced by Adobe Firefly 5. The original is aesthetically convincing and recognizably cricket-like, but it has no visible bowler, wicketkeeper, non-striker, or umpire; only one coherent wicket; an implausibly large ball; and an unexplained white cylinder on the pitch. Before editing, the experiment froze a contract requiring one striker, non-striker, overarm bowler, wicketkeeper, bowler's-end umpire, live ball, and two pitch-end wickets, while preserving the stadium, daylight, framing, and silver-robot style.

The single retained proposal appears to satisfy every primary visible obligation under the frozen manual audit. It constructs a bowler–ball–striker trajectory, places the wicketkeeper behind the striker’s wicket, separates the non-striker and umpire at the far end, adds the second wicket, rescales the ball, and removes the foreign object. The result is a strong structural repair, but it is not a masked local retouch: the editor recomposes much of the field. That preservation cost must remain part of the audit rather than disappearing behind the improved final image.

#### 18.14.3 *Three kinds of Nano Banana repair*

The Nano Banana Lite cases in Figure 18.14.3 separate three failure types that a scalar image-quality score would tend to collapse. In cricket, the objects are plausible but role binding is wrong: a bowler is missing, a second wicketkeeper-like figure occupies the bowling end, and the non-striker’s gesture mixes action phases. The retained reconstruction reassigns those roles into one delivery. In baseball, runners, bat contact, dust, and a misplaced coach combine incompatible moments. ARTISTIC chooses a pre-pitch bases-loaded state and reconstructs the diamond around that temporal declaration. In tennis, both opponents begin on the same side of the net, the principal ball intersects the robot’s strings, a second small ball is visible, and a holographic display substitutes for physical trajectory. The repair changes the opponent topology by transporting the human behind the net and restoring one-ball rally geometry.

These cases were edited with complete game-state declarations. They therefore demonstrate contract-guided proposal and admission, not autonomous acquisition of cricket, baseball, or tennis. Their importance is diagnostic: role binding, temporal phase, and relational topology require different observers even when all three failures are colloquially described as “bad generation.”

#### 18.14.4 *Strong generators change the opportunity*

Outputs obtained through the OpenAI Playground show why the ARTISTIC claim must remain provider-conditioned. A terse request for a robot playing tennis produced a coherent scene with two opponents separated by the net, one racket each, and one ball. A similarly terse pool request produced a plausible mid-game shot. Under their generic visible-scene contracts, the correct action is admission without editing.

The same interface also produced more subtle failures under stronger contracts. A wide cricket request asking that every player and both umpires be visible generated fewer than the required eleven fielders, placed both umpires side by side at the bowler’s end rather than assigning one to square leg, and appeared to attach the ball to a rod or



malformed object associated with an umpire. The nouns are present, but cardinality, spatial role, topology, and event provenance do not compose. A bases-loaded baseball image successfully placed three runners, yet assigned the catcher the blue uniform of the batting team while the pitcher wore the opposing red-and-white uniform. That failure is captured by the typed compatibility law

$$\text{team}(\text{catcher}) = \text{team}(\text{pitcher}) \neq \text{team}(\text{batter and runners}).$$

| Challenge case         | Structural reading                                                     | ARTISTIC decision                         |
|------------------------|------------------------------------------------------------------------|-------------------------------------------|
| Firefly cricket        | Missing roles, incomplete pitch, foreign object, and ball scale defect | Global repair; retain after manual audit  |
| Nano Banana cricket    | Duplicate role and missing bowler                                      | Rebind roles and reconstruct delivery     |
| Nano Banana baseball   | Incompatible action phases and excess agents                           | Select pre-pitch state; provisional admit |
| Nano Banana tennis     | Opponents on same side; duplicate and embedded ball                    | Repair topology and contact geometry      |
| OpenAI tennis and pool | Generic visible-scene obligations satisfied                            | Admit unchanged                           |
| OpenAI close cricket   | Locally coherent; global roles outside the frame                       | Admit with global uncertainty             |
| OpenAI wide cricket    | Global count, umpire-position, and ball-source violations              | Repair or regenerate                      |
| OpenAI baseball        | Bases loaded, but catcher bound to the offensive team                  | Localized attribute repair                |

These observations are case studies rather than matched estimates of provider quality. Prompts, seeds, aspect ratios, sampling policies, model revisions, and editing interfaces were not controlled across systems. A proper benchmark must freeze those variables where interfaces permit and report both raw competence and marginal ARTISTIC gain:

$$B_p = \text{score}(I_p, S), \quad \Delta_p = \text{score}(I_p^+, S) - B_p,$$

where  $p$  indexes the provider and  $I_p^+$  is the retained output after any selective intervention. A strong base model may have large  $B_p$  and small  $\Delta_p$ ; ARTISTIC may still be valuable as a certificate for a private, high-cost, or exact domain contract. In casual generation, the same additional verification cost may not be justified.

#### 18.14.5 The generator interface

A deployable system should isolate each image provider behind a small adapter:

$$\begin{aligned} \text{generate} &: (\text{prompt}, \text{references}, \text{controls}) \longrightarrow I, \\ \text{edit} &: (I, \text{contract}, \text{mask}, \text{controls}) \longrightarrow I^+. \end{aligned}$$

The ARTISTIC controller owns the sketch, observer, diagnosis, repair contract, and admission decision; the provider supplies image proposals. This boundary supports at least three depths of integration.

*Black-box prompting.* Compile a sketch into a more informative initial request and call an ordinary generation endpoint.

*Image-level assurance.* Inspect a returned image, localize violations, and call an editing endpoint with preservation and repair obligations. The sports cases above operate at this level.

*Diffusion-native enforcement.* Observe and steer intermediate latents, attention maps, score fields, or denoising steps. This deeper level is accessible only when the provider exposes the relevant internal state; it remains the most direct experimental bridge to the double infinitesimal geometry proposed earlier in the chapter.

The interface distinction prevents a misleading interpretation of the current results. In the challenge repairs, ARTISTIC supplied the domain theory and compiled instruction, while the image editor synthesized the corrected pixels. The experiments did not directly alter the hidden diffusion trajectory or prove that the editor had acquired a persistent theory of the game.

#### 18.14.6 *Movie Mode as a semantic audit trace*

Static before–after pairs conceal the reasoning boundary between diagnosis and generation. ARTISTIC Movie Mode therefore renders five recorded stages:

Declare → Observe → Compile → Repair → Verify.

The trace shows the original image, localized counter-witnesses, exact compiled contract, retained proposal, and clause-level audit. It also states whether the edit was local or reconstructive and which residuals remain uncertain. The visualization is explicitly labeled an *explanatory semantic reconstruction*. A black-box editor returns a completed image, so the movie is not a recording of hidden denoising steps and must not be presented as one.

#### 18.14.7 *Three practical axes*

The cross-provider benchmark must measure more than final correctness. *Generality* asks whether the same engine and constraint vocabulary transfer across providers, domains, styles, action phases, and failure types. *Ease of use* measures human authoring time, domain expertise, annotations, correction turns, and whether a rulebook or manual can replace a hand-written sketch. *Processing cost* measures generation and editing calls, latency, compute, image tokens or pixels, and monetary cost. A useful summary quantity is the gain in verified structure per

unit of additional burden,

$$\eta_p = \frac{\Delta_p}{\alpha C_p^{\text{money}} + \beta C_p^{\text{latency}} + \gamma C_p^{\text{human}}},$$

with weights fixed by the intended application rather than selected after the result is known.

The matched evaluation should compare raw generation, sketch-compiled generation, one global repair, localized masked repair, and—where available—diffusion-native enforcement. Primary endpoints are constraint satisfaction, false repair of valid images, preservation of unaffected content, calibrated abstention, provider and domain transfer, latency, and cost. The strongest practical result may be selective rather than maximal: a terse request is generated once, most valid images are certified cheaply, and only supported violations trigger another expensive call.

#### Design principle

ARTISTIC scales by changing the theory supplied to a common assurance engine, not by hiding a separate hand-written repair inside every image. A better generator should reduce the intervention rate; it should not change what the registered visual contract means.

### 18.15 Visual theory invention in Bongard problems

Bongard problems provide a sharper test of visual creativity than image repair. Each problem presents six positive and six negative drawings and asks for the intended rule that separates the two sides [Bongard, 1970, Foundalis, 2026]. The task is deliberately underspecified: twelve panels admit many accidental separators, whereas a successful explanation should name a reusable visual concept. Figure 18.15, for example, is not merely a binary classification problem. Its intended distinction is a relation between typed objects: a triangle lies above a circle on the left, while a circle lies above a triangle on the right.

This distinction turns Bongard solving into a compact experiment in theory construction. Let

$$O_k : X \longrightarrow \mathcal{R}_k$$

be the current visual observer and let  $\mathcal{H}_k$  be its executable rule language. Assimilation searches for and composes hypotheses  $h \in \mathcal{H}_k$  while keeping  $O_k$  fixed. A candidate extension enlarges the declaration—for example by adding object identity, topology, containment, or a typed spatial relation—to obtain  $(O_{k+1}, \mathcal{H}_{k+1})$ . It becomes an admitted accommodation only after counter-witness and transport tests. The mixed

or core question is whether the new observable is bound to the right object and remains meaningful under irrelevant transformations. Thus zero panel error is only the beginning of admission, not its conclusion.

#### Design principle

A separating rule is not yet a discovered concept. Admission requires the rule to survive transformations that preserve the intended meaning and to fail on a counter-witness that reverses that meaning.

This point is familiar in earlier visual-language approaches to Bongard problems: a flexible classifier can fit incidental properties without recovering the concept communicated by the examples [Depeweg et al., 2018]. The DIAL formulation makes the remedy operational. For a candidate  $h$ , the later stages used a conjunctive contract whose core conditions were

$$\begin{aligned}\hat{R}_{\text{panel}}(h) &= 0, & h(gx) &= h(x), \\ p_{\text{FWER}}(h) &\leq 0.05, & h(q) &= y_q.\end{aligned}$$

Here  $g$  ranges over registered nuisance transformations and  $q$  is an independent counter-witness. The final clause is especially important: a candidate-specific query can distinguish two rules that agree on all twelve displayed panels.

#### 18.15.1 Five evidence regimes for concept induction

Five evidence regimes connect registered scene graphs to historical pixel problems. Their results are summarized below. The rows are not a leaderboard: Corpus-2 is a sealed blind evaluation, while Corpus-4 deliberately uses the published target relation as an oracle for active admission and is therefore an upper-bound experiment.

| Stage    | Question                                                     | Result                                                                                                     | Diagnostic                                                                                                         |
|----------|--------------------------------------------------------------|------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------|
| Corpus-0 | Can a registered scene-graph language grow conservatively?   | 50/50 solved; 11 by reuse, 32 by composition, and 7 by accommodation; language size $3 \rightarrow 10$ .   | Typed extensions were reusable and caused no regression on earlier problems.                                       |
| Corpus-1 | What is missing on historical BP1-25?                        | 7 expressible; 7 required a rule constructor; 11 required a perceptual primitive.                          | Rule search and perceptual representation are distinct obstructions.                                               |
| Corpus-2 | Does a frozen scalar observer generalize blindly to BP26-50? | 16 rules admitted and 9 abstentions; after reveal, 0 exact, 2 related proxies, and 14 spurious separators. | Zero error concealed nuisance exploitation; 12/16 rules used horizontal centroid.                                  |
| Corpus-3 | Does typed accommodation repair that failure?                | 4 rules admitted: 3 semantic, 1 spurious, and 21 abstentions.                                              | Object-attribute binding removed most page-position shortcuts; conservative ignorance replaced unsupported claims. |
| Corpus-4 | Can candidate-specific queries improve admission?            | 6/25 semantic rules, 19 abstentions, 0 selected spurious rules, using 7 oracle queries.                    | Active counter-witnesses rejected accidental alternatives, but the oracle prevents an autonomous-discovery claim.  |

The failure of Corpus-2 is the most informative result in the sequence. Its scalar learner appeared successful before the answer key was opened: sixteen of twenty-five problems had a zero-error rule. Yet none expressed the intended invariant. Accommodation in Corpus-3 replaced aggregate statistics with typed objects and relations, added 48 translation and crop-jitter counter-witnesses for every admitted rule, and calibrated selection by an exact permutation family-wise test. Precision improved sharply, at the price of twenty-one honest abstentions.

Corpus-4 then treated abstention as a request for information. Rather than discard the previous observer, it formed a conservative atlas

$$O = (O_{\text{connected}}, O_{\text{topological}}),$$

combining connected-contour evidence with a topology chart for filled cores and holes. This is evidence transport across accommodation: a new chart may extend what can be seen without invalidating previously admitted observations. The resulting system recovered six intended rules with no selected spurious rule, although nineteen cases correctly remained outside the supported language.

Problem 34 illustrates the active step. The displayed panels supported both an accidental fill-count rule and the intended distinction based on hole area. A synthesized counter-witness made those explanations disagree. The oracle rejected `solid_minus_outline` and admitted `max_hole_ratio`; the system changed which structural observable it trusted rather than merely tuning a decision threshold.

These experiments advance ARTISTIC from correcting a declared image defect to a bounded form of visual theory invention. The creative act is not drawing another panel; it is extending the vocabulary in which

the distinction can be stated, transporting prior evidence into that extension, and admitting the new concept only after a discriminating probe.

Firefly 5 proposal



One-pass ARTISTIC repair



Figure 18.6: The sealed Firefly cricket challenge. The initial image has the appearance of cricket without a complete model of one delivery. A frozen integrity contract guides a single editing call that reconstructs the missing roles, two-ended pitch, and ball trajectory. This is a qualitative one-case challenge, not a provider benchmark.



**Role repair.** A duplicate wicketkeeper-like role and ambiguous non-striker action are replaced by a bowler and a waiting non-striker, yielding one temporally and spatially coherent delivery.



**Temporal-state repair.** A collage of contact, advancing runners, dust, and coaching is replaced by one provisionally admitted pre-pitch bases-loaded configuration.



**Topological repair.** The human is moved to the opposite half-court; one ball is placed beyond the robot's strings; and the duplicate ball and holographic overlay are removed.

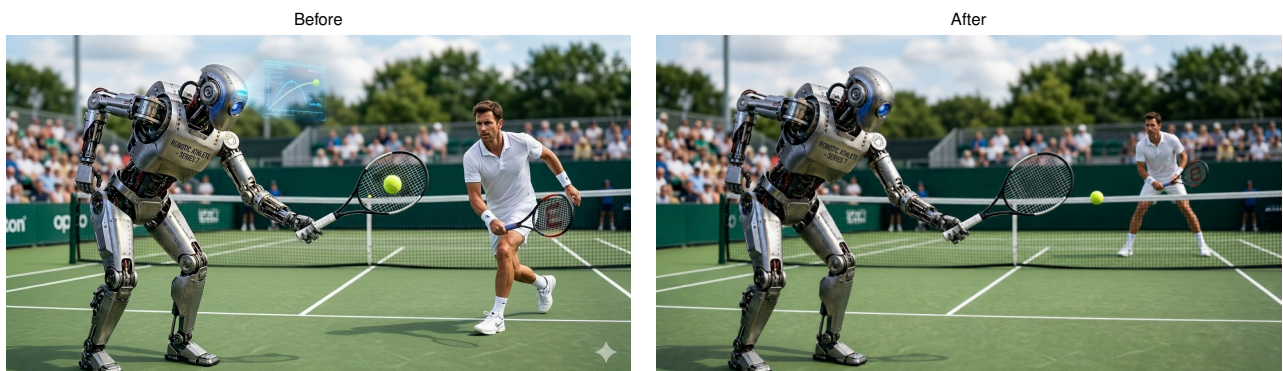


Figure 18.7: Three Nano Banana Lite challenge cases and their retained ARTIS-TIC proposals. The upper two source assets already contain side-by-side labels; the tennis pair is shown separately. The examples distinguish role, action-phase, and topology repair. Baseball and tennis remain provisional admissions because distant team binding, ball control, exact identity, and temporal velocity are not all independently observable in one frame.



Admit: coherent tennis scene



Repair required: global cricket contract



Figure 18.8: Selective assurance on two OpenAI Playground outputs. The tennis image satisfies its visible-scene contract and should be left unchanged. The wide cricket image responds impressively to a demanding prompt but fails global player cardinality, umpire placement, and ball-source constraints. The comparison illustrates two decisions by ARTISTIC, not a ranking of image providers.

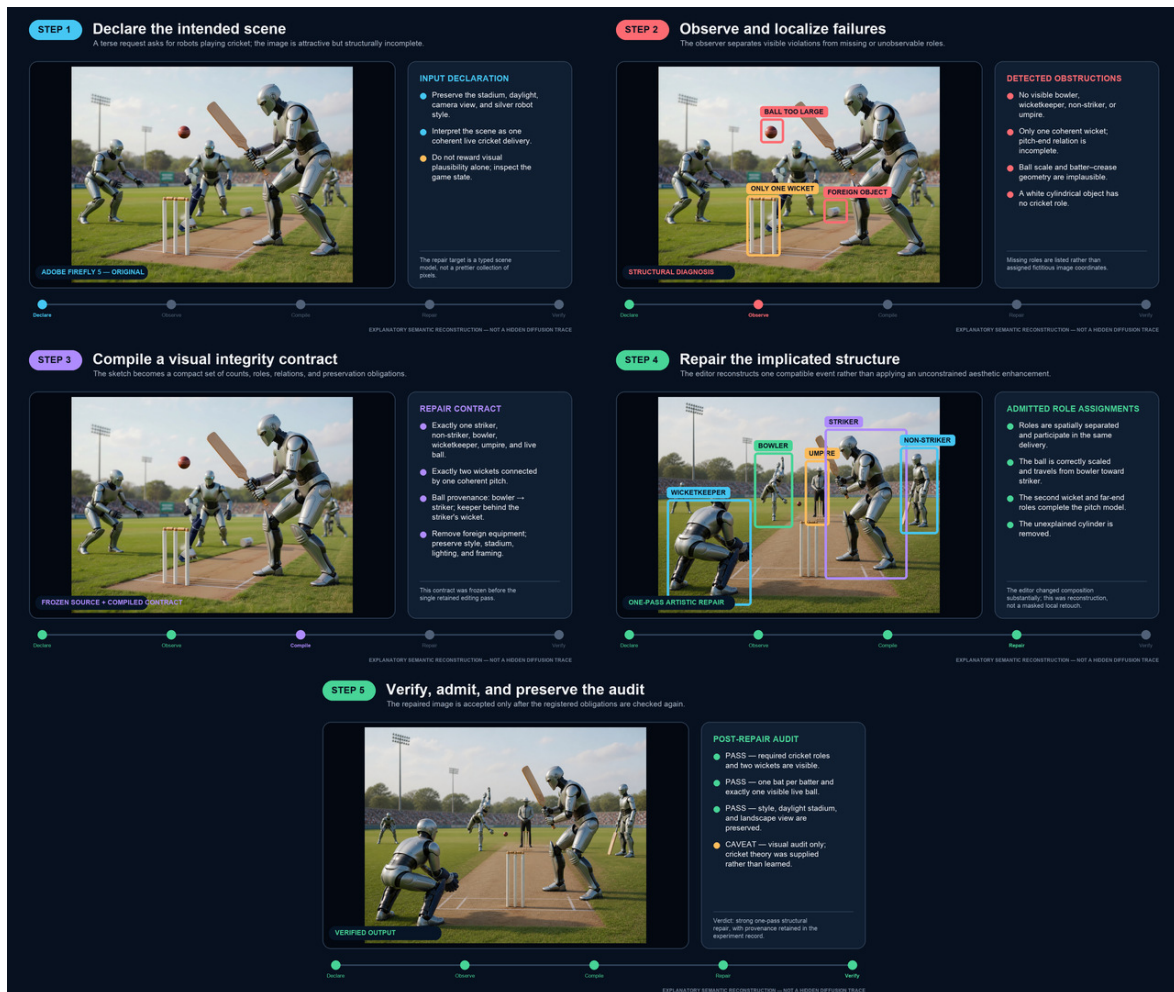


Figure 18.9: ARTISTIC Movie Mode for the Firefly cricket challenge. The five-stage storyboard makes the declaration, diagnosis, compiled contract, semantic reconstruction, and final audit visible. Equivalent traces were rendered for the Nano Banana cricket, baseball, and tennis cases. The persistent disclosure distinguishes this auditable semantic record from inaccessible diffusion internals.

