Intelligent Systems

The Adaptive Closure

Emad Mostaque

Intelligent Internet

August 2026

Abstract

An adaptive system never starts from nothing. It inherits a way of weighting the possibilities it can represent. Evidence changes what is credible; value, fitness, or policy changes what is selected; action unfolds through time; observation hides some distinctions; generation adds possibilities that the current reference could not select. The central question is whether one mathematical form can follow all of these changes without confusing their roles.

Two derivational routes lead to the same reference-relative law. One begins from what an answer may contain when it must follow from stated grounds. The other begins from bounded systems choosing under inherited expectations. Both yield

$$\boxed{\mathcal{T}_f[\mu](dx) = \frac{e^{f(x)}\mu(dx)}{\int e^f \, d\mu}, \qquad f = V/\tau.}$$

The reference μ records what the system already carries. The potential f represents evidence, value, fitness, loss, or another declared influence. When $f = V/\tau$, τ is the price of moving away from the reference.

The same form survives the principal transformations of adaptation. Repeated closed updates add in log space. Independent systems remain independent, while dependence has an exact information cost relative to a product reference. Observation and coarse-graining preserve the form but change the effective potential, revealing how much adaptive information remains visible and how much is hidden. On trajectories, the same law accounts for both the destination and the path taken to reach it. With transition or generation, open steps compose as positive operators before final normalisation.

The closure has a sharp boundary. Finite reweighting can redistribute probability within the current measure class, but it cannot give standing to a state outside it. Mutation, search, invention, and representation change therefore require generation. Commitment creates a second compression: action may be precise while belief remains unresolved, provided a retained record preserves what later revision needs.

These results apply to a broad class of reference-relative adaptive systems. Intelligent systems are the subclass in which represented information about hidden or future consequences guides inference, search, commitment, experiment choice, or revision. Because any law equivalent to a reference can be written as a tilt, explanation begins only when the reference, potential, channel, generator, path law, and resolution are fixed independently of the observed outcome.

Reading guide

The paper has one main argument and several technical branches. The formal statements are part of the theory; most proofs can be skipped on a first reading.

Three reading paths. **Orientation:** read the abstract, Sections 1, 5, 7–10, 17, 25, 38–39, 47–50, and 54. **Main argument:** read every section marked C or F and skip blocks labelled *Technical proof*. **Full technical reading:** include sections marked T and all proofs.

Difficulty key. C Core idea or interpretation. F Formal result; equations matter, proof optional. T Advanced mathematics. A Application or empirical test.

Symbol	Name	Plain meaning
μ	reference	what the system already carries: prior weight, inherited expectation, or baseline support
P	epistemic state	what the stated evidence and constraints support
ρ	selected law	what value, fitness, policy, or another selection rule makes more likely
f	potential	the evidence, value, fitness, or loss signal used to reweight the reference
τ	information price	how costly it is to depart from the reference when $f = V/\tau$
K	observation channel	what is seen, reported, measured, or retained
M	transition or generator	how the represented possibilities move or expand
Γ	commitment channel	how an unresolved state becomes a decision, action, or artefact

A useful rule. On a first reading, follow the theorem statement and the paragraph after the proof. Technical pages are flagged in the margin. The proofs remain available for verification without interrupting the conceptual line.

Contents

Part I
The Problem of Adaptive Complexity

Metaphor and analogy can be helpful, or they can be misleading. All depends on whether the similarities the metaphor captures are significant or superficial.
Herbert A. Simon, *The Architecture of Complexity*, 1962

1 The Missing Law

Complexity science offers mechanisms for feedback, regulation, morphogenesis, reproduction, hierarchy, search, selection, self-organisation, networks, and emergence. These programmes solve distinct problems. A common criterion is still needed to decide when one adaptive description survives movement across mechanism, time, or scale.

The two derivations meet at one quantitative object: a reference law changed by an exponential potential. The question is whether that form survives time, interaction, stochastic observation, path evolution, and scale while preserving every distinction left unresolved by the transformation.

Structural unity requires more than common notation. Parameters chosen after observation can make any resemblance appear structural. Closure instead determines transformed parameters from the objects and operations supplied beforehand.

Definition 1.1 (Unification by closure)**.** A mathematical form is **unified by closure** over a declared class of transformations when:

(i) every transformation returns an object of the same typed form;
(ii) the transformed parameters follow from the original object and the declared transformation;
(iii) information discarded by compression appears as an exact remainder;
(iv) information or support added by openness appears through an explicit additional operation;
(v) an application transfers at least one restriction beyond the common equation.

A shared equation gives correspondence, and a shared variational law adds structure. Closure also requires a transported reference, a derived effective drive, an account of hidden information, and an explicit boundary for the closed law.