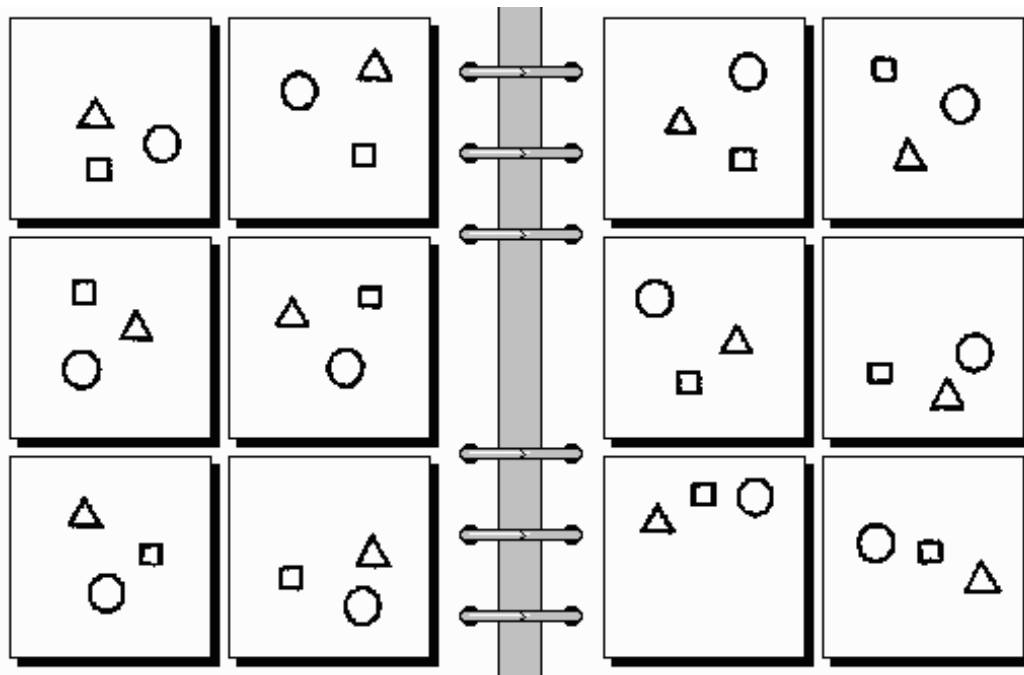


Figure 18.10: Historical Bongard problem 37. The relevant distinction is not the absolute location of any mark on the page, but the typed order relation between a triangle and a circle. Recovering this description requires an observer that binds shape to object identity and a language in which *above* can be expressed.

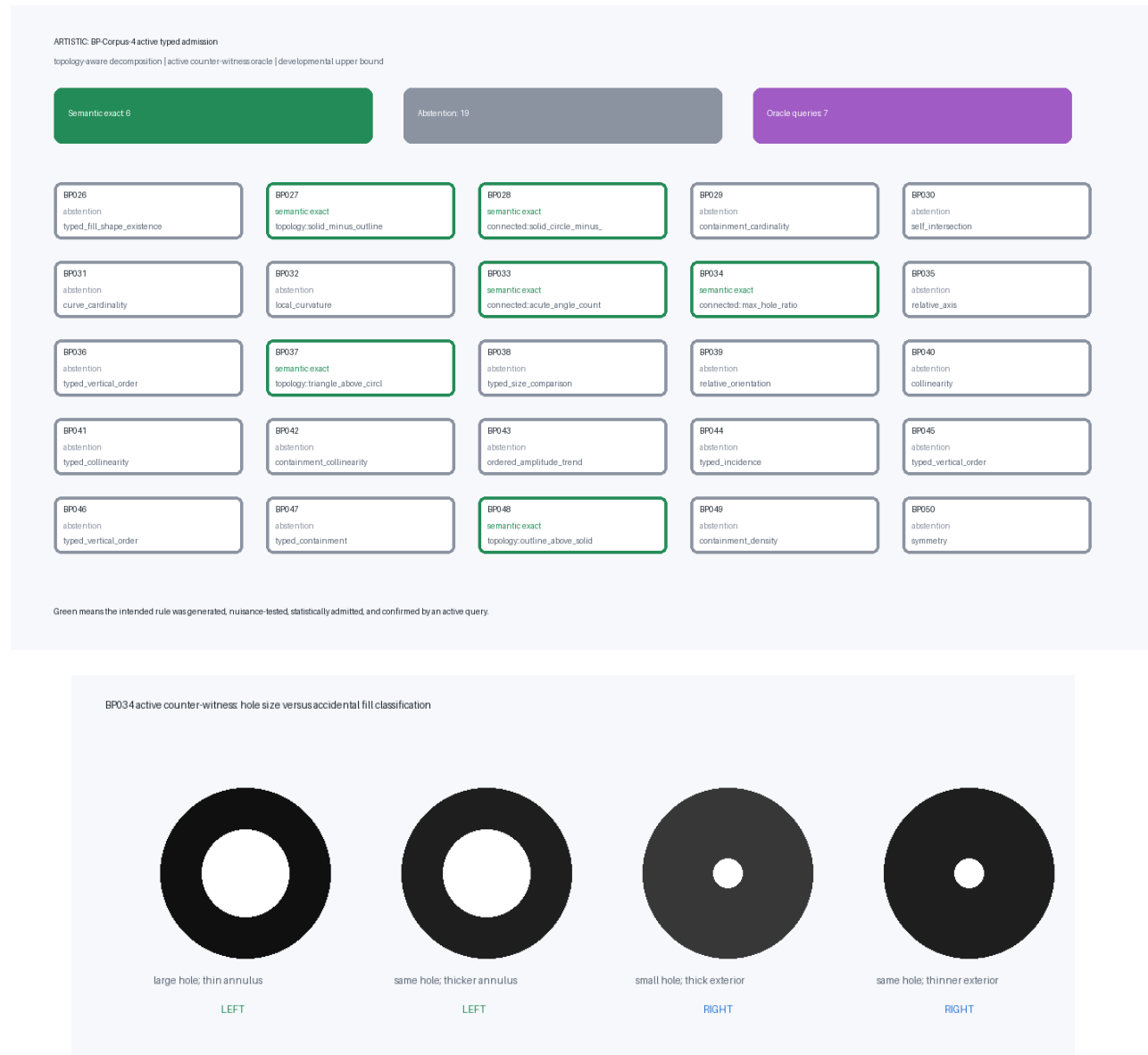


Figure 18.11: Corpus-4 active admission. *Top*: the dashboard for historical BP26–50; green entries mark six intended typed rules admitted by the observer atlas, while nineteen problems remain explicit abstentions. *Bottom*: the BP34 probe makes an accidental separator and the proposed topological account predict different labels. The experiment is a developmental upper bound because the known Bongard relation labels each requested counter-witness.

18.16 *What the experiments establish***Boundary**

The evidence has three different statuses. TI/NI supplies the preregistered declaration-and-repair ladder; the sports images are qualitative external challenge cases with disclosed prompts and manual audits; and Bongard Corpus-4 is an oracle-assisted upper bound. None of these, alone or together, shows that ARTISTIC has autonomously invented and retained a new visual theory.

The evidence sequence supports four bounded claims. A diffusion-based generation-and-repair workflow can maintain and evaluate a typed visual declaration rather than relying only on pixels or a scalar similarity; this does not imply that the declaration was enforced inside every denoising step. Structural failure can be localized to typed counter-witnesses and repaired with less area than blanket projection. The declaration itself can be estimated and evaluated against information withheld from selection and repair. Finally, uncertainty can trigger another probe that increases safe relational coverage without converting disagreement into certainty.

In DIAL language, ordinary denoising is the assimilative direction: it moves within a fixed visual and conditioning theory. The proto-accommodative direction probes changes to the declaration, observer factorization, repair operator, or acquisition policy; admitted finite accommodation installs one of those changes. The TI-4 and NI-2 evidence sequences show why these directions cannot be collapsed. Source and change observers have different jobs; point and interval relations support different inferences; a failed declaration can demand a new probe rather than another optimization step.

The experiments do *not* yet construct a double involution algebroid or measure its abstract interchange defect. They realize a finite operational shadow of the proposed geometry: two classes of change, a mixed failure, a typed extension, and an independent admission record. Nor have they shown visual-language invention in the strongest sense. The vocabulary of people, masks, parts, colors, and visible-area relations is supplied by the experimenter. The system constructs a control structure inside that language: adaptive admission, a factored observer, an uncertainty type, or an acquisition rule.

ARTISTIC's admitted empirical substrate defines a stronger test: withhold an object, relation, or compositional constructor and ask the system to propose the declaration type itself. Success would require communication to a fresh interpreter, transport across renderers or models, and counterfactual uses not enumerated in training examples.

### 18.17 *Measurements and failure modes*

Pixel similarity is inadequate because the same semantic construction may have many visual realizations. The benchmark therefore uses typed scene-graph checks in the procedural phase, registered vision models and human audits in the natural-image phase, and controlled contrastive or perturbation-based tests where the protocol supports them. These tests diagnose structural dependence; they acquire a causal interpretation only when the manipulated interface has a registered causal semantics.

Primary measures include semantic accuracy, exact relation and binding success, preservation of irrelevant attributes, productivity on unseen combinations, sample and compute cost, false extension, abstention, and transport. The three-way AGENTIC interaction asks whether joint structural diagnosis, executable language construction, and active probe selection exceed all component and pairwise systems under matched budgets.

Characteristic failures include a token that memorizes its exemplars, a relation that works only for one object pair, an adapter that improves visual quality while changing the proposed meaning, a private language that no fresh agent can acquire, and a false accommodation caused by limitations of the visual evaluator. Each failure should remain in the theory store as a typed counterexample.

### 18.18 *Open mathematics*

ARTISTIC creates three mathematical problems for the DIAL program.

First, determine conditions under which a conditioned probability-flow family integrates to a smooth groupoid over noise time and semantic context. Second, construct the language-side groupoid of admissible semantic transports and identify the correct interchange comparison with generative flow. Third, decide whether stochastic diffusion should be represented pathwise by random groupoids, through Markov categories, or by a second-order enlargement of involution algebroids.

A successful theory should explain when the mixed defect is invariant under changes of latent coordinates, how it relates to observed attribute binding, and when its vanishing implies transportable semantic control. Until those results are proved, the double-groupoid picture is a disciplined conjectural scaffold rather than an established theorem about diffusion models.

ARTISTIC primarily moves generated images toward a supplied theory. The next chapter reverses the pressure: it asks when a persistent failure of the current artistic language warrants extending that language with a reusable generative technique. Repair thereby becomes the

empirical and conceptual staging ground for DILATE, rather than being presented as creativity by itself.

### *Further Reading*

Score-based generative modeling through SDEs and the associated probability-flow ODE are developed by [Song et al. \[2021\]](#). Classical stochastic flows of diffeomorphisms provide the pathwise counterpart under strong regularity assumptions [[Kunita, 1984](#)]; recent work clarifies the additional well-posedness needed for a deterministic probability-flow representation [[Cont, 2026](#)]. Textual inversion is the principal baseline for learning a new conditioning “word” in a frozen text-to-image system [[Gal et al., 2022](#)]; the Diffusers training implementation makes its frozen components, learned token rows, checkpoints, and multi-vector variants directly available for the ARTISTIC calibration ladder [[Hugging Face, 2026](#)]. Structured diffusion guidance illustrates how linguistic structure and attention control can improve attribute binding and composition without claiming language invention [[Feng et al., 2022](#)]. ARTISTIC asks whether these image, language, and experimental-selection structures can be jointly extended and independently admitted. Bongard’s original collection [[Bongard, 1970](#)] and the maintained historical index [[Foundalis, 2026](#)] provide a demanding visual-concept corpus; [Depeweg et al. \[2018\]](#) develop a complementary account in terms of a visual language and pragmatic inference.





## *Artistic Theory Extension*

### DILATE: DOUBLE INVOLUTION LEARNING FOR ARTISTIC THEORY EXTENSION

Chapter 18 developed ARTISTIC as a system for maintaining and repairing a declared visual language. Its strongest experiments estimate visual declarations, localize failures, acquire counter-witnesses, and repair images while preserving protected content. Those capabilities are necessary for creative visual reasoning, but they do not yet explain how a system could invent a reusable artistic technique. A repair system assumes that the language needed to state the correction is already available. Transformational creativity begins when that assumption fails.

The two chapters therefore apply structural pressure in opposite directions. ARTISTIC moves a generated model toward an already declared visual theory: a cricket scene should satisfy cricket’s role, cardinality, and layout laws. DILATE asks whether the declared artistic theory itself is too small and should move toward a reusable operation not currently expressible in its language. The second direction is not permission for “anything goes.” A candidate extension must conserve the valid part of the old theory, explain a persistent quotient obstruction, generate a family rather than one fortunate image, and survive independent assessment of its control and value.

This chapter introduces DILATE—Double Involution Learning for Artistic Theory Extension—as the theory-extension layer of ARTISTIC. It starts from the Creative Adversarial Network (CAN), which encourages a generator to remain recognizable as art while deviating from established style classes [Elgammal et al., 2017]. Relative to CAN, DIAL-CAN changes the unit of creativity from a statistically unusual image to a reusable extension of the artistic operation language; DILATE names the resulting proposed paintbrush-discovery architecture. Its target is not merely an image lying between known styles, but an admitted extension of the visual theory by which styles and techniques are described. The central signal is therefore not classifier uncertainty. It is a persistent mixed obstruction that cannot be expressed by the

currently registered artistic operations.

The construction developed here is a proposed formal architecture. Lie algebroids, Weil probes, sketches, and Kan extensions are established mathematics; their assembly into DILATE is a proposed architecture. The chapter distinguishes definitions made under explicit regularity assumptions, operational tests instantiated by the reported studies, and bridge results that remain to be proved.

A DILATE artifact dossier would contain a reusable sampler deformation or “paintbrush,” the visual-theory map that types it, transport across prompts, seeds, subjects, and compatible generators, independent structural and admission certificates, independently sourced critical-value assessments, and a persistent implementation with provenance.

Its enclosing representational package  $\mathfrak{R}_{\text{DILATE}}$  also contains the sampler and paintbrush grammar, observer geometry, equivalence and irreducibility criteria, transport doctrine, proposal mechanism, and admission rules. A transformational claim therefore requires

$$Y_{\text{DILATE}} : \mathfrak{R}_{\text{DILATE}} \longrightarrow \mathfrak{R}_{\text{DILATE}}^+$$

with  $i : S \rightarrow S^+$  only its artistic-presentation component. This distinction matters whenever an experiment learns a router, connection, metric, or observer response without enlarging the artistic language itself.

The registered studies below estimate mixed interaction, assemble paintbrush programs inside supplied finite languages, route supplied subject theories, and test finite-sample gates. They do not yet construct and admit an artistic operation absent from the registered paintbrush language.

The artistic setting therefore has two different reference regimes. A controlled withheld-generator experiment supplies machine-held ground truth about an operation deliberately removed from the initial language. Open-ended artistic invention has no corresponding ground-truth aesthetic. There the admission burden shifts to reusable control, irreducibility to registered operations, transport, persistence, and independently assessed critical value. Neither regime licenses replacement of those criteria by a single style-classification or novelty score.

In this sense ARTISTIC supplies an integrity envelope for DILATE. The same observers that detect broken objects, bindings, and relations can prevent a putative new technique from earning novelty merely by destroying visual coherence. The extension may alter the artistic language, but it must state which older obligations it preserves, deliberately relaxes, or replaces.

*Chapter roadmap.* The chapter moves from CAN’s statistical baseline to a typed artistic-theory extension, develops the bracket, paintbrush, and video evidence, and closes with the withheld-generator test, baselines, and open mathematical obligations.

**Design principle**

DIAL supplies the double-involution geometry. DIAL-CAN identifies the conceptual step beyond CAN-style statistical deviation. DILATE is the operational target: an architecture intended to diagnose a missing artistic operation, propose a reusable paint-brush, transport it across generative states, and subject it to typed admission.

*19.1 What CAN makes creative*

Let a discriminator have an art head and a style head. Schematically, the CAN generator balances

$$\mathcal{L}_{\text{CAN}} = \mathcal{L}_{\text{art}} + \lambda \mathcal{L}_{\text{amb}},$$

where the first term rewards membership in the learned distribution of art and the second discourages confident assignment to an established style class. The important insight is that novelty can be sought relative to an explicitly learned repertoire rather than identified with unconstrained randomness.

The limitation is equally precise. The style taxonomy, discriminator interface, generator vocabulary, and admissible image operations remain fixed. High style entropy may identify a novel mixture or an unusual point near a learned boundary, but it does not say that the current language of styles is inadequate. Nor does it construct a new visual operation, explain its laws, or establish that the operation generates a coherent family of works. In Boden's terminology, CAN naturally supports combinational and exploratory readings of creativity. A claim of transformational creativity requires an explicit change in the representational space itself [Boden, 1998, 2004].

**Boundary**

Style ambiguity is evidence relative to a classifier. It is not evidence that the visual theory lacks a generator. A DILATE extension must be forced by a localized structural residual and then survive independent admission.

*19.2 The state space is a space of artistic theories*

Let  $S$  be a sketch presenting a visual theory. Its generators may name objects, colors, textures, parts, spatial relations, viewpoints, composition rules, and rendering operations. Its equations and designated

constructions state which composites agree and which visual structures must be preserved. A realization

$$m : S \longrightarrow \mathcal{E}$$

interprets this presentation in a semantic and generative environment  $\mathcal{E}$ . The realization may assign an image distribution rather than a single artifact to a semantic scene.

For a fixed sketch  $S$ , write  $M_S$  for the space of admissible models, modulo the chosen observational equivalence. This quotient is important: two neural parameter vectors that induce the same relevant artistic behavior should not count as different visual theories. Allowing the sketch itself to change produces a larger space whose points are pairs

$$(S, m).$$

It is generally unsafe to call this entire space a smooth manifold. Different sketches have different generators and model dimensions, and their moduli may contain singular strata. A more faithful global object is a fibred or Grothendieck-style assembly of the local model spaces  $M_S$ . The differential construction below is made on a regular stratum; finite sketch maps connect different strata.

This distinction separates two operations that a parameter-only account can collapse:

variation inside  $M_S$ , with  $S$  held fixed,

$S \xrightarrow{i} S^+$ , with a corresponding transport of models.

### 19.3 Two local artistic directions

Fix a regular model stratum  $M = M_S$ . DILATE places two interpreted algebroids over it:

$$A \longrightarrow M, \quad B \longrightarrow M.$$

The side  $A$  contains assimilatory skills: operations that refine a work or generator while preserving the current visual theory. Examples include registered color transformations, compositional adjustments, brush or denoising operations, and corrections that remain expressible in  $S$ .

The side  $B$  contains proto-accommodative probes: local variations that test possible reorganizations of the visual language. They may perturb token bindings, viewpoint conventions, compositional laws, geometric invariants, or the interface between a semantic declaration and its renderer. A section of  $B$  is not yet a new technique. It is an infinitesimal question asked of the current theory.

Their anchors

$$\rho_A : A \longrightarrow TM, \quad \rho_B : B \longrightarrow TM$$

map abstract artistic operations to observable local changes in the current model space. Sections  $a \in \Gamma(A)$  and  $b \in \Gamma(B)$  therefore induce vector fields  $\rho_A(a)$  and  $\rho_B(b)$  on the same stratum. The side interpretations are part of the target card; two arbitrary neural gradients do not become assimilation and accommodation merely by being named  $A$  and  $B$ .

The proposed double involution structure adds compatibility, core, and interchange data to these sides. At the classical smooth level, the picture is a candidate double Lie-algebroid geometry; internally, its relationship with involution algebroids in a tangent category remains one of the DIAL bridge problems described earlier in the book [Mackenzie, 2000b, Burke and MacAdam, 2019].