Definition 1.2 (Full answer)**.** A **full answer** is the entire result supported by a stated problem: a point, an equivalence class, a family, an empty class, or an unattained boundary. An unresolved family or boundary remains part of the answer unless the problem supplies a selector that resolves it.

2 A History of Adaptive Complexity

2.1 Organised complexity, signal, and feedback

Warren Weaver gave the modern problem its name. Between the few-variable problems of classical mechanics and the disorganised complexity treated by statistical averages lay systems in which many variables formed an organised whole (Weaver, 1948). The obstacle was not number alone. It was dependence.

Shannon supplied a calculus for signal and uncertainty (Shannon, 1948). Wiener supplied feedback and named cybernetics as the study of control and communication in animal and machine (Wiener, 1948). Ashby turned regulation into a problem of matching variety and, in *Design for a Brain*, treated adaptation as a higher-order change in the regulator itself: ultrastability restored essential variables by altering the organisation that controlled them (Ashby, 1952, 1956). Conant and Ashby later stated the regulator theorem: every effective regulator must embody a model of the system it regulates (Conant and Ashby, 1970).

These results already contain four elements of the later architecture: a represented range of states, a channel connecting system and environment, feedback that returns output as input, and an internal change that preserves viability. They do not yet fix one probability law for that change.

2.2 Pattern, reproduction, and self-production

Turing's morphogenesis paper showed how local reaction and diffusion can break symmetry and generate macroscopic form from an initially homogeneous field (Turing, 1952). The mechanism was developmental rather than Darwinian: one local dynamics generated organised pattern across different materials.

Von Neumann asked a different question: what organisation permits a machine to construct another machine of the same organisation? His theory of self-reproducing automata separated a description from the machinery that interprets and copies it, anticipating the distinction between a represented state and the operations that reproduce it (von Neumann, 1966). Waddington's epigenetic landscape placed development on a structured field of possible trajectories, where canalisation retained a phenotype against perturbation while genetic assimilation could alter the landscape itself (Waddington, 1957).

General system theory sought isomorphisms across organisms, machines, and societies (von Bertalanffy, 1968). Autopoiesis sharpened the living case by treating an organism as a network that continually produces the components and boundary constituting it as a unity (Maturana and Varela, 1980). These traditions made reproduction, development, and self-maintenance central. They also separated two operations that will remain distinct: changing weights inside a represented world and changing the represented possibilities themselves.

2.3 Simon: boundedness, hierarchy, and retained partial solutions

Herbert Simon joined the decision problem to the architectural one. Bounded rationality replaced the fiction of costless global optimisation with search under limited information and computation (Simon, 1955). Satisficing was not a defective version of omniscience; it was a different account of choice in a world whose full comparison could not be performed.

In *The Architecture of Complexity*, Simon argued that complex systems are commonly hierarchical and nearly decomposable: interaction is stronger within modules than between them (Simon, 1962). Stable intermediate forms make complex construction possible because failure at one level need not erase every partial success below it. Simon and Ando gave the corresponding dynamic programme: fast within-block relaxation and slower between-block motion can justify aggregation (Simon and Ando, 1961).

A module is memory embodied in organisation: it retains a partial answer while the larger system searches over arrangements. Any unified theory of adaptive systems should therefore say what a module retains, how much cross-module dependence costs, and when the macro-description closes.

2.4 Scale, stability, and self-organisation

Kadanoff and Wilson developed an exact calculus of scale change in statistical physics. Coarse-graining replaced microscopic degrees of freedom by effective variables, and renormalisation tracked how parameters changed under repeated reduction (Kadanoff, 1966; Wilson, 1971a,b). Anderson's conclusion, that more is different, did not deny the lower level. It held that an effective law at a higher level can contain organising principles absent from a list of microscopic parts (Anderson, 1972).

Nonequilibrium thermodynamics and synergetics showed how open systems can maintain order through flow. Prigogine and Nicolis studied dissipative structures; Haken studied order parameters and collective modes (Nicolis and Prigogine, 1977; Haken, 1977). Bak, Tang, and Wiesenfeld proposed self-organised criticality as one route to scale-free behaviour without external fine-tuning (Bak et al., 1987).

Ecology supplied the necessary warning. May showed that size and connectance can destabilise random interaction systems (May, 1972). Complexity does not purchase resilience by itself. Topology, interaction strength, modularity, and the distribution of alternatives determine whether a system remains viable.

2.5 Selection, information, and the computational turn

Darwin supplied the cumulative mechanism: inherited variation is differentially retained (Darwin, 1859). Fisher and Wright translated selection into changes in population frequencies and adaptive landscapes (Fisher, 1930; Wright, 1932). Eigen treated molecular evolution as selection on replicating information and identified an error threshold beyond which inheritance cannot preserve a type (Eigen, 1971). Evolutionary game theory later made frequency-dependent selection explicit through replicator dynamics (Taylor and Jonker, 1978; Hofbauer and Sigmund, 1998).

John Holland turned adaptation into a general computational programme. Genetic algorithms treated variation, selection, and recombination as a search process; classifier systems added rules and internal models; the later language of complex adaptive systems joined agents, tags, flows, building blocks, and adaptation (Holland, 1975, 1992, 1995). Artificial life extended the study of organised living behaviour to synthetic systems as well as evolved organisms (Langton, 1989).

The Santa Fe Institute gave that convergence an institutional home. Physicists, biologists, computer scientists, economists, and others treated evolution, computation, information, and nonlinear dynamics as parts of one research programme (Cowan, 1995). Kauffman studied rugged landscapes and autocatalytic order; Gell-Mann sought the regularities connecting fundamental law to biological and social complexity; Arthur recast the economy as an evolving, path-dependent process rather than a system exhausted by equilibrium (Kauffman, 1993; Gell-Mann, 1994; Arthur, 1999; Arthur et al., 1997).

Agent-based modelling and network science then made local-to-global construction operational. Schelling showed how local movement rules can generate aggregate segregation; Epstein and Axtell grew social structure from interacting agents; Watts, Strogatz, Barabási, and Albert supplied reusable descriptions of network topology (Schelling, 1971; Epstein and Axtell, 1996; Watts and Strogatz, 1998; Barabási and Albert, 1999).

2.6 What the history gives

The historical programmes classify different structures:

Programme	What it contributes	What remains free
Cybernetics	feedback, regulation, model-based control	the law of belief or selection change
Systems theory	cross-domain structural comparison	a restrictive quantitative operator
Morphogenesis and self-organisation	pattern formation and maintained order	inheritance, valuation, and support change
Simon	boundedness, hierarchy, near-decomposability	the exact information price and scale map
Evolution	variation, selection, inheritance	a common epistemic and social reading
Complex adaptive systems	adaptive agents, rules, flows, internal models	closure under time, channels, and aggregation
Renormalisation	effective description under scale change	the adaptive meaning of the effective parameters
Agent and network models	emergence from local interaction	a transferable law across implementations

A commuting law would make these comparisons exact. An adaptive form should remain recognisable when repeated, coupled, observed, and aggregated, with interaction cost, discarded information, macro-autonomy, and support expansion accounted for explicitly.