#### 19.4 From noncommutativity to structural obstruction

The commutator of the two realized flows measures the lowest-order difference between “assimilate then accommodate” and “accommodate then assimilate.” Under the usual smoothness and local-flow assumptions,

$$\Phi_{-\epsilon}^B \Phi_{-\epsilon}^A \Phi_{\epsilon}^B \Phi_{\epsilon}^A = \text{id} + \epsilon^2 [\rho_A(a), \rho_B(b)] + O(\epsilon^3).$$

The bracket is consequently an order-sensitive mixed witness. It is not yet a creative obstruction. Neural vector fields commonly fail to commute, and a large bracket can arise inside a perfectly adequate representation.

To ask whether the current artistic language is closed under this interaction, combine the anchors:

$$\rho_A \oplus \rho_B : A \oplus B \longrightarrow TM.$$

Assume first that its image has locally constant rank. Define the artistic obstruction bundle by the exact sequence

$$A \oplus B \xrightarrow{\rho_A \oplus \rho_B} TM \xrightarrow{q_{\text{art}}} Q_{\text{art}} \longrightarrow 0,$$

where

$$Q_{\text{art}} = \text{coker}(\rho_A \oplus \rho_B).$$

**Definition 19.1** (Artistic closure obstruction). For registered local sections  $a$  and  $b$ , the DILATE obstruction is

$$\Omega_{\text{art}}(a, b) = q_{\text{art}}([\rho_A(a), \rho_B(b)]) \in Q_{\text{art}}.$$

It is the component of the mixed bracket that cannot be represented by the currently registered assimilatory and proto-accommodative directions.

**Proposition 19.2** (Tensoriality of the quotient obstruction). *On a constant-rank stratum,  $\Omega_{\text{art}}$  is  $C^\infty(M)$ -bilinear in  $a$  and  $b$ . It therefore defines a local bundle morphism*

$$A \otimes_M B \longrightarrow Q_{\text{art}}.$$

*Proof.* Write  $X = \rho_A(a)$  and  $Y = \rho_B(b)$ . For  $f, g \in C^\infty(M)$ ,

$$[fX, gY] = fg[X, Y] + fX(g)Y - gY(f)X.$$

The last two terms lie in  $\text{im } \rho_B$  and  $\text{im } \rho_A$ , respectively, so  $q_{\text{art}}$  annihilates them. Hence  $q_{\text{art}}([fX, gY]) = fg q_{\text{art}}([X, Y])$ .  $\square$

If  $\Omega_{\text{art}}(a, b) = 0$ , the order effect closes inside the current language. It may still support valuable combinational or exploratory behavior, but it does not force a new basis element. If the quotient class is nonzero, the interaction exposes a direction for which the present artistic theory has no registered expression.

When the anchor image changes rank,  $Q_{\text{art}}$  need not be a vector bundle. One must work stratum by stratum or replace the ordinary cokernel by an appropriate quotient sheaf, stacky quotient, or derived cokernel. Singular rank changes may themselves mark creative frontiers, but the elementary bundle formula must not be applied across them without justification.

#### Design principle

DILATE does not equate creativity with a nonzero Lie bracket. It looks for the persistent component of that bracket that survives projection away from everything the current visual theory can already express.

### 19.5 The Weil profile of an obstruction

A single bracket estimate can spike because of minibatch noise, a chart boundary, a transient optimization state, or an unreliable observer. Theory extension therefore requires a profile across infinitesimal probes. Let  $W$  range over a registered family  $\mathcal{W}$  of Weil algebras, and let  $T_W$  denote the corresponding prolongation. The lifted residual has the form

$$\Omega_W(a, b) \in T_W Q_{\text{art}}.$$

For a morphism of probe shapes  $W \rightarrow W'$ , the induced comparison must carry the residuals compatibly. The profile

$$\mathcal{P}_{a,b} = \{\Omega_W(a, b)\}_{W \in \mathcal{W}}$$

should therefore behave as a natural family rather than as an unrelated list of scalar scores [Leung, 2017, 2018].

The bracket appears through an order- $\epsilon^2$  commutator. First-order dual numbers alone do not expose this mixed coefficient. The probe family must include a double infinitesimal such as  $D \otimes D$ , together with the flips and comparison maps needed to exchange its two infinitesimal directions. This is precisely where the double geometry earns its role: the mixed coefficient is not a second copy of either side derivative.

**Definition 19.3** (Weil-persistent artistic obstruction). A candidate obstruction is Weil-persistent relative to  $\mathcal{W}$  when it is natural under the registered probe maps, remains separated from the zero class across admitted scales, retains a coherently transported direction, localizes to the same deficient part of the sketch, and survives independent observer and data resampling.

This definition contains both structural and statistical obligations. The residual itself remains typed; a scalar persistence statistic may summarize evidence, but it cannot erase where the failure lives or which direction it selects. The finite-sample DIAL-CAN and DILATE experiments described below illustrate why this matters: an estimated bracket may carry a stable zero boundary even when a learned calibrator has never observed one side of that boundary.

### 19.6 Constructing a candidate technique

Suppose a persistent residual identifies a missing direction. DILATE then proposes a candidate generator bundle  $C \rightarrow M$ , an anchor  $\rho_C : C \rightarrow TM$ , and a local section  $c$  whose quotient action accounts for it:

$$q_{\text{art}}(\rho_C(c)) \simeq \Omega_{\text{art}}(a, b).$$

The equation is in  $Q_{\text{art}}$ : it does not select a canonical lift of the obstruction back into  $TM$ . The declared type, localization, and regularization of that lift are therefore part of the proposal record. Depending on the localization,  $c$  may be a new visual token, a relation, a spatial operation, a token-interaction law, a geometric invariant, or a multistage rendering procedure. The corresponding finite creative act is a sketch map

$$i : S \hookrightarrow S^+ = S[c].$$

The notation  $S[c]$  suppresses the relations required to type  $c$ . A bare symbol with no domain, codomain, equations, or observable consequences does not constitute a theory extension.

The first test of the proposal is explanatory closure. Let  $q_{\text{art}}^+$  project away the operations registered in the extended theory. On held-out probes, DILATE asks whether

$$q_{\text{art}}^+([\rho_A(a), \rho_B(b)]) \simeq 0.$$

An extension that adds capacity but leaves the localized quotient residual unchanged has not explained the frontier. Conversely, forcing the residual to zero by an unrestricted memorizing generator fails irreducibility and productivity.

### 19.7 *Transporting the old artistic theory*

The sketch map  $i : S \rightarrow S^+$  induces restriction

$$i^* : \text{Mod}(S^+, \mathcal{E}) \longrightarrow \text{Mod}(S, \mathcal{E}).$$

Assume that  $\mathcal{E}$  admits the pointwise colimits required for the Kan extension. Given an old model  $m : S \rightarrow \mathcal{E}$ , a left Kan extension then provides a canonical candidate for generative transport:

$$\text{Lan}_i m : S^+ \longrightarrow \mathcal{E}.$$

Its universal property explains how existing semantic material propagates through the new generators and relations [Mac Lane, 1998, Riehl, 2016]. It does not guarantee that the result preserves the products, limits, enrichment, tangent structure, or empirical integrity required by the visual model doctrine. Those are additional proof and admission obligations.

The universal construction carries its unit

$$\eta_m : m \longrightarrow i^* \text{Lan}_i m.$$

Conservative transport requires this unit to be an isomorphism, equivalence, or registered observational equivalence of the strength demanded by the model doctrine. The existence of the Kan extension alone does not supply that certificate. In a successful extension, established color, composition, object, and rendering knowledge need not be relearned from scratch. The new technique becomes teachable because it is presented as a finite addition to a shared theory, not because a private latent vector happens to reproduce one image.

### 19.8 *The admission firewall*

The proposed extension  $(i, c, m^+)$  is admitted only through independent criteria. Seven obligations are especially important.

*Conservativity.* Restriction to the old sketch recovers the previously admitted theory to the declared tolerance.

*Irreducibility.* The proposed generator is not adequately expressible as a composition of existing generators under the registered equivalence.



*Coherence.* The enlarged sketch has a consistent realization and its new equations commute where promised.

*Productivity.* The extension generates a diverse family of works rather than memorizing one anomalous image.

*Persistence.* The residual and its repair survive changes of subject, prompt, seed, scale, and admissible observer.

*Teachability.* A fresh interpreter can learn, execute, and evaluate the new operation from its finite declaration and examples.

*Critical value.* Independent human assessors judge the family meaningful or valuable relative to a declared artistic context. Carefully separated automated observers may report registered proxies, but do not become aesthetic judges merely by producing a score.

The first six criteria can be operationalized structurally. Critical value is less reducible to a universal automatic test and must retain provenance and plurality of judgment. A new artistic language may be coherent and productive without being valued by every community; DILATE should report that distinction rather than hide it inside one scalar loss.

### 19.9 The hierarchy from CAN to DILATE

| System            | Creative object                                  | Strongest licensed claim                                                                  |
|-------------------|--------------------------------------------------|-------------------------------------------------------------------------------------------|
| CAN               | Image outside confident registered style classes | Statistical style deviation while retaining learned art membership                        |
| ARTISTIC repair   | Corrected image under a fixed visual sketch      | Typed visual integrity can be diagnosed, enforced, and audited                            |
| Local DIAL-CAN    | Ordering of two registered update fields         | A mixed bracket predicts when local artistic updates interact                             |
| DILATE paintbrush | Reusable deformation of a generative trajectory  | A typed local operator family can be synthesized and transported across generative states |
| Full DILATE       | Extended visual sketch $S^+$                     | A missing visual generator is constructed, transported, and independently admitted        |

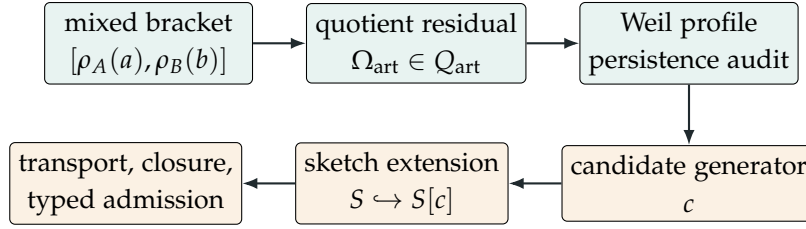


Figure 19.1: The DILATE theory-extension cycle. A mixed order effect becomes a creative candidate only after projection away from registered operations, persistence across Weil probes, construction of a typed generator, and independent admission of the resulting sketch extension.

### 19.10 Calibration of the mixed-order diagnostic

Before asking a system to extend an artistic theory, the mixed differential mechanism must be calibrated independently of that stronger claim. The following sequence tests whether a local bracket predicts the useful order of two registered updates, whether the estimate survives changes of observer and sample size, and when the correct response is abstention. It deliberately stops before theory extension.

*DIAL-CAN-0.* In a smooth two-dimensional CAN proxy, the finite order-difference converged to the exact Lie bracket with first-order slope approximately one. The bracket selected the better order in 97.7 percent of states, while an exactly commuting control remained null.

*DIAL-CAN-1.* A trained neural CAN with a LeakyReLU observer failed the registered differential audit. Rare activation-stratum crossings produced large finite-witness outliers; first-order convergence, norm correlation, matched-order accuracy, and the shuffled-control comparison all missed their frozen gates.

*DIAL-CAN-1B.* A separately registered Softplus ablation restored first-order convergence, near-perfect norm correlation, and perfect order prediction. It nevertheless failed admission because one order dominated every sampled probe, leaving the bracket controller no adaptive advantage over the best fixed order.

*DIAL-CAN-2.* Sampling three paired training checkpoints produced real decision switching. The smooth bracket predicted the better order on every registered probe and exceeded both fixed-order and shuffled controls. The registered study nevertheless missed its pre-registered aggregate diversity threshold because one order occurred on 16.7 rather than 20 percent of probes.

*DIAL-CAN-3.* A first-order finite-difference estimator was trained on checkpoints 200 and 600 and transported to the switching checkpoint 400. The supervised calibrator failed because its training regions contained no examples of one order. In a retrospective diagnostic, the

theory-supplied zero boundary improved from 77.1 to 87.5 percent held-out accuracy as the latent sample increased from 16 to 128; the diagnostic does not override the failed registered study.

Taken together, the calibrations license one narrow conclusion: in the smooth registered proxies, mixed geometry can carry useful local order information. They do not establish a successful learned finite-sample controller, a nonzero quotient residual, a Weil-persistent missing direction, or the construction of a visual generator. The final calibration instead supplies an epistemic constraint for the full architecture: when training data have no support on one side of a structurally defined boundary, a learned controller should abstain or retain the declared boundary rather than invent an exterior threshold.

### 19.11 *A paintbrush over the space of generative dynamics*

The bracket experiments operate on update fields after a generative model has already been fixed. A stronger interpretation changes the dynamics by which an image is made. Let  $\mathcal{S}$  denote a local space of admissible stochastic dynamics, with a point

$$S = (f, g, s_\theta)$$

specifying the drift, diffusion coefficient, and score field of a reverse-time image process. A conventional conditional sampler evolves according to

$$dX_t = [f(X_t, t) - g(t)^2 s_\theta(X_t, t, \gamma)] dt + g(t) d\bar{W}_t.$$

An elementary paintbrush inserts a registered state-level field  $\rho_C(c)$ :

$$dX_t = [f - g^2 s_\theta + \rho_C(c)(X_t, t, \gamma)] dt + g_c(X_t, t, \gamma) d\bar{W}_t. \quad (19.1)$$

This equation already permits a reusable operation to act throughout the trajectory rather than as a terminal image filter. It is nevertheless the lower-level shadow of the desired construction. A family of parameterized SDEs is only a bundle of possible tools. It becomes Lie-algebroidal only after one specifies admissible sections, an anchor into  $TS$ , a bracket, and the Jacobi, Leibniz, and anchor-morphism laws (or their declared internal involution-algebroid analogues).

The proposed meta-dynamics therefore separates the evolution of the visual state from the evolution of its generative law:

$$\frac{dS_\tau}{d\tau} = \rho_C(c_\tau)(S_\tau), \quad (19.2)$$

$$dX_t = b_{S_\tau}(X_t, t, \gamma) dt + \sigma_{S_\tau}(X_t, t, \gamma) d\bar{W}_t. \quad (19.3)$$

The image-time variable  $t$  traces one act of generation, while the meta-time variable  $\tau$  traces a change in the available mode of generation. In this sense, DILATE seeks a new paintbrush rather than merely a new picture. The creative artifact is a reusable deformation rule for a family of samplers.

This formulation also makes the double direction explicit. Assimilation changes an image or sampler inside the current theory; proto-accommodation probes a change to the family of admissible generative laws. A proposed law becomes accommodation only after finite realization, transport, and admission. For assimilatory section  $a$  and paintbrush section  $c$ , the corresponding mixed obstruction is

$$\Omega_{A,C}(a, c) = [\rho_A(a), \rho_C(c)] \pmod{\text{im } \rho_A + \text{im } \rho_C}.$$

A nonzero raw bracket records order sensitivity. A persistent nonzero quotient class says something stronger: the current registered image and sampler edits do not close under their interaction. Only the latter can motivate a finite extension of the paintbrush language.

#### Boundary

Equation (19.2) is a target semantics, not a theorem that the implemented brush family already forms a global Lie algebroid. The registered experiments below estimate a local anchor, metric, and connection-like transport rule in a fixed observer chart. Bracket closure, global integrability, and equivalence under changes of observer remain mathematical obligations.

### 19.12 Observer-induced metric and discrete transport

A fixed coefficient vector is not a geometrically meaningful way to transport an artistic intention across heterogeneous latent states. The same numerical edit can have very different visible effects at different noise levels, prompts, and discretizations. The PB2.2 experiment therefore represents an intent in observer coordinates and lifts it separately at each state.

For a candidate skeleton, let

$$\{b_{\text{focal}}, b_{\text{transition}}, b_{\text{peripheral}}\}$$

be a local basis of typed fields. Let  $E_i$  be the high-frequency energy measured by an independent observer in region  $i$ . Finite differences define the local response matrix

$$J_S(i, j) = \left. \frac{\partial}{\partial \epsilon_j} \log E_i(X + \epsilon_j b_j) \right|_{\epsilon_j=0}.$$

For the artistic contract “preserve focal detail, soften the transition by ten percent, and soften the periphery by twenty percent,” the frozen response coordinate is

$$u = (0, \log 0.90, \log 0.80).$$

With observer weights  $W = \text{diag}(3, 1.5, 2)$  and ridge  $\lambda = 10^{-2}$ , the regularized pullback form and weighted-response lift are

$$\begin{aligned} G_S &= J_S^T W J_S + \lambda I, \\ \alpha_S &= G_S^{-1} J_S^T W u, \\ F_S &= \sum_j \alpha_{S,j} b_{S,j}. \end{aligned}$$

Equivalently,  $\alpha_S$  minimizes

$$\|J_S \alpha - u\|_W^2 + \lambda \|\alpha\|_2^2.$$

The operational transport rule preserves the same observer-level intention, not the same coefficient vector. The resulting  $F_S$  changes with the local response geometry.

Finally, temporal weights are normalized,

$$\bar{w}_k = \frac{w_k}{\sum_j w_j}, \quad X_{k+1} = D_k(X_k) + \Lambda \bar{w}_k F_{S_k},$$

so that  $\Lambda$  denotes total integrated action rather than an edit applied anew at every denoising step. This normalization is necessary for comparing samplers with different step counts. The metric and state-adaptive transport rule are then responsible for matching the desired regional response as the state changes. Calling this rule a *discrete connection* records its intended role. A mathematical connection would additionally require coherent transformation under changes of observer chart and compositional parallel transport; neither property is established by PB2.2.

#### Design principle

DILATE transports artistic intent through observer geometry. It does not assume that a paintbrush is a globally fixed vector in neural parameter space.

### 19.13 Registered paintbrush evidence

The paintbrush experiments form a preregistered DILATE diagnostic sequence. Their archived artifacts retain the earlier DIAL-CAN-PB identifiers, while this chapter reports them as evidence about DILATE. A failed stage is retained because each failure identifies structure required by a stronger model.

| Stage | Admission | Change in the hypothesis class                                                 | Principal result                                                                                                                       |
|-------|-----------|--------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------|
| PBo   | Fail      | Select one member of a finite trajectory-operator grammar                      | Residual reduction 0.682, just below the frozen 0.70 gate                                                                              |
| PBo.1 | Pass      | Fit continuous parameters inside the declared family                           | Residual reduction 0.991; witness correlation 1.000; endpoint error 0.010                                                              |
| PB1   | Pass      | Select the missing operator family from a typed grammar                        | Selected edge-energy gated band-pass; residual reduction 0.976; all 14 gates passed                                                    |
| PB2   | Fail      | Synthesize a composite technique from a verbal artistic intent                 | Generalization failed: held-out composite/atomic ratio 1.581 and semantic retention 0.872                                              |
| PB2.1 | Fail      | Normalize total action over the denoising path                                 | Held-out ratio improved to 0.846, but semantic retention was 0.948 and focal response remained unstable across prompts and step counts |
| PB2.2 | Pass      | Add observer metric, state-adaptive transport, and rendered action calibration | All 15 gates passed; contract objective fell from 4.051 to 0.239                                                                       |

PBo.1 establishes controlled recovery of one withheld member of a declared operator family. PB1 goes one level higher: without access to that family or its fitted parameters, it searched a finite typed observer-field grammar and selected an edge-energy-gated band-pass skeleton. Its winning objective was 0.0153, compared with 0.3991 for the second candidate, and the shuffled screening control never exceeded cosine 0.0063. This is family proposal inside a declared grammar, not invention outside it.

PB2 removes the oracle target and begins from an artistic brief. Its failure is instructive. Applying a plausible composite at each denoising step accumulated too much effect, and merely normalizing the path in PB2.1 did not make a fixed brush act consistently across local SDE states. PB2.2 repaired that defect by transporting the requested observer response through the local metric and by selecting total action from rendered recovery cases before the held-out set was generated.

On four held-out prompts, PB2.2 produced focal-detail ratios

$$(0.970, 1.041, 0.974, 0.973),$$

a median peripheral-detail ratio of 0.776, CLIP semantic retention of 0.988, and median registered connection-response residual 0.0013. No



Figure 19.2: Held-out PB2.2 transport. Each column is a different prompt. The top row is the frozen base sampler, the middle row applies the PB2.1 normalized path with fixed coefficients, and the bottom row applies the PB2.2 observer-induced metric and connection-like transport. PB2.2 preserves the focal subject while producing bounded, radially ordered softening in the surrounding field. The figure is an observer-contract diagnostic, not a human judgment that a new artistic style has been invented.

connection evaluation activated the safety norm cap. Across 16-, 24-, and 32-step samplers, the maximum focal-effect and peripheral-effect ranges were 0.0535 and 0.0633, respectively.

**Experiment: DILATE–2.2: Connected Paintbrush**

**Artifact.** A reusable trajectory technique in a finite typed program language.

**Experimental provenance.** Archived as DIAL-CAN-PB2.2.

**Frozen intent.** Preserve focal detail while progressively softening the transition and peripheral regions.

**Geometry.** Observer-response pullback form, regularized weighted-response lift, normalized path action, and state-adaptive connection-like transport.

**Controls.** Atomic programs, null and shuffled cases, strength monotonicity, semantic and color preservation, prompt transport, and 16/24/32-step discretization.

**Outcome.** All 15 preregistered admission gates passed.

**Claim boundary.** The result supports a local empirical paintbrush geometry in a fixed observer chart. It does not establish a global Lie algebroid, integration to a Lie groupoid, open-ended style invention, or human artistic value.

The progression matters more than the final score. PBo–PB1 move from parameter recovery to typed family selection. PB2–PB2.2 move from oracle recovery to an intent-conditioned technique assembled inside a supplied finite program language, and show why normalization, a fiber metric, a state-adaptive transport rule, and image-level calibration are not optional details. The creative threshold beyond these studies remains unresolved and is not another gain on the same contract. It is the construction and admission of a generator that the original typed grammar did not contain.

### 19.14 *Theory routing across natural videos*

Admitting a paintbrush inside a supplied language is only the beginning. When a new video arrives, the system must decide whether the existing declaration and observer chart still apply. The DILATE-MOVIES study tests a deliberately narrower precursor to open-ended extension. A frozen observer maps a video to typed evidence, and a deterministic resolver selects one of four actions:

REUSE,      SPECIALIZE,      EXTEND,      ABSTAIN.

The reuse action retains the registered single-egret theory and chart. The specialization action retains relevant avian structure but requires a new cardinality, scale, or behavioral chart. Extension declares the subject outside the avian theory, while abstention records inadequate or conflicting evidence. These are routing decisions, not claims that a new paintbrush has already been constructed.



MOVIES-0 evaluated the first router on eight natural videos. Seven routes were correct, but a sea-lion colony was incorrectly specialized because a generic instance observer produced persistent false bird proposals. A second, taxonomic observer supplied strong sea-lion evidence. MOVIES-1 made this transverse disagreement explicit: persistent one-sided non-avian evidence could override generic bird-shaped proposals, while unresolved two-sided conflict produced a distinct abstention. This repaired the sea-lion route and conserved the other seven decisions, but it was a same-corpus repair rather than evidence of transport.

MOVIES-2 therefore froze the observer, thresholds, conflict resolver, code, model identities, and an independently collected eleven-video manifest before opening its route labels. The corpus contained four egret, five duck or pelican, and two sea-lion recordings. The following table reports the sealed result.

| MOVIES-2 endpoint         | Result      |
|---------------------------|-------------|
| Route accuracy            | 0.818       |
| Macro-F1                  | 0.886       |
| Non-abstaining coverage   | 0.909       |
| Selective accuracy        | 0.900       |
| Majority-class accuracy   | 0.455       |
| EXTEND precision / recall | 1.000/1.000 |
| REUSE precision / recall  | 1.000/0.750 |

Table 19.1: Independent transport of the frozen MOVIES-1 resolver. The resolver meets every preregistered gate on this eleven-video corpus. The result supports coarse theory routing in that registered setting, not open-set theory construction or paintbrush quality.

Both unseen sea-lion recordings were correctly routed to EXTEND, providing the first independent evidence that the MOVIES-1 conflict repair transports beyond the corpus that motivated it. No non-egret recording was incorrectly reused. The two misses reveal different structural limitations. A 4.7-second flying-duck clip yielded only one probe under the frozen temporal schedule and was conservatively rejected for insufficient evidence. The appropriate repair is duration-aware observation, not a weaker semantic threshold.

The other miss is more revealing. One recording supplied unambiguous egret evidence in every sampled frame, but the bird occupied 14.3% of the image, beyond the frozen chart's 7% reuse limit. The router returned SPECIALIZE, whereas the coarse corpus label called every egret recording REUSE. At the level of subject theory, reuse is correct; at the level of observation scale, specialization is correct. A more faithful target is therefore factorized:

(subject theory = REUSE, observer chart = SPECIALIZE).

This distinction is geometrically natural. Transporting a theory need not mean that the same local chart remains valid. In a mature DILATE system, the connection should transport the paintbrush across compatible charts while the admission layer separately decides whether the

underlying theory requires extension.

#### Experiment: DILATE-MOVIES–2: Independent Theory Routing

**Artifact.** A sealed route for each of eleven unseen natural videos.

**Frozen structure.** Visual observers, temporal sampling, thresholds, typed conflict resolver, code and model hashes, and corpus manifest.

**Controls.** Six counterfactual conflict and null cases, source-hash checks, sealing before label access, a majority baseline, and explicit abstention.

**Outcome.** All preregistered gates passed; accuracy 0.818, macro-F1 0.886, and perfect precision and recall for extension.

**Claim boundary.** Independent transport of a coarse finite router on one small wildlife corpus; not autonomous discovery of a new subject theory, domain-general video understanding, or artistic evaluation.

#### Boundary

No experiment in this chapter yet completes the full DILATE claim. PB2.2 admits a connected local paintbrush inside a supplied finite program language; MOVIES–2 admits transport of a frozen coarse router. Neither constructs a previously unavailable generator and then demonstrates its conservative, productive reuse. The withheld-generator protocol below states that stronger test as a controlled admission contract.

### 19.15 *The withheld-generator experiment*

A decisive calibration uses a controlled visual grammar in which the missing operation is known. Let  $S$  contain registered generators for color, shape, texture, layering, ordinary perspective, and composition. Construct an expanded theory

$$S^* = S[c^*]$$

with one additional operation. A useful initial choice is simultaneous multi-viewpoint projection: the same object may expose geometrically incompatible viewpoints inside one coherent composition. The renderer and held-out corpus use  $c^*$ , while the learner initially receives only  $S$ .

The experiment then asks five separate questions.

1. **Detection:** does a nonzero quotient residual distinguish examples using  $c^*$  from noise and unfamiliar compositions expressible inside  $S$ ?

2. **Localization:** does the residual identify the viewpoint and composition interface rather than an irrelevant color or texture component?
3. **Recovery:** does the proposed generator agree with  $c^*$  up to the registered observational and presentation equivalence?
4. **Closure:** does adjoining the proposal remove the residual on held-out examples without damaging old-theory cases?
5. **Productivity:** does the admitted operation generate coherent new works across unseen objects, palettes, layouts, and renderers?

The withheld operation gives ground truth without making exact syntax the only success criterion. A different presentation may be equivalent if it generates the same models, transports the old theory conservatively, and passes the same counterfactual probes.

#### Experiment: DILATE-TE-0: Withheld Visual Generator

**Epistemic status.** Proposed benchmark; no result is reported.

**Frozen reference.** A visual sketch  $S$ , its renderer, old-theory corpus, observer family, probe family, and presentation equivalence.

**Withheld structure.** One typed generator  $c^*$  and its defining relations in  $S^*$ .

**Positive worlds.** Works generated using  $c^*$  in subjects and compositions disjoint from proposal examples.

**Negative controls.** Noise, transient optimizer spikes, unfamiliar but expressible compositions, corrupted observers, and unrestricted capacity additions.

**Primary endpoints.** Detection, localization, equivalent recovery, residual closure, conservativity, productivity, and false extension.

**Claim boundary.** Recovery of one registered missing operation is a calibration of transformational theory extension, not unrestricted invention of an art-historical movement.

### 19.16 Baselines and ablations

A credible comparison must prevent visual quality from substituting for theory recovery. The corresponding ablation matrix includes:

- a CAN-style art-plus-style-ambiguity objective;
- reconstruction error without a typed quotient;
- raw bracket magnitude without projection into  $Q_{\text{art}}$ ;
- quotient residual without Weil persistence;
- persistent residual with an unrestricted proposal network;

- sketch extension without Kan transport or conservativity;
- full DILATE with all admission criteria; and
- an oracle supplied with  $c^*$ , which gives an upper bound rather than a deployable baseline.

The central false-positive control is an image family that looks unfamiliar but is exactly expressible by long compositions of existing generators. CAN may assign high style ambiguity, and a reconstruction model may report high error, but irreducibility-aware DILATE should refuse an extension. The central false-negative control varies the surface realization of  $c^*$  while retaining its structural action; a pixel-level detector should fail to transfer while a correct theory-level proposal should persist.

Together these baselines form  $\text{Ctl}_{\text{DILATE}}$ . The full system may propose  $Y_{\text{DILATE}}$  only when the unfamiliar family survives the quotient and Weil-persistence tests and the registered controls fail to express it inside the current paintbrush grammar. High novelty, bracket magnitude, reconstruction error, or observer response alone opens no theory-extension branch.

### 19.17 *Open mathematical obligations*

Several bridge results remain open. First, the quotient obstruction must be defined invariantly when anchor images change rank. Second, the internal Weil profile must be connected rigorously to the mixed core or interchange data of a double involution algebroid. Third, one needs conditions under which a persistent quotient class admits a finite sketch generator rather than only a local analytic direction. Fourth, Kan transport must be shown to remain inside the doctrine-preserving category of visual models. Finally, equivalent presentations of a recovered technique require a suitable Morita-style or model-category notion of sameness.

These obligations are not technical decoration. They separate three claims that an empirical system might otherwise conflate:

$$\begin{aligned} \text{unusual artifact} &\longrightarrow \text{persistent missing direction} \\ &\longrightarrow \text{admitted extension of artistic theory.} \end{aligned}$$

The DILATE program succeeds only when it can travel across all three while recording the evidence for each transition.

The final chapter gathers the lessons of these testbeds without treating DIAL as a universal theory of creativity. It separates what the present constructions demonstrate from what remains conjectural, and asks which parts of the framework could transfer to creative domains not studied experimentally here.

### *Further Reading*

The Creative Adversarial Network provides the direct statistical baseline for style deviation [Elgammal et al., 2017]. Boden’s distinction among combinational, exploratory, and transformational creativity supplies the conceptual hierarchy [Boden, 1998, 2004]. Sketches and functorial models originate in the categorical theory tradition of Lawvere [1963] and Ehresmann [1968]; Kan extensions and their transport semantics are developed systematically by Mac Lane [1998] and Riehl [2016]. Weil probes and tangent structure are treated by Leung [2017], while Burke and MacAdam [2019] generalize Lie algebroids to tangent categories. Classical double Lie algebroids and their Weil algebra appear in Mackenzie [2000b] and Meinrenken and Pike [2021]. These sources provide the established components. DIAL-CAN supplies the bridge from CAN’s statistical notion of deviation; the quotient-persistence-extension cycle and paintbrush system proposed here constitute the DILATE research construction.



## *Creativity as Compositional Theory Change*

### CONCLUSIONS, LIMITS, AND OPEN DOMAINS

This book began with a deliberately difficult conjunction. Creativity appears to involve a jump: a new concept, mechanism, experiment, proof strategy, or artistic technique enters a language that did not previously contain it. Infinitesimal reasoning appears to move in the opposite way, through local variations whose effects can be differentiated, compared, and controlled. The central question has been whether a disciplined interaction between these two scales can make some creative transitions intelligible without reducing them either to an unexplained leap or to optimization of a scalar novelty score.

DIAL is the book's proposed answer, but only within a carefully bounded scope. It treats a creative system as maintaining a compositional theory, varying its models in two interacting directions, diagnosing a persistent obstruction, and admitting a finite extension when the old theory cannot express a reusable solution. This is a categorical framework for *one important class* of creative processes. It is not a universal definition of creativity, nor does it claim that every human discovery was produced by double involution.

#### 20.1 *The argument in one view*

The book's argument can be compressed into the following chain:

$$\begin{aligned} \text{acquire } \mathfrak{R} \supset \mathbb{S} &\longrightarrow \text{realize } D \longrightarrow \text{probe } (A, B) \\ &\longrightarrow \text{localize } \Omega \longrightarrow \text{test } \gamma \\ &\longrightarrow \text{test less-disruptive controls} \\ &\longrightarrow \text{propose } Y : \mathfrak{R} \rightarrow \mathfrak{R}^+ \\ &\longrightarrow \text{transport and test independently} \\ &\longrightarrow \text{admit and persist, or reject and retract.} \end{aligned}$$

The initial representational package  $\mathfrak{R}$ , including its theory  $\mathbb{S}$ , realization semantics, observers, doctrine, proposal language, and admission

interface, may be supplied explicitly, learned from demonstration, or acquired through an apprenticeship in which a teacher corrects the learner's interpretation. A model  $D$  realizes that theory in a particular domain. Assimilative probes  $A$  improve behavior while preserving the current declaration; proto-accommodative probes  $B$  test how the declaration might have to change. Their mixed interaction produces a diagnostic class, schematically

$$\Omega(a, b) = [\rho_A(a), \rho_B(b)] \pmod{\text{im } \rho_A + \text{im } \rho_B}.$$

A nonzero bracket is not automatically creative. The relevant signal is a persistent quotient obstruction: an interaction that cannot be explained by the already admitted directions and survives the chosen Weil probes, uncertainty model, and negative controls. An integration observer must then test whether the candidate local repair composes into a coherent finite realization. Before a package change is submitted, registered no-change, parameter, stronger in-language, measurement, observer, and domain-specific alternatives must be tested where applicable. Only after those gates does the system submit a comparison  $Y : \mathfrak{R} \rightarrow \mathfrak{R}^+$ . If the theory presentation changes,  $J : S \rightarrow S^+$  is one component of  $Y$ , not the whole delta. The proposal must preserve what remains valid, support new models or actions, and pass an independent admission test before it is persisted; otherwise it is retracted while its counterwitnesses remain in the audit record.

#### Design principle

Creativity is not identified with surprise. In DIAL, a creative transition is a typed and auditable extension of a maintained representational language, prompted by a persistent compositional obstruction, supported by a coherent finite realization or a counterwitness to one inside the old theory, and retained because the extension is conservative, productive, and reusable.

This view changes the unit of learning. A conventional learner usually returns parameters, predictions, actions, or samples. A DIAL system may return those things, but its distinctive artifact is a change in the language that makes them possible: a latent causal alternative, a new skill constructor, a new state abstraction, a new experiment, or a new visual operation. The extension matters only if subsequent reasoning can use it.

Not every admitted package comparison is therefore a creative theory extension. A change of parameters, observer, inference routine, or enforcement policy may be operationally important while leaving the declared language fixed. ARTISTIC repairs chiefly occupy this category. The stronger representational-language claim requires a registered enlargement of the presentation or generative operations, together with evidence that the added structure is irreducible, productive, transportable, and reusable. DILATE is designed to test that still-open step.



## 20.2 *What the present evidence supports*

The book contains three kinds of named object: component systems, testbed ladders, and proposed integrative architectures. Each targets a characteristic artifact, but they do not carry the same empirical status. CLIC, OPTIC, and RELIC isolate component constructions; IPC, AF, and DIAL-URL supply controlled mathematical construction ladders; SGTE and CTTE organize scientific evidence sequences; AGENTIC specifies their composition; ARTISTIC supplies visual repair and assurance infrastructure; and DILATE proposes the stronger transition to artistic theory extension. The names make this division of labor traceable, provided they are not mistaken for equally mature algorithms.

| Named object | Maintained package                                            | Target artifact                                             | Admission evidence                                                       |
|--------------|---------------------------------------------------------------|-------------------------------------------------------------|--------------------------------------------------------------------------|
| CLIC         | Causal mechanisms and intervention operators                  | Latent alternative, confounder, or missing causal direction | Counter-witness interventions, uncertainty, and held-out mechanism tests |
| OPTIC        | Skill vocabulary and its Lie-bracket organization             | New reusable skill, edit, or composition rule               | Transfer, cost reduction, closure, and negative controls                 |
| RELIC        | Sequential model, schema, and decision interface              | New state feature, regime, option, or transition schema     | Return, sample efficiency, transport, and abstention                     |
| AGENTIC      | Typed workflow connecting diagnosis, construction, and action | New workflow edge or allocation policy                      | End-to-end gain beyond isolated components and oracle comparisons        |

| Named object      | Maintained package                                               | Target artifact                                                     | Admission evidence                                                                                                   |
|-------------------|------------------------------------------------------------------|---------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------|
| IPC, AF, DIAL-URL | Finite presentations, symbolic grammars, and behavioral functors | Latent presentation, executable generator, or typed comparison cell | Hidden-point semantics, exact typing, transport certificates, and false-extension controls                           |
| SGTE              | Simulator-grounded scientific theory                             | New mechanism or discriminating experiment                          | Simulator-relative intervention outcomes under a registered interface and held-out regimes                           |
| CTTE              | Claims and relations extracted from a scientific corpus          | New hypothesis and executable study design                          | Corpus-relative provenance, consistency, typed feasibility obligations, and an explicit external-admission interface |
| ARTISTIC          | Scene grammar and image integrity constraints                    | Localized visual repair or enforcement operation                    | Independent structural observer, contract satisfaction, and preservation of unaffected content                       |
| DILATE            | Visual techniques and diffusion-time controls                    | New reusable “paintbrush” acting on generation dynamics             | Cross-prompt reuse, irreducibility, persistence, and critical value                                                  |

**Boundary**

**Evidence-status ledger.** CLIC, OPTIC, and RELIC have controlled component realizations. AGENTIC has bounded workflow-composition demonstrations, while its full domain benchmarks remain specifications. SGTE and CTTE are registered testbed ladders with admitted intermediate rungs, not single finished systems. IPC, AF, and DIAL-URL reach variable-size presentations, typed generator synthesis, and finite comparison-generator recovery under supplied grammars; they do not invent an unrestricted primitive or verify a new distributive law. ARTISTIC has operational assurance and repair evidence under supplied or estimated declarations. DILATE has local bracket, paintbrush, and routing evidence, while construction and admission of a genuinely missing artistic generator remain open. No name in this ledger denotes a demonstrated general creative intelligence.

These distinctions determine which claims the following summary can support; in particular, ARTISTIC's admitted repairs provide an assurance substrate for testing visual theory invention, not evidence that such invention has already occurred.

The experiments reported here provide bounded demonstrations at different levels of maturity. Under their registered contracts, the strongest bounded results show that explicit structural declarations can expose failures that a global loss or fluent generator leaves hidden; active counter-witnesses can improve diagnosis; selective systems can abstain when evidence is insufficient; and admitted repairs can improve planning, causal recovery, proposal construction, or visual coherence. The ARTISTIC studies make the point especially visible: a powerful image generator can produce a plausible cricket, baseball, or tennis scene while violating the rules that make it an instance of that game. A compiled scene theory turns those violations into localized obligations rather than an invitation to regenerate the entire image.

These results do *not* establish general autonomous creativity. Most experiments use controlled worlds, bounded proposal languages, registered observers, or human-curated doctrines. Even the most ambitious studies test whether the architecture can construct a missing reusable structure under a known evaluation contract. That is a necessary calibration for a theory of creative learning, not its completion.

### 20.3 *Why two interacting directions matter*

Piaget's distinction between assimilation and accommodation supplies the book's cognitive intuition. Assimilation improves a model's fit or

competence within its current organization. Accommodation changes that organization when continued assimilation is no longer adequate. Neither direction alone is creative in the sense developed here. Unconstrained accommodation merely changes the rules whenever they become inconvenient; assimilation alone can become increasingly competent inside a theory whose vocabulary is wrong.

The computational hierarchy is more explicit than this psychological analogy. Proto-accommodation probes or proposes a changed organization; accommodation names the finite package change after independent admission; and theory extension is the special case that changes the presentation or generative language.