3 The Proof Tree

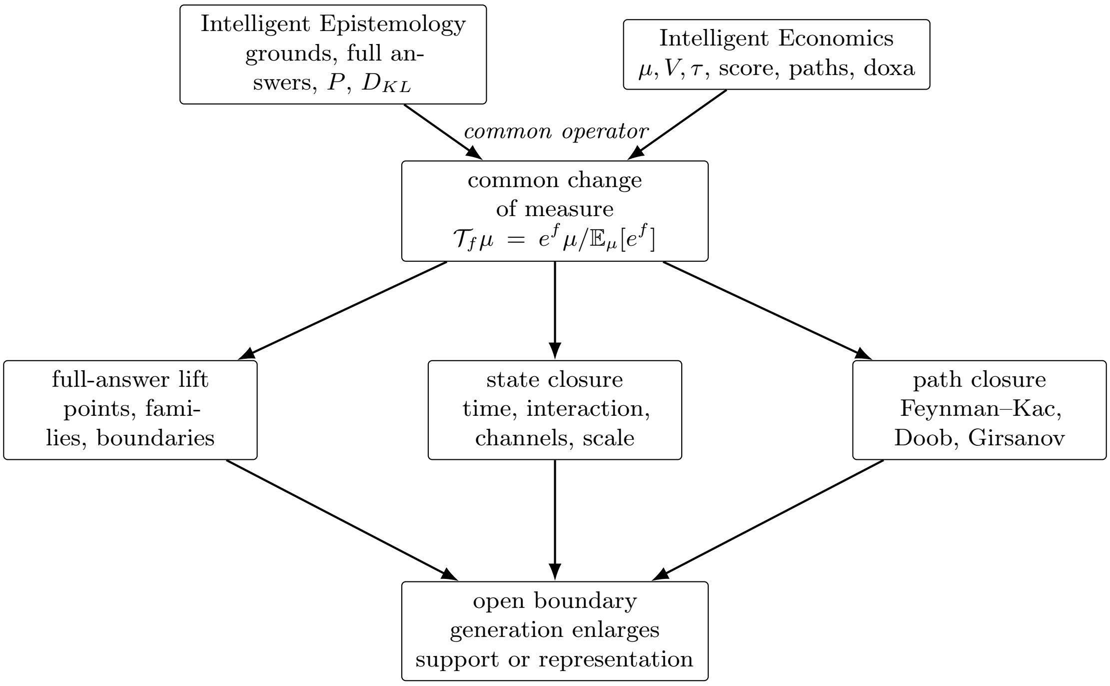

Figure 1: The proof tree. Two independent foundations converge on the reference tilt. The lower branches prove closure for full answers, state laws, and path laws, then locate the operation no closed branch can perform: giving standing to an unrepresented possibility.

Part II
Two Foundations, One Operator

More is different.

P. W. Anderson, 1972

4 Two Derivational Routes

The derivations begin from different premises. One constrains inference by dependence on grounds; the other constrains bounded valued choice under persistence. Their quantitative branches meet at the same change of measure.

4.1 The epistemic route

Inherited Result 4.1 (Reference-relative choice)**.** Within the finite and measurable domains stated in *Intelligent Epistemology*, dependence on grounds generates probability and exact informational accounting generates relative entropy. A permitted reference and constraints generate least-informative completion; a current state and new constraints generate retention by KL projection; and an additional value V with departure price $\tau > 0$ gives the unique choice law

$$\rho^* = \arg\max_{\rho \ll \mu} \left\{ \mathbb{E}_\rho[V] - \tau D_{\mathrm{KL}}(\rho \| \mu) \right\}, \qquad \frac{d\rho^*}{d\mu} = \frac{e^{V/\tau}}{\int e^{V/\tau}\, d\mu}.$$

Finite value preserves the measure class and therefore the support of the reference (Mostaque, 2026b).

The epistemic result fixes the geometry. A represented state begins from a reference. Completion or revision occurs relative to it, and KL measures the informational departure. Value tilts the reference while preserving its measure class.

4.2 The economic route

Inherited Result 4.2 (Bounded persistent choice)**.** Within the agent domain and consistency requirements stated in *Intelligent Economics*, bounded valued comparison against a reference forces the same exponential choice law, its score decomposition

$$\nabla \log \rho^* = \nabla \log \mu + \frac{1}{\tau} \nabla V,$$

its KL-regularised variational dual, and a canonical reversible relaxation with ρ^* as stationary law. Across adaptive rounds, realised choice can be retained as the next reference, and interacting agents couple through reference and value (Mostaque, 2026a).

The economic result gives the persistent interpretation. The reference is inherited structure; value is present direction; temperature prices information; and adaptation writes present choice into future conditions.

4.3 Convergence of the routes

One derivation fixes what an answer may contain; the other fixes how bounded valued choice departs from an inherited reference. Once value is supplied, both yield the same reference-relative operator.

5 Reference, Potential, and Tilt

5.1 The represented state

Let $(\mathcal{X}, \mathcal{F})$ be a standard Borel measurable space. The assumption ensures that the conditional laws used later admit regular versions. A probability measure μ on $\mathcal{X}$ is the system's *reference*: the state against which change is measured before the present evidence, value, or fitness acts. Two probability measures are *equivalent*, written $\rho \sim \mu$, when they have the same null sets; their equivalence class is the *measure class*.

The reference may be a population distribution, a base policy, an inherited institution, a physical ensemble, a current model state, or a macro-description of represented possibilities. The domain supplies its meaning; the mathematical role remains fixed.

Definition 5.1 (Admissible potential)**.** A measurable function $f : \mathcal{X} \to \mathbb{R}$ is **μ-admissible** when

$$0 < Z_\mu(f) := \int_{\mathcal{X}} e^{f(x)}\, \mu(dx) < \infty.$$

The function f is dimensionless. When a domain supplies a value V and a positive information price τ, one writes $f = V/\tau$. The term *potential* denotes the mathematical function; *drive* denotes its domain-specific interpretation.

Definition 5.2 (Exponential tilt)**.** For a probability measure μ and a μ-admissible potential f, define

$$\boxed{\mathcal{T}_f\mu(dx) = \frac{e^{f(x)}}{Z_\mu(f)}\, \mu(dx).}$$

The normalising term

$$\psi_\mu(f) = \log Z_\mu(f)$$

is the log-partition functional.

The pair (μ, f) separates inheritance from direction. Absorbing μ into f can reproduce one density in one coordinate system, but it erases the distinction between what the system brought and what the present signal changed. The closure results need that distinction.

5.2 The inherited variational identity

Theorem 5.1 (Reference–tilt identity)**.** *Let f be μ-admissible and let ρ be a probability measure with $\rho \ll \mu$, finite $D_{\mathrm{KL}}(\rho\|\mu)$, and finite $\mathbb{E}_\rho[|f|]$. Then*

$$\boxed{\mathbb{E}_\rho[f] - D_{\mathrm{KL}}(\rho\|\mu) = \psi_\mu(f) - D_{\mathrm{KL}}(\rho\|\mathcal{T}_f\mu)\,.}$$

Consequently $\mathcal{T}_f\mu$ is the unique maximiser of

$$\rho \longmapsto \mathbb{E}_\rho[f] - D_{\mathrm{KL}}(\rho\|\mu).$$

In value units, with $f = V/\tau$,

$$\mathcal{T}_{V/\tau}\mu = \arg\max_\rho \left\{\mathbb{E}_\rho[V] - \tau D_{\mathrm{KL}}(\rho\|\mu)\right\}.$$

optional proof *Proof.* The density of the tilted measure is

$$\log \frac{d\mathcal{T}_f\mu}{d\mu} = f - \psi_\mu(f).$$

Therefore

$$\log \frac{d\rho}{d\mathcal{T}_f\mu} = \log \frac{d\rho}{d\mu} - f + \psi_\mu(f).$$

Integrating with respect to ρ gives the identity. Relative entropy is nonnegative and vanishes only when the two measures agree almost everywhere. ☐

The identity is the Gibbs variational principle in reference-relative form and also appears as the one-step soft-control optimum and exponential-family completion (Kullback and Leibler, 1951; Jaynes, 1957; Csiszár, 1975; Shore and Johnson, 1980; Haarnoja et al., 2018). Retaining the reference as part of the represented object makes its transformation, rather than the isolated optimum, the object of study.

6 Established Forms and Their Roles

The exponential change of measure is classical. Statistical mechanics knows it as the Gibbs law; probability as a variational representation of log moments; statistics as the natural-parameter form of an exponential family; online learning as exponentiated gradient or mirror descent; stochastic control as a risk-sensitive or KL-regularised change of law; trajectory theory as Feynman–Kac weighting, Doob transformation, Schrödinger bridging, and driven processes (Gibbs, 1902; Donsker and Varadhan, 1975; Kivinen and Warmuth, 1997; Amari, 2016; Todorov, 2006; Del Moral, 2004; Doob, 1957; Léonard, 2014; Chetrite and Touchette, 2015).

The relevant component identities come from established traditions. Their roles differ according to the object being transformed.

Object	Established tradition	Adaptive role
$\mathcal{T}_f\mu$	Gibbs measures, exponential families, variational log-moments	common state-space operator
$\mathcal{T}_f(\mu M)$	Feynman–Kac and selection–mutation flows	open adaptive recursion
$dP^*/dQ \propto e^F$	path tilting, risk-sensitive control, Schrödinger problems	path-space closure and kinetic cost
D_{KL} **chain rules**	information theory and probability	exact coordination and scale remainders
conditional log-moment	effective potentials and marginalisation	derived drive under channels and scale

The closure problem therefore joins five requirements: independent derivation of the reference-relative form, preservation of full answer shapes, closure of state and path laws, exact information remainders under coupling and compression, and an explicit operation for support expansion.

7 Representation, Gauge, and Empirical Content

7.1 What the tilt can represent

Theorem 7.1 (Finite representation theorem)**.** *Let μ and ρ be probability measures on the same measurable space. There exists a finite μ-admissible potential f such that*

$$\rho = \mathcal{T}_f\mu$$

if and only if ρ and μ are equivalent measures. In that case every representing potential has the form

$$\boxed{f = \log \frac{d\rho}{d\mu} + c \quad \mu\text{-almost everywhere}}$$

for some constant c.

optional proof *Proof.* If $\rho = \mathcal{T}_f\mu$ with finite f, then $d\rho/d\mu = e^f/Z_\mu(f)$ is positive and finite μ-almost everywhere, so $\rho \sim \mu$. Conversely, if $\rho \sim \mu$, then $r = d\rho/d\mu$ is positive and finite μ-almost everywhere. Set $f = \log r$. Since $\int e^f d\mu = \int r d\mu = 1$, the potential is admissible and $\mathcal{T}_f\mu = \rho$. Additive constants give the same tilt, and equality of two tilted densities forces their potentials to differ by a constant almost everywhere. □

Remark 7.1 (Extended potentials and boundary laws). If only $\rho \ll \mu$ is required, the same representation holds with the extended potential $f = \log(d\rho/d\mu)$ and $f = -\infty$ on the part of the reference removed by ρ. Finite tilts preserve support; extended or limiting tilts may contract it. Those cases remain distinct.