The double structure records their interaction. Order can matter: adapting a skill and then revising its interface need not agree with revising the interface and then adapting the skill. The mixed bracket makes that order-sensitivity observable, while the quotient removes changes already available along either side separately. Weil probes then ask whether the residual is stable under a controlled family of infinitesimal tests. The result is not a mystical detector of creativity. It is a mechanism for asking a precise question: *does the interaction reveal a direction missing from the current theory?*

This also explains the book's use of the word *infinitesimal*. A finite conceptual change may be proposed because many local comparisons point toward the same obstruction. The infinitesimal layer supplies evidence and localization; a finite package comparison records the proposal, with a sketch extension when the language changes. DIAL therefore does not assert that every breakthrough is a tiny parameter update. It asks whether a discontinuous revision can be made accountable to a structured family of local failures.

#### 20.4 *What DIAL does not explain*

Any account of creativity that begins from a sketch inherits an observability assumption. Someone or something must decide which objects, maps, equations, and probes matter. Initial sketch acquisition partly addresses this issue, but it does not eliminate it. An apprentice can inherit a flawed doctrine; an observer can miss the decisive variable; a proposal language can make the needed extension literally inexpressible.

Several further dimensions lie outside the formalism unless they are added explicitly:

- **Embodiment and material practice.** A sculptor, experimentalist, chef, or athlete learns from resistance, fatigue, instruments, and materials, not only from symbolic constraints.
- **Affect and aesthetic judgment.** Pleasure, tension, beauty, humor,

grief, and cultural resonance are not consequences of categorical coherence.

- **Historical and social contingency.** Communities decide which problems, techniques, and artifacts deserve attention. Institutions, power, resources, and accidents shape what can be discovered and admitted.
- **Analogy and serendipity.** Darwin's use of Malthus, or an unexpected experimental artifact, may depend on a distant relationship that no local probe of the current sketch reveals.
- **Value and responsibility.** A coherent, novel, and reusable theory may still be trivial, harmful, deceptive, or unjust. Admission requires normative governance as well as structural tests.

#### Boundary

DIAL is a compositional framework for creative processes expressible as an interaction between refinement of a maintained theory and controlled extension of that theory. It is not a definition of creativity in general. The framework can organize evidence for a revision; it cannot, by itself, decide what a culture should value or what a person will experience as beautiful, meaningful, or profound.

This boundary is a strength rather than an embarrassment. A useful formalism should state what it makes visible and what it leaves to other sciences. The neuroscience, psychology, sociology, history, and philosophy of creativity remain indispensable partners rather than phenomena to be compiled away.

### 20.5 *Music as an open compositional domain*

Music is a particularly revealing open domain because it is simultaneously formal, embodied, temporal, cultural, and affective. A modest musical sketch can contain pitches, durations, timbres, motifs, voices, chords, and sections. Its operations may include transposition, inversion, augmentation, orchestration, repetition, variation, and voice leading. Equations and limits can express metrical alignment, harmonic expectations, contrapuntal constraints, formal return, and instrument ranges. Models of the same declaration may live as scores, performances, MIDI events, symbolic analyses, or audio distributions.

This makes music a richer target than sequence continuation. Assimilation could improve a melody, accompaniment, orchestration, or performance within a declared musical grammar. Proto-accommodation

could propose a new transformation, tuning relation, rhythmic grammar, timbral operation, or large-scale form; admitted accommodation would install it as a package change. Their interaction might reveal that an artistic intention is not expressible by the admitted operations. But an unfamiliar chord, high model loss, or surprising sound is not yet evidence for a new musical language. The residual must persist across examples and support a reusable operation.

| Layer       | Existing declaration                        | Possible extension          | Admission question                                                       |
|-------------|---------------------------------------------|-----------------------------|--------------------------------------------------------------------------|
| Motif       | Interval and rhythmic transformations       | New motif operator          | Does it generate a coherent family beyond memorized phrases?             |
| Harmony     | Chord and voice-leading grammar             | New relation or tuning rule | Is the rule stable under transposition, voicing, and instrumentation?    |
| Form        | Sections, repetition, contrast, and return  | New organizing operation    | Does it guide multiple complete works rather than one successful sample? |
| Timbre      | Instrument and synthesis controls           | New production “brush”      | Can performers or models reuse it while preserving musical identity?     |
| Performance | Gesture, timing, dynamics, and articulation | New expressive control      | Does it transfer across passages and remain perceptually distinct?       |

### 20.5.1 *A controlled benchmark proposal*

A disciplined DIAL-MUSIC-o experiment would withhold one reusable musical operation from a small declared theory. The corpus would include positive examples generated with the operation, negative examples that are merely unusual, and surface variants that preserve the operation under transposition, tempo, voicing, or instrumentation. The learner would have to:

1. detect a persistent quotient obstruction rather than high prediction error alone;
2. propose an operation equivalent to the withheld generator;

3. transport previously admitted motifs and constraints into the extended theory;
4. generate held-out works that use the operation coherently; and
5. pass independent evaluation by musicians as well as structural and perceptual tests.

Baselines should include a sequence model, a novelty-reward system, a style-confusion objective, and an oracle supplied with the missing operation. Success would not mean producing one pleasant tune. It would mean recovering a teachable musical operation that supports sustained creation. This is the musical analogue of inventing a paintbrush rather than producing a single new image.

Music also exposes DIAL's limits unusually well. A grammar may certify that a piece is coherent without explaining why a performance moves a listener. Conversely, practices dismissed as violations by one theory may be central to another musical culture. The observer, doctrine, and admission community must therefore be recorded as part of the experiment rather than hidden behind a supposedly universal theory of music.

## 20.6 Other domains not explored here

Several domains could test whether the framework transports beyond the applications in this book.

In the table, *creative artifact* names the central construction, not a standalone output. Each row inherits the artifact-dossier obligation of Section 10.4.1: structural delta, transport, independent admission, persistence, and provenance must accompany any strong creativity claim.

| Domain                       | Maintained theory                                               | Creative artifact                                   | Primary difficulty                                          |
|------------------------------|-----------------------------------------------------------------|-----------------------------------------------------|-------------------------------------------------------------|
| Language                     | Lexicon, syntax, pragmatics, and discourse conventions          | New word, metaphor, construction, or genre          | Meaning is negotiated socially and changes with use         |
| Architecture and engineering | Functions, loads, materials, codes, and manufacturing processes | Reusable design principle or fabrication method     | Physical safety and lifecycle cost constrain admission      |
| Cuisine                      | Ingredients, transformations, sensory profiles, and service     | New technique rather than a novel ingredient list   | Taste and embodied practice are difficult to simulate       |
| Software                     | Types, interfaces, tests, and execution semantics               | New abstraction, protocol, or programming pattern   | Apparent novelty may be a renamed familiar construction     |
| Scientific instruments       | Observables, calibration, interventions, and data pipelines     | New measurement or experimental protocol            | The extension changes what evidence can exist               |
| Collaborative creativity     | Distributed theories, roles, and communication channels         | New division of labor or shared conceptual language | Credit, governance, and disagreement are part of the object |

Mathematical field construction remains perhaps the most demand-



ing case. A frontier model may prove a collection of difficult theorems while failing to create a persistent conceptual structure comparable to number theory, Fourier analysis, or Lie theory. In DIAL terms, the artifact would not be an isolated proof but a new presentation: definitions, examples, invariants, theorems, proof techniques, interfaces to prior fields, and a research program that other mathematicians can inhabit. The DIAL-URL experiments provide a small step toward this target, but no experiment in this book approaches its full historical scale.

### 20.7 *Evaluation beyond novelty*

Novelty is cheap when the proposal space is large. Random corruption, adversarial optimization, and unconstrained sampling can all produce artifacts that are unlike a reference set. The more demanding question is whether a candidate extension earns a stable place in a theory. Across domains, at least eight tests recur:

1. **Structural necessity:** does an independently measured obstruction motivate the extension?
2. **Irreducibility:** is the proposal more than an obscure composition or renaming of admitted generators?
3. **Conservativity:** does valid prior knowledge survive?
4. **Closure:** does the proposal remove the targeted obstruction without creating larger unexplained failures?
5. **Productivity:** does it generate a family of useful artifacts, predictions, or actions?
6. **Transport:** does it work across tasks, contexts, models, or media not used to propose it?
7. **Independent value:** do external observers find the extension scientifically, practically, or artistically worthwhile?
8. **Economy and safety:** are the gains worth the acquisition, verification, and error costs?

These criteria turn evaluation from a contest of unusual outputs into an audit of theory change. They also make negative results informative. A system that abstains because the obstruction is transient, the extension is reducible, or the evidence does not transport has learned something about the boundary between novelty and warranted invention.

## 20.8 *Open obligations*

The theoretical construction remains incomplete. A rigorous development must clarify the relationship between classical double Lie algebroids and involution algebroids internal to tangent categories; define quotient obstructions when ranks and presentations vary; establish conditions under which local residuals integrate into finite sketch extensions; and specify when Kan transport is conservative in the chosen doctrine. Equivalent theory presentations require a principled notion of sameness, perhaps Morita-like, so that rediscovering a generator under a new name is not scored as a new theory.

The algorithmic agenda is equally substantial. Systems must acquire sketches and observers rather than receive all of them by hand; estimate uncertainty and abstain under ambiguity; search variable-size proposal languages without rewarding gratuitous complexity; preserve provenance across composed workflows; and separate the cost of diagnosis from the cost of acting in the world. Human and multi-agent settings require explicit protocols for teaching, criticism, disagreement, attribution, and governance.

Finally, benchmarks must test capabilities that existing generative models do not solve by surface fluency. Withheld-generator worlds remain valuable calibrations because their ground truth is known. Domain-grounded simulators, independently annotated images, executable proofs, and prospective scientific experiments supply distinct admission environments with different authorities and blind spots. The long-term test is not whether DIAL can attach a new name to a latent direction, but whether the resulting theory changes what a community can reliably explain, construct, or discover.

## 20.9 *Conclusion*

The framework developed in this book starts from a modest but consequential claim: a learning system can maintain explicit compositional commitments, and their failures can carry more structure than a scalar loss. Tangent and Weil probes make local variation inspectable. Double involution organizes the interaction between improving a theory's models and probing changes to the maintained package. Lie brackets and quotient residuals localize missing directions. A package comparison records the finite proposal; sketch extension records its presentation-changing case. Transport and admission ask whether the change deserves to persist.

This does not solve creativity. It supplies a grammar for asking better questions about a subset of creative acts and an engineering discipline for building systems whose revisions can be inspected. The

resulting picture is less romantic than an inexplicable leap and more ambitious than sampling a surprising artifact. It treats creativity as the construction of new, reusable compositional structure under evidence.

#### Design principle

A creative machine should not merely generate alternatives. It should record which theory generated them, where that theory failed, what extension was proposed, what prior knowledge survived, what evidence admitted the change, and whether others can reuse it.

The question left open is whether this discipline can scale from controlled repairs and withheld generators to discoveries that found new scientific, mathematical, or artistic practices. DIAL is not offered as the last word on that question. It is a categorical starting point from which the question can be stated compositionally, tested experimentally, and revised when the theory itself fails.

#### *Further Reading*

Boden's distinctions among combinational, exploratory, and transformational creativity remain the conceptual point of departure [Boden, 1998, 2004]. Computational criteria for creative systems are developed by Wiggins [2006], Ritchie [2007], Colton [2008], and Jordanous [2012]; the Association for Computational Creativity maintains a broader and continually developing bibliography [Association for Computational Creativity, 2026b]. The psychological and neural scope of creativity is surveyed by Abraham [2018], while Piaget's account of equilibration motivates the assimilation–accommodation distinction used throughout this book [Piaget, 1985].

Chapter 6 situates three complementary surveys at the points where their explanatory scales enter the argument: Sawyer [2012] connects individual cognition to collaborative and cultural processes, Wilson [2018] offers an evolutionary and humanistic perspective, and Miller [2019] records the cross-domain landscape of AI-powered creative practice before the current frontier-model era. They define important phenomena and historical baselines; none is a computational model of DIAL-style theory change.

For musical structure, Lerdahl and Jackendoff [1983] develop a generative theory of tonal organization, Xenakis [1992] treats composition through mathematical constructions, Cope [1996] studies computational composition through reusable stylistic structure, and Tymoczko [2011] gives a geometric account of harmony and voice leading. These works do not instantiate DIAL; they illustrate the depth of the musical theories, practices, and evaluative communities that a credible DIAL-MUSIC system would have to respect.



# Bibliography

Pieter Abbeel and Andrew Y. Ng. Apprenticeship learning via inverse reinforcement learning. In *Proceedings of the Twenty-First International Conference on Machine Learning*, pages 1–8, 2004. DOI: 10.1145/1015330.1015430. URL <https://ai.stanford.edu/~ang/papers/icml04-apprentice.pdf>.

Anna Abraham. *The Neuroscience of Creativity*. Cambridge Fundamentals of Neuroscience in Psychology. Cambridge University Press, 2018. DOI: 10.1017/9781316816981.

Louis J. Aronne, Naveed Sattar, David B. Horn, Harold E. Bays, Sean Wharton, Wan-Ting Lin, Nadia N. Ahmad, Shuyu Zhang, Ruijun Liao, Mathijs C. Bunck, and Irina Jouravskaya. Continued treatment with tirzepatide for maintenance of weight reduction in adults with obesity: The SURMOUNT-4 randomized clinical trial. *JAMA*, 331(1): 38–48, 2024. DOI: 10.1001/jama.2023.24945.

Association for Computational Creativity. Association for computational creativity. Research community and bibliography, 2026a. URL <https://computationalcreativity.net/>. Accessed 2026-08-05.

Association for Computational Creativity. Bibliography of computational creativity. Online bibliography, 2026b. URL <https://computationalcreativity.net/home/resources/bibliography/>. Accessed 2026-08-05.

Association for Computational Creativity. International conference on computational creativity, 2026c. URL <https://computationalcreativity.net/iccc26/iccc/>. Accessed 2026-08-05.

Michael Barr and Charles Wells. *Category Theory for Computing Science*. Centre de Recherches Mathématiques, 3 edition, 1999.

Andrew G. Barto, Marco Mirolli, and Gianluca Baldassarre. Novelty or Surprise? *Frontiers in Psychology*, 4:907, 2013. DOI: 10.3389/fpsyg.2013.00907. URL <https://doi.org/10.3389/fpsyg.2013.00907>.

Margaret A. Boden. Creativity and artificial intelligence. *Artificial Intelligence*, 103(1–2):347–356, 1998. DOI: 10.1016/S0004-3702(98)00055-1.

Margaret A. Boden. *The Creative Mind: Myths and Mechanisms*. Routledge, 2 edition, 2004.

Mikhail M. Bongard. *Pattern Recognition*. Spartan Books, New York, 1970. English translation of the Russian edition first published in 1967.

Garyk Brixi, Matthew G. Durrant, Jerome Ku, Michael Poli, Greg Brockman, Daniel Chang, Gabriel A. Gonzalez, Samuel H. King, David B. Li, Aditi T. Merchant, et al. Genome modeling and design across all domains of life with Evo 2. *bioRxiv*, 2025. DOI: 10.1101/2025.02.18.638918. URL <https://www.biorxiv.org/content/10.1101/2025.02.18.638918v1>.

Fiona Brunk. The development of Galois theory. MacTutor History of Mathematics, University of St Andrews, n.d. URL <https://mathshistory.st-andrews.ac.uk/Projects/Brunk/>. Accessed August 10, 2026.

Matthew Burke and Benjamin MacAdam. Involution algebroids: A generalisation of Lie algebroids for tangent categories, 2019. URL <https://arxiv.org/abs/1904.06594>.

Mingguang Chen, Licheng Wang, and Bo Qu. Recursive self-improvement in AI: From bounded self-refinement to autonomous research loops, 2026.

Noam Chomsky, Ian Roberts, and Jeffrey Watumull. The false promise of ChatGPT. *The New York Times*, March 2023. URL <https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html>.

J. R. B. Cockett and G. S. H. Cruttwell. Differential structure, tangent structure, and synthetic differential geometry. *Applied Categorical Structures*, 22(2):331–417, 2014. DOI: 10.1007/s10485-013-9312-0.

J. R. B. Cockett, G. S. H. Cruttwell, and Jean-Simon Pacaud Lemay. Differential equations in a tangent category I: Complete vector fields, flows, and exponentials, 2021. URL <https://arxiv.org/abs/1911.12120>.

Simon Colton. Creativity versus the perception of creativity in computational systems. In *Proceedings of the AAAI Spring Symposium on Creative Intelligent Systems*, pages 14–20. AAAI Press, 2008.

- Rama Cont. Mimicking diffusion processes with differential equations, 2026. URL <https://arxiv.org/abs/2607.25685>.
- David Cope. *Experiments in Musical Intelligence*. A-R Editions, Madison, WI, 1996.
- Marius Crainic and Rui Loja Fernandes. Integrability of Lie brackets. *Annals of Mathematics*, 157(2):575–620, 2003. DOI: 10.4007/annals.2003.157.575. URL <https://annals.math.princeton.edu/2003/157-2/p06>.
- Charles Darwin. *On the Origin of Species by Means of Natural Selection*. John Murray, London, 1859. URL [https://darwin-online.org.uk/converted/pdf/1859\\_Origin\\_F373.pdf](https://darwin-online.org.uk/converted/pdf/1859_Origin_F373.pdf).
- Darwin Correspondence Project. Darwin in letters, 1837–1843: The London years to natural selection. University of Cambridge, 2015a. URL <https://www.darwinproject.ac.uk/letters/darwins-life-letters/darwin-letters-1837-1843-london-years-natural-selection>. Accessed August 10, 2026.
- Darwin Correspondence Project. Natural selection. University of Cambridge, 2015b. URL <https://www.darwinproject.ac.uk/commentary/evolution/natural-selection>. Accessed August 10, 2026.
- Richard Dawkins. *The Blind Watchmaker: Why the Evidence of Evolution Reveals a Universe Without Design*. W. W. Norton & Company, New York, 1986. URL <https://wnorton.co.uk/books/9780393351491-the-blind-watchmaker>.
- Stefan Depeweg, Constantin A. Rothkopf, and Frank Jäkel. Solving Bongard problems with a visual language and pragmatic reasoning, 2018. URL <https://arxiv.org/abs/1804.04452>.
- René Descartes. *The Geometry of René Descartes*. Dover Publications, 1954. French original published in 1637.
- Gary L. Drescher. The schema mechanism: A conception of constructivist intelligence. Master’s thesis, Massachusetts Institute of Technology, Cambridge, MA, February 1985.
- Gary L. Drescher. Genetic AI—translating Piaget into Lisp. *Instructional Science*, 14(3–4):357–380, 1986. DOI: 10.1007/BF00051828. Also issued as MIT Artificial Intelligence Laboratory Memo 890.
- Gary L. Drescher. *Made-Up Minds: A Constructivist Approach to Artificial Intelligence*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, September 1989.

Gary L. Drescher. *Made-Up Minds: A Constructivist Approach to Artificial Intelligence*. MIT Press, Cambridge, MA, 1991. URL <https://mitpress.mit.edu/9780262041201/made-up-minds/>.

V. G. Drinfel'd. Hamiltonian structures on Lie groups, Lie bialgebras and the geometric meaning of the classical Yang–Baxter equations. *Soviet Mathematics Doklady*, 27:68–71, 1983. URL <https://www.mathnet.ru/eng/dan45858>. English translation of Doklady Akademii Nauk SSSR 268(2), 285–287.

Frank Watson Dyson, Arthur Stanley Eddington, and Charles Rundle Davidson. A determination of the deflection of light by the sun's gravitational field, from observations made at the total eclipse of may 29, 1919. *Philosophical Transactions of the Royal Society of London. Series A*, 220:291–333, 1920. DOI: 10.1098/rsta.1920.0009.

Charles Ehresmann. Esquisses et types des structures algébriques. *Buletinul Institutului Politehnic din Iași*, 14:1–14, 1968.

Albert Einstein. Die grundlage der allgemeinen relativitätstheorie. *Annalen der Physik*, 354(7):769–822, 1916. DOI: 10.1002/andp.19163540702.

Ahmed Elgammal, Bingchen Liu, Mohamed Elhoseiny, and Marian Mazzone. CAN: Creative adversarial networks, generating “art” by learning about styles and deviating from style norms. In *Proceedings of the Eighth International Conference on Computational Creativity*, pages 96–103, 2017. URL <https://arxiv.org/abs/1706.07068>.