The representation theorem also sets the empirical burden. Inside one measure class, every positive density change is a tilt. A post hoc potential can therefore redescribe almost any observed transition. Closure alone is a mathematical classification, not an empirical explanation.

7.2 Three gauges

Proposition 7.2 (Additive gauge). *For every constant $c \in \mathbb{R}$,*

$$\mathcal{T}_{f+c}\mu = \mathcal{T}_f\mu.$$

Thus potentials are identified only up to an additive constant.

optional proof *Proof.* Both numerator and normalising denominator are multiplied by e^c. □

Proposition 7.3 (Reference–potential gauge). *Let h be a finite μ-admissible potential and set $\mu^h = \mathcal{T}_h\mu$. Whenever the terms are admissible,*

$$\boxed{\mathcal{T}_{f-h}\mu^h = \mathcal{T}_f\mu.}$$

More generally, changing to any equivalent reference transfers the log-density ratio into the potential.

optional proof *Proof.* The unnormalised measure is

$$e^{f-h}\mu^h \propto e^{f-h}e^h\mu = e^f\mu.$$

Normalisation gives the result. □

Proposition 7.4 (Value–temperature gauge). *For every $a > 0$,*

$$\frac{aV}{a\tau} = \frac{V}{\tau}.$$

Hence state choice identifies the dimensionless drive $f = V/\tau$, not the absolute scale of V and τ separately.

optional proof *Proof.* The ratio is unchanged by common positive rescaling. □

The reference–potential gauge governs the interpretation of interaction. Inherited dependence can be written as reference coupling or moved into a present interaction potential. Those descriptions become causally distinct only when temporal order, intervention, institutional provenance, training history, or another independent ground fixes which structure was already present before the current drive acted.

7.3 Finite tilts preserve the represented world

Proposition 7.5 (Finite tilts preserve the measure class). *If f is finite μ-almost everywhere, then*

$$\mathcal{T}_f\mu \ll \mu \qquad \text{and} \qquad \mu \ll \mathcal{T}_f\mu.$$

Hence the two measures have the same null sets and, on a regular topological state space, the same support.

optional proof *Proof.* The Radon–Nikodym derivative $e^f/Z_\mu(f)$ is positive and finite μ-almost everywhere. □

A finite potential can reverse rankings, create extreme concentration, and change every nonzero probability while preserving the immediate measure class. The reference boundary sets which states can receive standing.

7.4 Where empirical content enters

The empirical condition. A tilt model is explanatory when the objects that make it restrictive are fixed independently of the selected law or constrained before testing. At least one of the following must occur:

(i) the reference is measured before the intervention or selected by declared symmetry, history, apparatus, or institution;

(ii) the potential is experimentally manipulated, supplied by a payoff or likelihood, or restricted to a low-dimensional family;

(iii) the information price is calibrated through independent choice variation or resource cost;

(iv) the channel, generator, or coarse-graining is known from the mechanism or measurement design;

(v) the transported object is predicted at another time, interaction strength, channel, or scale without refitting the hidden components.

Post hoc fitting represents the law; explanation requires prior inputs.

In plain language. If the potential is inferred from the outcome, the model has described that outcome. A prediction begins only when the reference and potential are fixed before the outcome is observed.

Non-Smuggling requires an output to depend only on distinctions supplied by its grounds.

8 Reference, Belief, and Realised Law

8.1 Three probability roles

The same probability measure can play different roles. The role, rather than the symbol alone, fixes its interpretation.

Definition 8.1 (Three probability roles)**.** (i) μ is the **reference**: inherited comparison structure before the current evidence, value, or selection step.

(ii) P is an **epistemic state**: a probability law or credal answer supported by stated grounds after evidence is taken into account.

(iii) ρ is a **selected or realised law**: the distribution produced when value, fitness, policy, or another declared selection rule acts on its immediate reference.

Evidence determines what is supported; value, fitness, or policy determines what is selected or done. Both may use the same mathematics, but they answer different questions.

In plain language. The same probability law can be a prior, a supported belief, or a selected outcome. Its role is fixed by the problem, not by the symbol used to write it.

8.2 Revision, assimilation, and memory

Three operations carry information between rounds. They act on different inputs and should not be collapsed under one name.

Operation	Definition	Role
Constraint revision	$P^{+} = \arg\min_{Q\in\mathcal{C}} D_{\mathrm{KL}}(Q\|P)$	incorporates new constraints while changing no more than required
Reference assimilation	$\mu_{t+1} = \rho_t$	makes the realised law of one round the inherited reference of the next
Partial memory	$\dfrac{dR_\alpha(\mu,\rho)}{d\mu} = \dfrac{(d\rho/d\mu)^\alpha}{\int (d\rho/d\mu)^\alpha d\mu}$	controls how much of the selected law is retained, $0 \le \alpha \le 1$

The three operations coincide only in special cases. For a reference and its tilt, partial memory remains in the tilt family.