Euclid. *The Thirteen Books of Euclid's Elements*. Dover Publications, 2 edition, 1956. Originally composed circa 300 BCE.

Damon Falck, Samer Sabri, Anja Surina, Thom Foster, Anya Sims, Sam Devlin, Dylan Rogers, Tantum Collins, Kaloyan Aleksiev, Louis Kirsch, and Edward Hughes. Training AI scientists to replicate research, 2026. URL <https://arxiv.org/abs/2608.13331>. Faraday and the Replica benchmark, developed at Inherent Laboratories.

Weixi Feng, Xuehai He, Tsu-Jui Fu, Varun Jampani, Arjun Akula, Pradyumna Narayana, Sugato Basu, Xin Eric Wang, and William Yang Wang. Training-free structured diffusion guidance for compositional text-to-image synthesis, 2022. URL <https://arxiv.org/abs/2212.05032>.

Richard P. Feynman, Robert B. Leighton, and Matthew Sands. *The Feynman Lectures on Physics, Volume I: Mainly Mechanics, Radiation, and Heat*. Addison-Wesley, Reading, MA, 1963. URL [https://www.feynmanlectures.caltech.edu/I\\_01.html](https://www.feynmanlectures.caltech.edu/I_01.html).



Harry E. Foundalis. Index of Bongard problems, 2026. URL <https://www.foundalis.com/res/bps/bpidx.htm>. Accessed 15 August 2026.

Rosalind E. Franklin and Raymond G. Gosling. Molecular configuration in sodium thymonucleate. *Nature*, 171(4356):740–741, 1953. DOI: 10.1038/171740a0.

Rinon Gal, Yuval Alaluf, Yuval Atzmon, Or Patashnik, Amit H. Bermano, Gal Chechik, and Daniel Cohen-Or. An image is worth one word: Personalizing text-to-image generation using textual inversion, 2022. URL <https://arxiv.org/abs/2208.01618>.

Richard Garner and Stephen Lack. Grothendieck quasitoposes. *Journal of Algebra*, 355:111–127, 2012. DOI: 10.1016/j.jalgebra.2012.01.006.

W. Timothy Garvey, Rachel L. Batterham, Manoj Bhatta, Silvio Buscemi, Louise N. Christensen, Juan P. Frias, Niklas R. Jørgensen, Klaus Kandler, George Rigas, Thomas A. Wadden, and Sean Wharton. Two-year effects of semaglutide in adults with overweight or obesity: The STEP 5 trial. *Nature Medicine*, 28:2083–2091, 2022. DOI: 10.1038/s41591-022-02026-4.

A. Gracia-Saz, M. Jotz Lean, K. C. H. Mackenzie, and R. A. Mehta. Double Lie algebroids and representations up to homotopy, 2014. URL <https://arxiv.org/abs/1409.1502>.

Gustav A. Hedlund. Endomorphisms and automorphisms of the shift dynamical system. *Mathematical Systems Theory*, 3:320–375, 1969. DOI: 10.1007/BF01691062.

HKUST KnowComp. Awesome LLM scientific discovery. Continuously updated research bibliography and software repository, 2026. URL <https://github.com/HKUST-KnowComp/Awesome-LLM-Scientific-Discovery>. Accessed 2026-08-06.

Kexin Huang, Serena Zhang, Hanchen Wang, Yuanhao Qu, Yingzhou Lu, Yusuf Roohani, Ryan Li, Lin Qiu, Gavin Li, Junze Zhang, et al. Biomni: A general-purpose biomedical AI agent. *bioRxiv*, 2025. DOI: 10.1101/2025.05.30.656746. URL <https://www.biorxiv.org/content/10.1101/2025.05.30.656746v1>.

Hugging Face. Textual inversion in Diffusers. Diffusers documentation and reference implementation, 2026. URL [https://huggingface.co/docs/diffusers/en/training/text\\_inversion](https://huggingface.co/docs/diffusers/en/training/text_inversion). Accessed 14 August 2026.

Emily C. Isko, Clifford E. Harpole, Xiaoyue Mike Zheng, Huiqing Zhan, Martin B. Davis, Anthony M. Zador, and Arkarup Banerjee. Selective

expansion of motor cortical projections in the evolution of vocal novelty. *bioRxiv*, 2024. DOI: 10.1101/2024.09.13.612752.

Nikki Goth Itoi. How AI is accelerating scientific discovery. Stanford Institute for Human-Centered Artificial Intelligence, July 2026. URL <https://hai.stanford.edu/news/how-ai-is-accelerating-scientific-discovery>. Published 8 July 2026; accessed 2026-08-06.

Ania M. Jastreboff, Louis J. Aronne, Nadia N. Ahmad, Sean Wharton, Laura Connery, Breno Alves, Arihiro Kiyosue, Shuyu Zhang, Bing Liu, Mathijs C. Bunck, and Anne Stefanski. Tirzepatide once weekly for the treatment of obesity. *New England Journal of Medicine*, 387(3): 205–216, 2022. DOI: 10.1056/NEJMoa2206038.

Anna Jordanous. A standardised procedure for evaluating creative systems: Computational creativity evaluation based on what it is to be creative. *Cognitive Computation*, 4(3):246–279, 2012. DOI: 10.1007/s12559-012-9156-1.

G. M. Kelly. *Basic Concepts of Enriched Category Theory*, volume 64 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, 1982. URL <https://tac.mta.ca/tac/reprints/articles/10/tr10.pdf>.

Ross D. King, Jem Rowland, Stephen G. Oliver, Michael Young, Wayne Aubrey, Emma Byrne, Maria Liakata, Madeleine Markham, Pinar Pir, Larisa N. Soldatova, Andrew Sparkes, Kenneth E. Whelan, and Amanda Clare. The automation of science. *Science*, 324(5923):85–89, 2009. DOI: 10.1126/science.1165620.

Dexter Kozen and Nicholas Ruoizzi. Applications of metric coinduction. *Logical Methods in Computer Science*, 5(3), 2009. DOI: 10.2168/LMCS-5(3:10)2009.

Hiroshi Kunita. Stochastic differential equations and stochastic flows of diffeomorphisms. In *'Ecole d'Eté de Probabilités de Saint-Flour XII–1982*, volume 1097 of *Lecture Notes in Mathematics*, pages 143–303. Springer, 1984. DOI: 10.1007/BFb0099433.

Stephen Lack. Composing PROPs. *Theory and Applications of Categories*, 13(9):147–163, 2004. URL <https://tac.mta.ca/tac/volumes/13/9/13-09abs.html>.

Imre Lakatos. *Proofs and Refutations: The Logic of Mathematical Discovery*. Cambridge University Press, 1976.

- Pat Langley, Herbert A. Simon, Gary L. Bradshaw, and Jan M. Żytkow. *Scientific Discovery: Computational Explorations of the Creative Process*. The MIT Press, Cambridge, MA, 1987. ISBN 9780262121163. URL <https://mitpress.mit.edu/9780262121163/scientific-discovery/>.
- F. William Lawvere. Functorial semantics of algebraic theories. *Proceedings of the National Academy of Sciences*, 50(5):869–872, 1963. DOI: 10.1073/pnas.50.5.869. URL <https://pmc.ncbi.nlm.nih.gov/articles/PMC221940/>.
- Douglas B. Lenat. *AM: An Artificial Intelligence Approach to Discovery in Mathematics as Heuristic Search*. PhD thesis, Stanford University, 1976. Stanford Artificial Intelligence Laboratory Memo AIM-286; Computer Science Report STAN-CS-76-750.
- Fred Lerdaahl and Ray Jackendoff. *A Generative Theory of Tonal Music*. The MIT Press, Cambridge, MA, 1983.
- Poon Leung. Classifying tangent structures using Weil algebras. *Theory and Applications of Categories*, 32(9):286–337, 2017. URL <http://www.tac.mta.ca/tac/volumes/32/9/32-09abs.html>.
- Poon Leung. Tangent bundles, monoidal theories and Weil algebras. *Bulletin of the Australian Mathematical Society*, 98(1):175–176, 2018. DOI: 10.1017/S0004972718000102. Abstract of the author’s 2017 Macquarie University PhD thesis.
- A. Michael Lincoff, Kirstine Brown-Frandsen, Helen M. Colhoun, John Deanfield, Scott S. Emerson, Sille Esbjerg, Soren Hardt-Lindberg, G. Kees Hovingh, Steven E. Kahn, Robert F. Kushner, Ildiko Lingvay, Elif A. Oral, Mette M. Michelsen, Jorge Plutzky, Christoffer W. Tornøe, and Donna H. Ryan. Semaglutide and cardiovascular outcomes in obesity without diabetes. *New England Journal of Medicine*, 389(24):2221–2232, 2023. DOI: 10.1056/NEJMoa2307563.
- Yu Liu, Hongbo Fu, Hongliang Zhang, Yunhang Wang, Rajan K. Chakrabarty, Xiang Tu, Xu Tang, Alexander Laskin, Gregory R. Carmichael, Jianmin Chen, Joseph S. Francisco, and Drew T. Shindell. Atmospheric warming contributions from airborne microplastics and nanoplastics. *Nature Climate Change*, 2026. DOI: 10.1038/s41558-026-02620-1.
- Saunders Mac Lane. *Categories for the Working Mathematician*, volume 5 of *Graduate Texts in Mathematics*. Springer, 2 edition, 1998. DOI: 10.1007/978-1-4757-4721-8.

Saunders Mac Lane and Ieke Moerdijk. *Sheaves in Geometry and Logic: A First Introduction to Topos Theory*. Springer, 1992. DOI: 10.1007/978-1-4612-0927-0.

Benjamin MacAdam. The functorial semantics of Lie theory, 2023. URL <https://arxiv.org/abs/2301.00305>.

Kirill C. H. Mackenzie. Notions of double for Lie algebroids, 2000a. URL <https://arxiv.org/abs/math/0011212>.

Kirill C. H. Mackenzie. Double Lie algebroids and second-order geometry, II. *Advances in Mathematics*, 154(1):46–75, 2000b. DOI: 10.1006/aima.1999.1892.

Kirill C. H. Mackenzie. *General Theory of Lie Groupoids and Lie Algebroids*, volume 213 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, 2005.

Sridhar Mahadevan. Imagination machines: A new challenge for artificial intelligence. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, pages 7988–7993. AAAI Press, 2018. URL <https://ojs.aaai.org/index.php/AAAI/article/view/12214>.

Sridhar Mahadevan. Latent confounded causal discovery via Lie bracket geometry, 2026a. URL <https://arxiv.org/abs/2606.19610>.

Sridhar Mahadevan. Categories for AGI. Book manuscript, 2026b. URL <https://people.cs.umass.edu/~mahadeva/papers/catagi.pdf>.

Sridhar Mahadevan. Infinitesimal causality. *arXiv preprint arXiv:2606.24621*, 2026c. URL <https://arxiv.org/abs/2606.24621>.

Sridhar Mahadevan. Agentic skill optimization over Lie algebroids. *arXiv preprint arXiv:2607.11493*, 2026d. URL <https://arxiv.org/abs/2607.11493>.

Sridhar Mahadevan. Machine learning from enforcing compositionality: A categorical framework for structural diagnosis and repair. Book manuscript, 2026e.

Sridhar Mahadevan. PROMETHEUS: Automating deep causal research integrating text, data and models, 2026f. URL <https://arxiv.org/abs/2605.12835>.

Sridhar Mahadevan, Tom M. Mitchell, Jack Mostow, Lou Steinberg, and Prasad V. Tadepalli. An apprentice-based approach to knowledge acquisition. *Artificial Intelligence*, 64(1):1–52, 1993. DOI: 10.1016/0004-3702(93)90059-K.

- John McCarthy. Challenges to machine learning: Relations between reality and appearance. In *Inductive Logic Programming: 16th International Conference, ILP 2006, Revised Selected Papers*, volume 4455 of *Lecture Notes in Computer Science*, pages 2–9. Springer, 2007. DOI: 10.1007/978-3-540-73847-3\_2.
- Eckhard Meinrenken and Jeffrey Pike. The Weil algebra of a double Lie algebroid. *International Mathematics Research Notices*, 2021(11): 8550–8622, 2021. URL <https://arxiv.org/abs/1901.00230>.
- Arthur I. Miller. *The Artist in the Machine: The World of AI-Powered Creativity*. The MIT Press, Cambridge, MA, 2019. ISBN 9780262042857. DOI: 10.7551/mitpress/11585.001.0001. URL <https://mitpress.mit.edu/9780262042857/the-artist-in-the-machine/>.
- Tom M. Mitchell, Sridhar Mahadevan, and Louis I. Steinberg. LEAP: A learning apprentice for VLSI design. In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, pages 573–580, 1985. URL <https://www.ijcai.org/Proceedings/85-1/Papers/109.pdf>.
- Tom M. Mitchell, Richard M. Keller, and Smadar T. Kedar-Cabelli. Explanation-based generalization: A unifying view. *Machine Learning*, 1(1):47–80, 1986. DOI: 10.1007/BF00116251.
- Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518:529–533, 2015. DOI: 10.1038/nature14236.
- Raymond J. Mooney and Scott W. Bennett. A domain independent explanation-based generalizer. In *Proceedings of the Fifth National Conference on Artificial Intelligence*, pages 551–555, 1986. URL <https://cdn.aaai.org/AAAI/1986/AAAI86-093.pdf>.
- National Institutes of Health. Simplified peer review framework. NIH Grants and Funding, 2026. URL <https://www.grants.nih.gov/policy-and-compliance/policy-topics/peer-review/simplifying-review/framework>. Accessed 20 August 2026.
- Open Conference of AI Agents for Science. Call for papers: Open Conference of AI Agents for Science 2025. Stanford University, 2025. URL <https://agents4science.stanford.edu/call-for-papers.html>. Accessed 20 August 2026.

Oxford University Press. What's new in the Oxford English Dictionary? June 2026 update, June 2026. URL <https://corp.oup.com/news/oxford-english-dictionary-update-june-2026/>. Quarterly update describing new and revised words, senses, and entries.

William Paley. *Natural Theology; or, Evidences of the Existence and Attributes of the Deity*. R. Faulder, London, 1802. URL <https://quod.lib.umich.edu/g/genpub/paleynatur>.

Günther Palm. *Novelty, Information and Surprise*. Information Science and Statistics. Springer, Berlin, Heidelberg, 2012. DOI: 10.1007/978-3-642-29075-6.

Deepak Pathak, Pulkit Agrawal, Alexei A. Efros, and Trevor Darrell. Curiosity-driven exploration by self-supervised prediction. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 2778–2787, 2017. URL <https://proceedings.mlr.press/v70/pathak17a.html>.

Linus Pauling and Robert B. Corey. A proposed structure for the nucleic acids. *Proceedings of the National Academy of Sciences*, 39(2):84–97, 1953. DOI: 10.1073/pnas.39.2.84.

Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2 edition, 2009.

Jean Piaget. *The Origins of Intelligence in Children*. International Universities Press, New York, 1952.

Jean Piaget. *The Construction of Reality in the Child*. Routledge, 1954. URL <https://www.routledge.com/The-Construction-Of-Reality-In-The-Child/Piaget/p/book/9780415210003>.

Jean Piaget. *The Equilibration of Cognitive Structures: The Central Problem of Intellectual Development*. University of Chicago Press, Chicago, 1985.

Jean Piaget and Bärbel Inhelder. *The Child's Construction of Quantities: Conservation and Atomism*. Routledge & Kegan Paul, London, 1974. Originally published in French in 1941.

Jeffrey Pike. *Weil Algebras and Double Lie Algebroids*. Ph.d. thesis, University of Toronto, 2020. Theses Canada OCLC 1334505821.

Recursive Superintelligence. Recursive superintelligence, 2026. URL <https://www.recursive.com/>. Accessed 2026-08-05.

Ricursive Intelligence. Recursive self-improvement via AI for chip design and chip design for AI, 2026. URL <https://www.ricursive.com/>. Accessed 2026-08-05.

Emily Riehl. *Category Theory in Context*. Dover Publications, 2016. URL <https://math.jhu.edu/~eriehl/161/context.pdf>.

Graeme D. Ritchie. Some empirical criteria for attributing creativity to a computer program. *Minds and Machines*, 17(1):67–99, 2007. DOI: 10.1007/s11023-007-9066-2.

Domenica Rubino, Niclas Abrahamsson, Melanie Davies, Dorte Hesse, Frank L. Greenway, Christian Jensen, Ildiko Lingvay, Ofri Mosenzon, Julio Rosenstock, Miguel A. Rubio, Gottfried Rudofsky, Shadi Tadayon, Thomas A. Wadden, and Dror Dicker. Effect of continued weekly subcutaneous semaglutide vs placebo on weight loss maintenance in adults with overweight or obesity. *JAMA*, 325(14):1414–1425, 2021. DOI: 10.1001/jama.2021.3224.

Jan J. M. M. Rutten. Universal coalgebra: A theory of systems. *Theoretical Computer Science*, 249(1):3–80, 2000. DOI: 10.1016/S0304-3975(00)00056-6.

Sakana AI. Introducing Sakana AI’s recursive self-improvement (RSI) lab, 2026. URL <https://sakana.ai/rsi-lab/>. Accessed 2026-08-05.

R. Keith Sawyer. *Explaining Creativity: The Science of Human Innovation*. Oxford University Press, New York, 2nd edition, 2012. ISBN 9780199737574. URL <https://global.oup.com/academic/product/explaining-creativity-9780199737574>.

Michael Schmidt and Hod Lipson. Distilling free-form natural laws from experimental data. *Science*, 324(5923):81–85, 2009. DOI: 10.1126/science.1165893.

Rico Sennrich, Barry Haddow, and Alexandra Birch. Neural machine translation of rare words with subword units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1715–1725, Berlin, Germany, 2016. Association for Computational Linguistics. DOI: 10.18653/v1/P16-1162. URL <https://aclanthology.org/P16-1162/>.

Mohit Shridhar, Xingdi Yuan, Marc-Alexandre Cote, Yonatan Bisk, Adam Trischler, and Matthew Hausknecht. ALFWorld: Aligning text and embodied environments for interactive learning. In *International Conference on Learning Representations*, 2021. URL <https://openreview.net/forum?id=0IOX0YcCdTn>.

Ana Sokolova. Probabilistic systems coalgebraically: A survey. *Theoretical Computer Science*, 412(38):5095–5110, 2011. DOI: 10.1016/j.tcs.2011.05.008.



- Hiren Solanki, Vikrant Jain, Kaustubh Thirumalai, Balaji Rajagopalan, and Vimal Mishra. River drought forcing of the Harappan metamorphosis. *Communications Earth & Environment*, 6:926, 2025. DOI: 10.1038/s43247-025-02901-1.
- Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021. URL <https://arxiv.org/abs/2011.13456>.
- Zhangde Song, Jieyu Lu, Yuanqi Du, Botao Yu, Thomas M. Pruyn, Yue Huang, Kehan Guo, Xiuzhe Luo, Yuanhao Qu, et al. Evaluating large language models in scientific discovery, 2025. Version 2, 8 May 2026.
- Luca Stefanini. *On Morphic Actions and Integrability of LA-Groupoids*. PhD thesis, Università degli Studi di Roma Tor Vergata, 2009. URL <https://arxiv.org/abs/0902.2228>.
- Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA, 2 edition, 2018. URL <http://incompleteideas.net/book/the-book-2nd.html>.
- Richard S. Sutton, Doina Precup, and Satinder Singh. Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence*, 112(1–2):181–211, 1999. DOI: 10.1016/S0004-3702(99)00052-1.
- Dmitri Tymoczko. *A Geometry of Music: Harmony and Counterpoint in the Extended Common Practice*. Oxford University Press, New York, 2011.
- Silviu-Marian Udrescu and Max Tegmark. AI Feynman: A physics-inspired method for symbolic regression. *Science Advances*, 6(16): eaay2631, 2020. DOI: 10.1126/sciadv.aay2631.
- John von Neumann. *Theory of Self-Reproducing Automata*. University of Illinois Press, Urbana, IL, 1966. Edited and completed by Arthur W. Burks.
- Ziyan Wang and Longlong Ma. Large language models and impossible language acquisition: “false promise” or an overturn of our current perspective towards AI, 2026. URL <https://arxiv.org/abs/2602.08437>.
- James D. Watson. *The Double Helix: A Personal Account of the Discovery of the Structure of DNA*. Atheneum, New York, 1968.
- James D. Watson and Francis H. C. Crick. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, 171(4356): 737–738, 1953. DOI: 10.1038/171737a0.