In plain language. Revision changes a belief, assimilation changes the next reference, and partial memory controls how much of the change is retained.

9 The Adaptive Class

Definition 9.1 (Closed adaptive step)**.** A **closed adaptive step** on $\mathcal{X}$ is

$$\mu_t \longmapsto \rho_t = \mathcal{T}_{f_t}\mu_t,$$

followed by a declared retention or assimilation rule that produces the next reference without enlarging represented support.

Here *closed* refers to the represented world: the step reweights possibilities already carried by the reference. Thermodynamic isolation is a separate property.

Definition 9.2 (Reference-relative adaptive system)**.** A reference-relative adaptive system consists of:

(i) a represented state space $(\mathcal{X}, \mathcal{F})$;

(ii) a time-indexed reference law or, when the grounds remain unresolved, a family of permitted references;

(iii) a dimensionless potential f_t, or value and information price (V_t, τ_t);

(iv) a retention rule;

(v) declared interaction, observation, actuation, and generation channels where present;

(vi) an output object type: state law, path law, action, macrostate, parameter, or model;

(vii) an answer shape: point, family, equivalence class, empty case, or boundary.

The system is open to novelty when an operation enlarges support or the represented state space. A generation channel is any declared operation with that role.

Definition 9.3 (Intelligent adaptive system)**.** An adaptive system is **intelligent** when represented information about hidden or future consequences contributes to its inference, potential, generation policy, commitment, experiment choice, or revision, and that contribution is answerable to prediction, regulation, or persistence.

The closure results apply to the broader adaptive class. A replicating population may satisfy the same probability law without possessing one central predictive model.

In plain language. Adaptation is the broad mathematical class. Intelligence is the narrower case in which represented information helps the system infer, search, commit, or revise.

9.1 The adaptive loop

A static tilt becomes adaptive when its output changes the next problem. Let Π_t contain the current language, candidate space, reference, constraints, channels, and output question. A typed cycle is

$$\Pi_t \longrightarrow \bar{\Pi}_t \longrightarrow P_t \longrightarrow \rho_t \longrightarrow \text{action and observation} \longrightarrow \Pi_{t+1}.$$

The first arrow is generation when the candidate space must be enlarged and the identity otherwise. Evidence determines the epistemic answer P_t. A declared selection rule may then produce a state or path law ρ_t. Action changes the world; observation, retention, support expansion, or representation change determines the next problem.

Remark 9.1 (Why agents are not primitive). An agent is one carrier of adaptation, but not the only one. Centralised decision, distributed selection, population replacement, Bayesian revision, and institutional reproduction can share a law of probability change while differing in mechanism. The common object is the typed law and the conditions under which it transfers, not psychological similarity among implementations.

10 Closure as a Commuting Problem

Let C be an operation such as taking a product, passing through a channel, or coarse-graining. Closure asks whether there is an effective potential f_C such that

$$C(\mathcal{T}_f\mu) = \mathcal{T}_{f_C}(C\mu).$$

For temporal composition, C is another closed tilt and the potentials add. For a channel, f_C is a conditional log-moment. For a product, additive potentials factor and interaction potentials measure departure from factorisation.

$$
\begin{tikzcd}
\mu \arrow[r, "\mathcal{T}_f"] \arrow[d, "C"'] \arrow[dr, phantom, "\substack{\textit{closure means}\\ \textit{the square commutes}}"] & \mathcal{T}_f\mu \arrow[d, "C"] \\
C\mu \arrow[r, dashed, "\mathcal{T}_{f_C}"'] & C(\mathcal{T}_f\mu)
\end{tikzcd}
$$

Figure 2: The closure problem. The transformed potential f_C is derived for temporal composition, products, stochastic channels, and deterministic coarse-graining.

Part III
The Algebra of Adaptation

On those who step into the same rivers, different and again different waters flow.