Geraint A. Wiggins. A preliminary framework for description, analysis and comparison of creative systems. *Knowledge-Based Systems*, 19(7): 449–458, 2006. DOI: 10.1016/j.knosys.2006.04.009.

Samuel Wilberforce. Review of *On the Origin of Species*. *Quarterly Review*, 108:225–264, 1860. URL <https://darwin-online.org.uk/content/frameset?itemID=A19&viewtype=text&pageseq=1>.

John P. H. Wilding, Rachel L. Batterham, Salvatore Calanna, Melanie Davies, Luc F. Van Gaal, Ildiko Lingvay, Barbara M. McGowan, Julio Rosenstock, Marie T. D. Tran, Thomas A. Wadden, Sean Wharton, Koutaro Yokote, Niels Zeuthen, and Robert F. Kushner. Once-weekly semaglutide in adults with overweight or obesity. *New England Journal of Medicine*, 384(11):989–1002, 2021. DOI: 10.1056/NEJMoa2032183.

Edward O. Wilson. *The Origins of Creativity*. Liveright Publishing Corporation, New York, 2018. ISBN 9781631494857. URL <https://wnorton.com/books/9781631494857>.

Stephen Wolfram. *A New Kind of Science*. Wolfram Media, Champaign, IL, 2002. ISBN 9781579550080. URL <https://www.wolframscience.com/nks/>.

Iannis Xenakis. *Formalized Music: Thought and Mathematics in Composition*. Pendragon Press, Stuyvesant, NY, revised edition, 1992.

Yifan Yang, Ziyang Gong, WeiQuan Huang, Qihao Yang, Ziwei Zhou, Zisu Huang, Yan Li, Xuemei Gao, Qi Dai, Bei Liu, Kai Qiu, Yuqing Yang, Dongdong Chen, Xue Yang, and Chong Luo. Skil-IOpt: Executive strategy for self-evolving agent skills, 2026. URL <https://arxiv.org/abs/2605.23904>.

Tom Zahavy. Position: LLMs can’t jump. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026. URL <https://openreview.net/forum?id=klU4737opt>. Position paper.

Tianshi Zheng, Zheyue Deng, Hong Ting Tsang, Weiqi Wang, Jiaxin Bai, Zihao Wang, and Yangqiu Song. From automation to autonomy: A survey on large language models in scientific discovery, 2025.



# Glossary of Notation

| Term or symbol                                                       | Meaning in this book                                                                                                                                                                                            |
|----------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| $\mathcal{S}$                                                        | A declared conceptual space or algebraic theory.                                                                                                                                                                |
| $D$                                                                  | A current realization or instance of a declared theory.                                                                                                                                                         |
| $\mathcal{D}$                                                        | A versioned collection of scientific documents and structured sources in Chapter 17; distinguished from the theory realization $D$ .                                                                            |
| $P_D$                                                                | The provenance-bearing presentation extracted from $D$ : candidate objects, arrows, contexts, equations, evidence types, and source-span links.                                                                 |
| $S_D$                                                                | A source-bearing sketch obtained by reconciling the local presentation $P_D$ .                                                                                                                                  |
| $\text{Compile}_\theta(D)$                                           | The ontology- and extractor-relative family of candidate source-bearing sketches, support values, and provenance maps compiled from a document collection. It is not assumed to be a unique functor from prose. |
| $\text{Test}_D$                                                      | The admission interface compiled from a document-grounded theory: an observation protocol, intervention, simulator, manipulable model, or locked evidence query with discriminating outcomes.                   |
| $\Lambda$                                                            | A family of observers through which failures or consequences become detectable.                                                                                                                                 |
| $\mathfrak{R}, \mathfrak{R}_{\text{diff}}$                           | A maintained representational package and its differential refinement $(\mathcal{S}, \mathcal{D}, \mathcal{M}, F, \Lambda, \mathcal{G}, \mathcal{A})$ .                                                         |
| $\mathcal{E} = (\mathcal{I}, \mathcal{D}, \mathcal{F}, \mathcal{X})$ | Initial sketch-acquisition evidence: explicit instruction, expert demonstrations, corrective feedback, and the learner's own interventions.                                                                     |
| $(\hat{\mathfrak{R}}_0, \hat{D}_0, \Gamma_0, U_0)$                   | An acquired provisional representational package, a current realization typed by it, its evidential support and provenance record, and its unresolved or unsupported alternatives.                              |
| Imitation                                                            | Open-loop sketch acquisition from observed expert demonstrations, without corrections conditioned on the learner's own attempts.                                                                                |
| Apprenticeship                                                       | Closed-loop sketch acquisition in which a teacher observes the learner's attempts and supplies adaptive corrections, contrasts, explanations, or counterexamples.                                               |

| Term or symbol                                        | Meaning in this book                                                                                                                                                                                                                                                                                                                                                                              |
|-------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| $\mathfrak{R}_{\text{dbl}}$                           | A double repair package: a differential representational package together with a candidate DIAL object and, when supplied, its classical cochain realization.                                                                                                                                                                                                                                     |
| $\mathfrak{D}, \mathcal{M}, \mathcal{G}, \mathcal{A}$ | Respectively, the preservation doctrine, semantic universe, generative operations, and admission procedures.                                                                                                                                                                                                                                                                                      |
| $\mathfrak{C}_X$                                      | The complete DIAL-X algorithmic contract: target, creative mandate when applicable, maintained realization, two repair theories and their operations, observers, mixed witness and localization record, finite-realization test, proposal language, less-disruptive control suite, package comparison, admission rule, and memory. It is distinct from the preservation doctrine $\mathfrak{D}$ . |
| $\text{Loc}_X$                                        | The observer-relative localization record for the mixed witness $\Theta_X$ . It is distinct from the semantic-transport functor $K$ , and may be empty when the registered observers do not support a meaningful localization.                                                                                                                                                                    |
| $S^+$                                                 | A proposed extension or transformation of $S$ .                                                                                                                                                                                                                                                                                                                                                   |
| $J$                                                   | A candidate presentation-extension map, $J : S \rightarrow S^+$ , together with an interpretation of its new generators and relations. It records the presentation component of a proposed package change, not by itself the semantic, observer, doctrine, or admission comparison, and it remains proposed until transport and consequences are independently admitted.                          |
| $Y_X : \mathfrak{R}_X \rightarrow \mathfrak{R}_X^+$   | The package comparison for DIAL-X, recording changes to the presentation, semantics, doctrine, observers, generative operations, proposal grammar, or admission interface. It is the identity for an unchanged package; a sketch map $J_X$ is one component when the presentation changes.                                                                                                        |
| $\omega$                                              | A localized obstruction motivating a possible theory change.                                                                                                                                                                                                                                                                                                                                      |
| $\gamma$                                              | A finite-realization record or integration observer testing whether a candidate local repair persists and composes coherently beyond its diagnostic neighborhood.                                                                                                                                                                                                                                 |
| $\Xi$                                                 | The status-bearing extension dossier containing prior and proposed representational packages and realizations, obstruction, finite-realization record, componentwise package comparison $Y$ , transport, predictions, and epistemic status. A theory map $J$ appears when the presentation or generative language changes.                                                                        |
| $\text{Weil}_1^{\text{tan}}$                          | Leung's monoidal theory of algebraic probe shapes classifying ordinary tangent structure. It is distinct from the cochain Weil algebra of a Lie algebroid.                                                                                                                                                                                                                                        |

| Term or symbol             | Meaning in this book                                                                                                                                                                                                                                                                                                                   |
|----------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| $F$                        | The strong monoidal functor realizing Weil probes as endofunctors of a conceptual universe.                                                                                                                                                                                                                                            |
| $\Theta_{\text{int}}$      | The proposed internal interchange defect between two tangent-categorical algebroid directions. It is meaningful before choosing a classical cochain realization.                                                                                                                                                                       |
| Real                       | A supplied cochain realization sending the proposed internal interchange defect to its classical graded-commutator observer, $\text{Real}(\Theta_{\text{int}}) = \Omega_{\text{cre}}$ . It is distinct from the frozen DIAL-X reference space $\mathcal{R}_X$ .                                                                        |
| $\text{DIAL}(\mathcal{M})$ | The proposed category of Double Involution ALgebroids internal to a suitable tangent category $\mathcal{M}$ ; constructing it and proving its realization properties are open objectives of this book.                                                                                                                                 |
| DIAL-X                     | An algorithmic specialization of DIAL to a base system $X$ , defined by an operational target, structural declaration, two typed repair operations, mixed witness, finite-realization observer, optional creative mandate, proposal language, less-disruptive controls, package comparison, admission contract, and persistent memory. |
| DIAL-SkillOpt              | A training method for an external, typed discovery skill. A frozen agent executes the skill over DIAL states and curve-object flows; a separate optimizer accepts bounded edits by strict held-out validation improvement; a locked split supplies final independent admission.                                                        |
| CLIC                       | Causal Learning with Infinitesimal Creativity: a DIAL-X realization that uses uncertain intervention geometry, active counter-witnesses, explicit unsupported alternatives, and typed causal package comparison. A causal-sketch map is required only when the presentation changes.                                                   |
| OPTIC                      | Optimizing Programs Through Infinitesimal Creativity: a DIAL-X realization that combines involution-algebroid skill geometry, versioned external programs, obstruction-triggered workflow construction, and independent execution admission.                                                                                           |
| RELIC                      | Reinforcement Learning with Infinitesimal Creativity: a DIAL-X realization that augments policy learning with obstruction-triggered construction, independent admission, persistence, and transport of decision structure.                                                                                                             |
| AGENTIC                    | Acquiring Generalizable Novel Theories with Infinitesimal Creativity: the proposed integrated architecture that composes CLIC, OPTIC, and RELIC through typed interfaces, structural packets, and independent component and interface admission.                                                                                       |

| Term or symbol                                        | Meaning in this book                                                                                                                                                                                                                                                                                                                            |
|-------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ARTISTIC                                              | Acquiring Representational Theories through Infinitesimal Semantic Transport in Image Creation: the AGENTIC text-to-image realization that couples generative transport to visual-language diagnosis, repair, bounded extension tests, and independent structural admission.                                                                    |
| DIAL-CAN                                              | The conceptual and experimental bridge from CAN-style statistical deviation to DIAL-based artistic theory extension. Its archived bracket and paintbrush experiments retain this prefix for provenance.                                                                                                                                         |
| DILATE                                                | Double Involution Learning for Artistic Theory Extension: the proposed paintbrush-discovery architecture that projects mixed obstructions away from registered visual operations, proposes reusable trajectory techniques, transports artistic intent through local observer geometry, and submits visual-theory extensions to typed admission. |
| $Q_{\text{art}} = \text{coker}(\rho_A \oplus \rho_B)$ | The artistic obstruction quotient on a regular model stratum: local visual-model directions modulo those generated by registered assimilatory and proto-accommodative operations. When anchor rank varies, a stratified or sheaf-theoretic replacement is required.                                                                             |
| $q_{\text{art}} : TM \rightarrow Q_{\text{art}}$      | The quotient map that removes directions already generated by the registered assimilatory and proto-accommodative anchors. It is distinct from the query component $q$ of a structural packet.                                                                                                                                                  |
| $\Omega_{\text{art}}(a, b)$                           | The artistic closure obstruction $q_{\text{art}}([\rho_A(a), \rho_B(b)])$ , retaining only the mixed bracket component that the current visual theory cannot express. On a constant-rank stratum it is tensorial in $a$ and $b$ .                                                                                                               |
| $(C, \rho_C, c)$                                      | A candidate artistic-generator bundle, its anchor into observable model-space directions, and a local section. The proposal condition is $q_{\text{art}}(\rho_C(c)) \simeq \Omega_{\text{art}}(a, b)$ ; this quotient equality does not itself choose a canonical lift or type the new generator.                                               |
| $\eta_m : m \rightarrow i^* \text{Lan}_i m$           | The unit comparing an old artistic model with the restriction of its proposed left-Kan-extended transport. Conservativity requires this map to meet the registered equivalence standard; existence of the Kan extension alone is insufficient.                                                                                                  |
| Weil-persistent artistic obstruction                  | A natural family of lifted quotient residuals that remains nontrivial, directionally and locally coherent, and observer-robust across a registered family of Weil probes.                                                                                                                                                                       |
| $S_{\text{COR}}$                                      | The meta-sketch presenting legal interfaces among the current CLIC causal sketch, OPTIC skill sketch, and RELIC decision sketch.                                                                                                                                                                                                                |

| Term or symbol                                                 | Meaning in this book                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|----------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Structural packet                                              | The typed exchange object $(\tau, \mathfrak{R}, Y, \omega, U, \gamma, h, q, c, s, \kappa, v)$ carrying packet type, prior representational package, proposed package comparison, localized obstruction, uncertainty, finite-realization record, registered-control record, query, status-bearing admission record, epistemic status, interface contract, and version between creative workflows. A changed sketch map $J$ is stored as the presentation component of $Y$ . |
| $\Delta_{COR}$                                                 | The descriptive three-way factorial interaction among CLIC, OPTIC, and RELIC; evidence of complementarity or interference, but not by itself a certificate of creativity.                                                                                                                                                                                                                                                                                                  |
| $S_{RL}$                                                       | A declared reinforcement-learning problem, including state and action types, transition and reward interfaces, discount convention, and observer.                                                                                                                                                                                                                                                                                                                          |
| Target <sub>RELIC</sub>                                        | The RELIC target card $(Y_{dec}, S_{dec}, C_{dec})$ : deployable decision object, structural decision declaration, and epistemic certificate family.                                                                                                                                                                                                                                                                                                                       |
| Target <sub>X</sub><br>$(Y_X, S_X, C_X)$                       | = The three-level target of DIAL-X: deployable object $Y_X$ , structural declaration $S_X$ , and epistemic certificate family $C_X$ .                                                                                                                                                                                                                                                                                                                                      |
| Art <sub>X</sub><br>$(y, Y_X, Tr_X, c_X, m_X)$                 | = A versioned DIAL-X artifact dossier: operational construction, package comparison, transport record, status-bearing admission record produced from independent evidence, and persistent provenance or replay state. The sketch map $J_X$ , when present, is the presentation component of $Y_X$ . A rejected or unsupported proposal can retain a dossier without being admitted.                                                                                        |
| Cr <sub>X</sub><br>$(\beta_X, \mathcal{R}_X, \nu_X, \sigma_X)$ | = The creative mandate of DIAL-X when creativity is claimed: Boden mode, frozen reference space, novelty observer, and surprise or explanatory-gain measure. Value remains part of the independent epistemic certificate.                                                                                                                                                                                                                                                  |
| Creative delta                                                 | The named difference between LINC-S-X and Creative DIAL-X: the frozen language and reference repertoire, the visible obstruction, the artifact or rule that may be constructed, and the independent test of value. Combinational and exploratory deltas may remain within the language; transformational deltas change it.                                                                                                                                                 |
| Target card                                                    | The preregistered interface declaring a DIAL-X target, side semantics, mixed witness, observability gate, finite-realization test, creative mandate when applicable, proposal language, less-disruptive controls, package comparison, admission evidence, and persistence policy.                                                                                                                                                                                          |

| Term or symbol                                                        | Meaning in this book                                                                                                                                                                                                                                                                                                                                                                                 |
|-----------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Dualizability doctrine                                                | The declared conditions under which the internal side and core objects admit enough duality to reproduce, or explicitly replace, Mackenzie's finite-rank double-vector-bundle duality.                                                                                                                                                                                                               |
| $\mathbb{D}_{\text{rep}}$                                             | A double vector bundle organizing two interacting classes of infinitesimal repair over a common theory-state space.                                                                                                                                                                                                                                                                                  |
| $A_{\text{asm}}, B_{\text{acc}}$                                      | The interpreted side bundles of assimilatory and proto-accommodative repair directions. These meanings are supplied by the application.                                                                                                                                                                                                                                                              |
| $K$                                                                   | The core of $\mathbb{D}_{\text{rep}}$ , containing mixed directions invisible from either side bundle alone.                                                                                                                                                                                                                                                                                         |
| $\mathcal{W}_{\text{Lie}}^{\bullet,\bullet}(\mathbb{D}_{\text{rep}})$ | Meinrenken–Pike's bigraded cochain Weil algebra associated with a classical repair double vector bundle; not Leung's classifying theory $\text{Weil}_1^{\text{tan}}$ .                                                                                                                                                                                                                               |
| $d_{\text{asm}}, d_{\text{acc}}$                                      | Candidate horizontal and vertical differentials representing the two local repair processes.                                                                                                                                                                                                                                                                                                         |
| $\Omega_{\text{cre}}$                                                 | The classical cochain realization $[d_{\text{asm}}, d_{\text{acc}}]_{\text{gr}}$ of the interchange defect; its vanishing characterizes compatibility of the candidate classical double Lie algebroid structure.                                                                                                                                                                                     |
| $(C, c_0, c_1)$                                                       | A contextual curve object supplying the abstract time parameter for differential equations in a tangent category, with distinguished initial point $c_0$ and vector field $c_1$ . For algebroid transport, the book separately declares the required linear-completeness doctrine; Burke–MacAdam's earlier complete curve object packages a restricted version of that doctrine into its definition. |
| $V_{\text{asm}}, V_{\text{acc}}$                                      | Realized vector fields for assimilation and proto-accommodation on a common theory-state object, when the relevant anchors, sections, and integrability data have been supplied.                                                                                                                                                                                                                     |
| $\Phi_{\text{asm}}, \Phi_{\text{acc}}$                                | Partial or complete curve-object flows integrating the two realized DIAL directions.                                                                                                                                                                                                                                                                                                                 |
| $r_{\Sigma}^{\text{DIAL}}$                                            | The nine-component residual of the matched-pair conditions for the two 2-term representations up to homotopy induced by a chosen splitting $\Sigma$ . Its coordinates depend on the splitting, while its vanishing does not.                                                                                                                                                                         |
| $\phi_A : A \rightarrow A^+$                                          | The skill-bundle component of transport across a proposed theory change, covering the semantic base map $\bar{J} : M \rightarrow M^+$ . It is distinct from the partial flows denoted by $\Phi$ .                                                                                                                                                                                                    |
| $\chi_A$                                                              | The comparison between transporting an $A$ -probe and probing after a finite theory extension.                                                                                                                                                                                                                                                                                                       |



| Term or symbol                                   | Meaning in this book                                                                                                                                                                                                                                                                                |
|--------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| $\text{Md}_{\text{adm}}$                         | The category of admissible typed Markdown skill programs and interface-preserving revisions within one fixed language version. Grammar or interface changes require registered maps between versioned artifact categories.                                                                          |
| $\text{Md}_{\text{sec}}, \text{Md}_{\text{pol}}$ | The typed subcollections whose artifacts compile, respectively, to local skill sections and to policies over admissible typed actions.                                                                                                                                                              |
| $\text{compile}_E^{\text{sec}}$                  | The partial semantic compiler from local skill artifacts to sections in environment $E$ .                                                                                                                                                                                                           |
| $\text{compile}_E^{\text{pol}}$                  | The partial semantic compiler from workflow artifacts to policies in environment $E$ ; either compiler may reject an ill-typed program.                                                                                                                                                             |
| $A \rightarrow M$                                | A Lie algebroid whose base contains theory states and whose fibers contain locally available skill directions.                                                                                                                                                                                      |
| $s_m$                                            | The local skill section obtained as $\text{compile}_E^{\text{sec}}(m)$ in a declared environment.                                                                                                                                                                                                   |
| $\rho : A \rightarrow TM$                        | The anchor mapping a skill direction to its observable infinitesimal effect on a theory state.                                                                                                                                                                                                      |
| $\ker \rho$                                      | Latent skill directions with no immediate visible displacement, though they may affect traces or later compositions.                                                                                                                                                                                |
| $\mathbb{L}_{\text{vis}}$                        | The visual-language sketch in Chapter 18; its sorts and arrows present visual objects, attributes, relations, scenes, transformations, and their compositional constraints.                                                                                                                         |
| $\Phi_{s,t}^\ell$                                | The conditioned probability-flow transport from diffusion time $s$ to $t$ under visual-language state $\ell$ , when existence, uniqueness, and invertibility conditions hold.                                                                                                                       |
| $\text{Mod}(\mathbb{T}, \mathcal{C})$            | The category of models of theory $\mathbb{T}$ in a semantic category $\mathcal{C}$ , with the preservation doctrine understood.                                                                                                                                                                     |
| $\mathbb{S}$                                     | A sketch presenting a theory by generators, equations, and designated universal constructions.                                                                                                                                                                                                      |
| $\mathfrak{D}^+, \delta_{\mathfrak{D}}$          | A proposed preservation doctrine and the declared comparison with the old doctrine, including the compatibility needed to compare their model categories.                                                                                                                                           |
| $\mathfrak{T}_{\text{str}}^+$                    | The structural extension subrecord containing a proposed presentation, doctrine, registered model semantics, theory map, doctrine comparison, and restriction specification. It refines fields of the status-bearing dossier $\Xi$ ; it does not contain or control independent admission evidence. |
| $\iota^*$                                        | Restriction of expanded models along a theory or sketch map $\iota : \mathbb{S} \rightarrow \mathbb{S}^+$ , when the doctrine comparison makes that reduct functor well-defined.                                                                                                                    |
| $\mathcal{C}, \mathcal{E}, \mathcal{T}$          | Combinational, exploratory, and transformational creativity.                                                                                                                                                                                                                                        |

| Term or symbol         | Meaning in this book                                                                                                                                                                                                                       |
|------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| IPC                    | Infinitesimal Probes of Creativity; the controlled mathematical benchmark family.                                                                                                                                                          |
| AF                     | The AI-Feynman-inspired symbolic-law construction family.                                                                                                                                                                                  |
| DIAL-URL               | The DIAL specialization of Universal Reinforcement Learning.                                                                                                                                                                               |
| SGTE                   | Simulator-Grounded Theory Extension.                                                                                                                                                                                                       |
| CTTE                   | Corpus-to-Theory Extension: the controlled document-grounded ladder that progressively withholds supplied entity, context, query, mechanism-language, and domain structure.                                                                |
| TI, NI                 | Textual-inversion calibration and natural-image transport phases in the ARTISTIC evidence ladder.                                                                                                                                          |
| Admission              | Independent assignment of an epistemic status to a proposed extension using an appropriate, named certificate. Admission is relative to that contract; it is not automatically proof, external validity, or general scientific acceptance. |
| Experimental support   | Compatibility of an observed outcome with a bounded claim under stated controls. Support does not by itself establish the claim outside the registered environment.                                                                        |
| Formal proof           | A consequence derived from stated assumptions in a specified formal system. Proof establishes that formal consequence, not the empirical adequacy, novelty, or value of the surrounding theory.                                            |
| Conceptual space       | A typed language of objects, relations, laws, observers, and admissible constructions.                                                                                                                                                     |
| Persistent theory      | A conceptual structure that can be communicated, criticized, reused, and extended by other reasoners.                                                                                                                                      |
| Generative compression | Economy achieved by a compact theory whose objects and laws generate explanations, predictions, interventions, and new questions; not merely a short encoding of observed data.                                                            |
| Creative object        | The unit whose novelty is being assessed, such as an artifact, equation, experiment, search trajectory, or presented theory.                                                                                                               |
| Symbolic regression    | Recovery of a symbolic expression from numerical observations, usually balancing empirical fit and expression complexity.                                                                                                                  |
| Novelty                | Difference relative to a declared repertoire, description language, observer, or conceptual space.                                                                                                                                         |
| Surprise               | The unexpectedness of an event or description relative to a specified probabilistic or information-theoretic model.                                                                                                                        |
| Theory surgery         | An explicit change to a theory together with an account of transport, preservation, and loss.                                                                                                                                              |

| Term or symbol                                                                      | Meaning in this book                                                                                                                                                                                                                                                                                                                                                                  |
|-------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Proto-accommodation                                                                 | An infinitesimal probe of a possible representational reorganization inside the current double repair package; not yet an admitted theory change.                                                                                                                                                                                                                                     |
| Involution algebroid                                                                | A tangent-categorical presentation of Lie-algebroid structure, supplying the established first-order bridge between tangent categories and classical Lie algebroids.                                                                                                                                                                                                                  |
| Yang–Baxter-style flip law                                                          | Burke–MacAdam’s braid relation for the involution on an anchored bundle’s prolongation. It is the internal coherence from which Jacobi is recovered on sections, and is not identical to Drinfel’d’s classical $r$ -matrix Yang–Baxter equation.                                                                                                                                      |
| 2-term representation up to homotopy                                                | A representation of a Lie algebroid on a two-term complex, encoded by a boundary map, compatible connections, and curvature data satisfying coherent differential identities.                                                                                                                                                                                                         |
| Matched pair of 2-representations                                                   | The nine compatibility conditions which, after choosing a linear splitting, characterize a classical double Lie algebroid.                                                                                                                                                                                                                                                            |
| DIAL                                                                                | Double Involution ALgebroid: the proposed internal double-algebroid structure used in this book, consisting of two involution-algebroid directions, a dualizability doctrine, and an interchange comparison. Its relationship with Mackenzie’s classical double Lie algebroids and their matched-pair homotopy presentation remains a bridge requirement, not a theorem assumed here. |
| $\text{DIAL}^{(n)}$                                                                 | A conjectural $n$ -fold involution-algebroid object with $n$ typed repair directions, DIAL structure on each two-dimensional face, and coherent higher interchange comparisons. The case $n = 2$ is ordinary DIAL.                                                                                                                                                                    |
| $T^k(\text{DIAL}^{(n)})$                                                            | The proposed two-parameter hierarchy separating tangent depth $k$ , which probes higher-order behavior of fixed repair directions, from repair arity $n$ , which counts independently typed directions.                                                                                                                                                                               |
| $\mathbb{I}_{\text{Lie}}$                                                           | Proposed enriched limit sketch or classifying tangent-categorical theory for one involution-algebroid direction.                                                                                                                                                                                                                                                                      |
| $\mathbb{D}_{\text{inv}} = \mathbb{I}_{\text{Lie}} \otimes \mathbb{I}_{\text{Lie}}$ | Candidate tensor-product presentation of two internal involution-algebroid directions and their interchange. Identifying its smooth finite-rank models with Mackenzie double Lie algebroids is a research objective, not an assumed theorem.                                                                                                                                          |
| $\mathbb{D}_{\text{inv}}^{(n)} = \mathbb{I}_{\text{Lie}}^{\otimes n}$               | Candidate enriched tensor-sketch presentation of $n$ internal involution-algebroid directions. Its existence and classical realization are part of the higher-DIAL conjecture.                                                                                                                                                                                                        |

| Term or symbol           |           | Meaning in this book                                                                                                                                                                                                                                                                                                                                           |
|--------------------------|-----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Sketch tensor product    |           | A presentation whose models are iterated models: under the required limit and enrichment hypotheses, $\text{Mod}(\mathcal{A} \otimes \mathcal{B}, \mathcal{M}) \simeq \text{Mod}(\mathcal{A}, \text{Mod}(\mathcal{B}, \mathcal{M}))$ . Its symmetry presents order-independence but does not imply Lie integrability.                                          |
| Integration observer     | ob-       | A domain-specific finite-realization observer: a declared mathematical or empirical test asking whether an infinitesimal repair composes into a coherent finite path. Examples include monodromy criteria, approximate holonomy, multi-step rollout, recurrence, intervention closure, and continuation across covers; these examples are not interchangeable. |
| Four-level bridge        | DIAL      | The ordered distinction among presenting a tensor-product repair theory, diagnosing an interchange obstruction in one of its models, integrating a finite repair trajectory, and transporting models along a separately proposed domain-theory extension.                                                                                                      |
| Double broid             | Lie alge- | A double vector bundle with compatible horizontal and vertical VB-algebroid structures; equivalently here, one whose Weil-algebra differentials super-commute.                                                                                                                                                                                                 |
| Infinitesimal creativity | cre-      | The architecture that diagnoses representational inadequacy through interacting local repairs, tests their persistent finite realization, and records creative jumps as admitted finite package changes. A sketch extension is claimed only when the presentation or generative language changes.                                                              |
| Differential extension   | theory    | A finite sketch map, semantic transport, and coherent comparison maps relating its old and new Weil actions. These data specify a candidate; admitted differential accommodation additionally requires coherent finite realization, transport, and independent testing of new consequences.                                                                    |
| Markdown contract        | skill     | A versioned, typed program declaring a skill's inputs, outputs, invariants, probes, and admission rule.                                                                                                                                                                                                                                                        |
| Lawvere theory           |           | A categorical algebraic theory whose models are finite-product-preserving functors.                                                                                                                                                                                                                                                                            |
| Finite-limit theory      |           | A small finitely complete category whose models preserve finite limits.                                                                                                                                                                                                                                                                                        |
| PROP                     |           | A one-generated strict symmetric monoidal category; its algebras are symmetric monoidal functors.                                                                                                                                                                                                                                                              |
| Grothendieck topos       |           | A category equivalent to sheaves on a small site, equivalently a left-exact reflective localization of a presheaf category.                                                                                                                                                                                                                                    |

# Subject Index

- ablation
  - integrated creativity, 242
- abstention
  - interval-valued relation, 362
- abstraction
  - mathematical creativity, 26
- accommodation, 37, 56
  - artistic, 388
  - combinational creativity, 149
  - computational creativity, 126
  - differential form, 82
  - exploratory creativity, 156
  - Lie algebroidal form, 97
  - transformational creativity, 163
- active acquisition
  - AGENTIC, 244
  - visual, 362
- active experiment
  - symbolic theory, 262
- active teaching, 46
- admission, 182
  - locked split, 100
- admission contract
  - initial sketch, 50
- admission record, 182
- affordance, 45
- AGENTIC, 235
  - controlled demonstration, 244
  - corpus testbed, 307
  - mathematical testbed, 252
  - simulator testbed, 286
  - text-to-image, 353
- AI Feynman
  - categorical ladder, 262
- ALFWorld, 223
  - information contract, 224
- algebraic theory, 107
- ALLORA, 190
- anchor map, 93
- artifact dossier, 181
- ARTISTIC, 349
  - evaluation axes, 370
  - Firefly cricket challenge, 367
  - generator adapter, 369
  - interpretive boundary, 381
  - Movie Mode, 370
  - Nano Banana challenge, 368
  - natural images, 359
  - provider-conditioned utility, 368
  - staged evidence, 357
  - structural assurance, 365
- artistic creativity
  - theory extension, 385
- artistic model space, 387
- artistic technique
  - generator construction, 391
- ARTISTIC-T2I-o, 356
- assimilation, 37, 56
  - artistic, 388
  - combinational creativity, 149
  - computational creativity, 126
  - differential form, 82
  - exploratory creativity, 156
  - Lie algebroidal form, 97
- BACON, 132
- Beck–Chevalley condition, 115
- blackbody radiation, 281
- blind watchmaker, 137
- Boden modes
  - workflow composition, 239
- Bongard problems, 371
- BRIDGE, 197
- category, 25
- causal discovery
  - structural accommodation, 197
- causal learning
  - infinitesimal creativity, 191, 197
- cellular automaton, 135
  - theory induction, 264
- certificate
  - empirical, 288
  - formal, 288
  - simulator-relative, 288
  - workflow exchange, 237
- Chomsky, Noam, 128
- CLIC, 191, 197, 235
  - target card, 197
- closure failure
  - artistic, 389
- coaching, 42
- COCO, 359
- cognitive development, 56
  - algebraic theories, 118
- combinational creativity
  - DIAL interpretation, 151
- completeness
  - vector field, 79
- composition
  - sequential, parallel, choice, and feed-back, 238
- computational creativity, 123
- computational irreducibility, 135
- concept invention, 57
- conceptual space
  - exploration, 155
- connection
  - paintbrush transport, 396
- conservation
  - developmental psychology, 56
  - sketch acquisition, 47
- conservation law
  - theory repair, 259
- conservativity
  - artistic, 392
- core

- double vector bundle, 35, 76
- corpus-to-theory construction, 303
- counter-witness
  - pedagogical, 46
- Creative Adversarial Network, 385
- creative artifact
  - artifact dossier, 181
- creative delta, 179
- creative exploration
  - optimization landscape, 93
- creative interchange obstruction, 35, 57, 76
- creative mandate, 178
- creative object, 124
- creative workflows
  - composition, 235
- creativity
  - combinational, 147
  - differential geometry, 65
  - exploratory, 155, 387
  - in DIAL-X, 178
  - in reinforcement learning, 219
  - lexical, 31
  - limits of formalization, 412
  - open domains, 416
  - related traditions, 123
  - scope of DIAL, 407
  - theory construction, 53
  - transformational, 163
- curve object, 79
- declaration accommodation, 257
- demonstration, 42
- developmental tower of theories, 118
- DIAL, 18, 35, 69
  - algorithmic realization, 175
  - empirical status, 409
  - four-level implementation invariant, 177
  - four-level semantics, 113
  - higher, 87
  - initialization, 48
  - Kan transport, 115
  - limitations, 407
  - music, 413
  - open problems, 418
  - readiness boundary, 49
- DIAL-ALLORA, 190
- DIAL-CAN, 385
  - calibration sequence, 394
- DIAL-GIRL, 188, 222
- DIAL-SKFM, 191, 197
- DIAL-SkillOpt, 100, 191
  - experiments, 213
- DIAL-URL, 266
- DIAL-X, 175
- differential theory extension, 65, 166
- Diffusers, 354
- diffusion model
  - DIAL, 349
  - SDE deformation, 395
- DILATE, 385
  - natural-video routing, 400
  - paintbrush evidence, 397
- discovery skill
  - training, 100
- document-to-sketch compilation, 306
- domain theory, 44
- double involution
  - interpretation, 411
  - RELIC interpretation, 227
  - simulator-grounded discovery, 289
  - sketch semantics, 70
  - visual language, 351
- Double Involution ALgebroid, 18
- double involution algebroid, 69
- Double Involution Learning for Artistic Theory Extension, 385
- double Lie algebroid, 25, 35, 57, 65, 76
- double vector bundle, 35, 76
- Drescher, Gary L., 57
- dynamical system
  - in a tangent category, 79
- enriched sketch, 70
- equilibration
  - creative systems, 126
- evaluation
  - creative theory extension, 417
- evidential status
  - admitted, 182
  - constructed, 182
  - DIAL-X, 182
  - estimated, 182
  - learned, 182
  - supplied or registered, 182
- evolution
  - as creative process, 137
- experiment
  - sketch acquisition, 51
- experimental support, 183
- explanation
  - versus prediction, 128
- explanation-based generalization, 44
- explanation-based learning, 41
- failure localization
  - RELIC, 229
- false repair, 365
- fiber metric
  - paintbrush, 396
- field building, 53
- finite realization, 177
- finite symmetry worlds, 260
- finite-limit theory, 107
- flow
  - assimilation and accommodation, 79
- formal proof, 183
- formal status
  - DIAL, 67
- functor, 25
- genotype–phenotype map, 137
- geometry
  - Euclidean and Cartesian, 26
- GIRL, 188, 222
- Grothendieck topos, 107
- ground truth
  - artistic theory extension, 402
- groupoid
  - diffusion flow, 351
- image generator
  - provider interface, 365
- imitation learning, 41
- impossible language, 128
- Indus Valley Civilization, 294
- infinitesimal probe, 33
- initialization
  - theory, 41
- instruction, 42
- integrability
  - creativity evaluation, 133
  - exploratory creativity, 157
  - theory repair, 113
- integrated creativity episode, 240
- integration observer, 177
- interface discovery, 147
- intrinsic motivation
  - versus RELIC, 225
- intrinsic property
  - generated discovery, 259
- involution algebroid, 69

- internal witness, 209
- n-fold, 87
- skill optimization, 207
- Kan extension, 115
  - artistic theory, 392
  - left, 115
  - right, 115
- knowledge-acquisition bottleneck, 45
- language invention
  - visual, 352
- large language model
  - critique, 128
  - scientific discovery, 141
- latent factor
  - presentation construction, 260
- Lawvere theory, 30, 107
- Lea, 257
- LEAP system, 45
- learning apprenticeship, 41
- Lie algebroid, 93
  - artistic skill, 388
  - extension, 97
  - paintbrush, 395
- Lie bracket
  - CLIC and OPTIC, 209
  - quotient residual, 389
- lossy functor, 118
- Mackenzie theory, 70
- MAPseq, 291
- Markdown
  - skill program, 93
- mathematical creativity
  - construction hierarchy, 254
- meta-creativity, 235
- meta-sketch, 236
- microplastics, 295
- model
  - functorial, 30
  - of an algebraic theory, 107
- monoidal theory, 65
- music
  - DIAL testbed, 413
- natural transformation, 25
- neologism, 31
- novelty, 123
- object permanence, 56
- observability
  - visual contract, 365
- observer chart
  - paintbrush, 396
- obstruction
  - cross-workflow, 240
- operationality, 44
- OPTIC, 191, 207, 235
  - experiments, 213
  - target card, 207
- Oxford-IIIT Pet dataset, 359
- paintbrush
  - generative dynamics, 395
- paired observation, 359
- persistence
  - Weil profile, 390
- Piaget, Jean, 56, 82
  - developmental stages, 118
- Planck, Max, 281
- possibility
  - statistical versus structural, 128
- presheaf, 32
- probability-flow ODE, 350
- problem solving, 53
- Prometheus, 290
- PROP, 107
- proto-accommodation, 57
- radiative forcing, 295
- reinforcement learning
  - coalgebraic accommodation, 266
  - infinitesimal creativity, 219
- RELIC, 219, 235
  - active evidence acquisition, 244
  - combinational construction, 228
  - frozen schema, 224
  - schema accommodation, 226
  - stage construction, 229
  - target card, 220
- repair arity, 87
- Schema Mechanism, 57
- schema mechanism, 223
- scientific discovery, 123
  - autonomous agents, 141
  - BACON, 132
  - foundation models, 141
- score-based diffusion, 350
- semantic accommodation, 351
- SGTE, 297
- sheaf, 32
- simulator-grounded theory extension, 285
- singing mice, 291
- sketch, 25
  - acquisition, 41
  - implicit document presentation, 306
  - presentation of a theory, 107
  - tensor product, 70, 113, 151, 167
- sketch extension
  - artistic, 391
- SKFM, 197
- skill optimization, 93
  - infinitesimal creativity, 191, 207
- SkillOpt, 100
- status-bearing extension dossier, 62
- structural accommodation
  - reinforcement learning, 219
- structural obstruction
  - artistic, 385
- structural packet, 237
- style ambiguity, 387
- surprise, 123
- symbolic regression, 123
- synergy
  - creative workflows, 242
- tangent category, 25, 65
- tangent depth, 87
- tangent probe
  - interaction, 150
- target card, 184
- task theory, 45
- testbed
  - integrated creativity, 242
  - mathematics, 251
  - scientific corpus, 303
  - simulation, 285
- text-to-image generation, 349
- textual inversion, 352
  - ARTISTIC calibration, 354
  - empirical calibration, 358
- theory
  - causal, skill, and decision, 236
- theory change
  - DIAL workflow, 407
- theory extension
  - Lean calibration, 257
- lexicon, 31
- theory revision
  - dynamics, 79

- theory routing
  - reuse, specialize, extend, 400
- theory surgery, 163
- theory transport, 42
- theory-extension package, 62
- tokenizer, 31
- transformational creativity
  - DIAL interpretation, 167
- typed admission
  - artistic theory, 392
- universal coalgebra
- theory construction, 266
- VB-algebroid, 76
- VIC hydrology model, 294
- visual declaration
  - estimated, 359
- visual language
  - construction, 349
  - empirical admission, 357
- visual theory invention, 371
- visual-language sketch, 387
- von Neumann, John, 135
- Weil algebra, 25, 65
  - combinational creativity, 150
  - transformational creativity, 166
  - two uses, 69
- Weil probe
  - artistic obstruction, 390
- Weil profile
  - exploratory creativity, 157
- Wien's law, 281
- withheld-generator experiment, 402
- Wolfram, Stephen, 135
- workflow algebra, 238