Heraclitus, fragment B12

11 Tilts Form an Additive Action

Theorem 11.1 (Tilt algebra). *Let μ be a probability measure and let f, g be drives for which the displayed tilts are admissible. Then*

$$\boxed{\mathcal{T}_g[\mathcal{T}_f[\mu]] = \mathcal{T}_{f+g}[\mu].}$$

Moreover $\mathcal{T}_0$ is the identity, adding a constant to a drive changes nothing, and whenever f is finite μ-almost everywhere,

$$\mathcal{T}_{-f}[\mathcal{T}_f[\mu]] = \mu.$$

Thus admissible drives modulo additive constants act abelianly on each measure class.

optional proof *Proof.* Write $Z_f = \int e^f \, d\mu$. Then

$$\mathcal{T}_g[\mathcal{T}_f\mu](dx) = \frac{e^{g(x)}e^{f(x)}\mu(dx)/Z_f}{\int e^g \, d(\mathcal{T}_f\mu)} = \frac{e^{f(x)+g(x)}\mu(dx)}{\int e^{f+g} \, d\mu}.$$

The identity and gauge claims are immediate. For the inverse,

$$\int e^{-f} \, d(\mathcal{T}_f\mu) = \frac{1}{Z_f},$$

so the second tilt restores μ. □

Repeated adaptation therefore remains in the class and accumulates in the drive.

Corollary 11.2 (History compression). *If*

$$\mu_{t+1} = \mathcal{T}_{f_t}[\mu_t], \qquad t = 0, \ldots, T-1,$$

then

$$\boxed{\mu_T = \mathcal{T}_{F_T}[\mu_0], \qquad F_T = \sum_{t=0}^{T-1} f_t.}$$

Equivalently,

$$\frac{d\mu_T}{d\mu_0}(x) = \frac{\exp\left(\sum_{t<T} f_t(x)\right)}{\int \exp\left(\sum_{t<T} f_t\right) d\mu_0}.$$

Technical proof — optional on a first reading. Induct on T using the tilt algebra. □

History is multiplicative in measure space and additive in log space. The present state records the cumulative potential of every retained round, except for distinctions erased by later coarse-graining or support contraction.

In plain language. Repeated closed updates do not require a new rule at every round. Their log-potentials add, so the entire retained history can be represented by one cumulative potential.

11.1 Bayes, selection, and multiplicative weights

If $f_t(x) = \log \ell_t(x)$, then

$$\mu_T(dx) \propto \mu_0(dx) \prod_{t<T} \ell_t(x),$$

which is sequential Bayesian updating when ℓ_t is a likelihood factor. If $f_t = \eta V_t$, it is multiplicative weights. If f_t is fitness accumulated across generations, it is discrete selection. The algebra is the same because each operation is reference-relative exponential reweighting.

12 The Geometry of a Tilt Path

FIRST READING The theorem statement and the paragraph after its proof carry the main idea. The differential details may be skipped.

Fix a reference μ and a potential f with finite exponential moments on an interval containing $[0, 1]$. Define

$$\mu_s = \mathcal{T}_{sf}[\mu], \qquad \psi(s) = \log \int e^{sf}\, d\mu.$$

Theorem 12.1 (Tilt-path identities)**.** *For this path,*

$$\psi'(s) = \mathbb{E}_{\mu_s}[f], \qquad \psi''(s) = \mathrm{Var}_{\mu_s}(f) \geq 0,$$

and, for every observable h with locally finite tilted absolute moments for h and hf,

$$\boxed{\frac{d}{ds}\mathbb{E}_{\mu_s}[h] = \mathrm{Cov}_{\mu_s}(h, f).}$$

In particular,

$$\frac{d}{ds}\mathbb{E}_{\mu_s}[f] = \mathrm{Var}_{\mu_s}(f) \geq 0.$$

Technical proof — optional on a first reading. Differentiate numerator and denominator of the tilted expectation. The first two identities are the corresponding derivatives of the log-partition function. □

A tilt changes the expectation of h according to its covariance with the potential. The expected potential rises at its variance rate. When $f = V/\tau$, expected value rises at $\mathrm{Var}_{\mu_s}(V)/\tau$. With a fixed potential, improvement slows as the selected law loses variance in f.

Corollary 12.2 (Information along the tilt path)**.** *For $0 \leq r < s$,*

$$D_{\mathrm{KL}}(\mu_s \| \mu_r) = (s - r)\mathbb{E}_{\mu_s}[f] - \psi(s) + \psi(r).$$

The infinitesimal information metric is

$$D_{\mathrm{KL}}(\mu_{s+ds} \| \mu_s) = \frac{1}{2}\,\mathrm{Var}_{\mu_s}(f)\, ds^2 + o(ds^2).$$

Technical proof — optional on a first reading. Use $d\mu_s/d\mu_r = \exp((s - r)f - \psi(s) + \psi(r))$ and integrate its logarithm under μ_s. The infinitesimal form follows from a second-order Taylor expansion. □

13 Retention Rates

Full replacement, $\mu_{t+1} = \rho_t$, is one retention rule. Many adaptive systems retain only part of a realised state. The interpolation compatible with the multiplicative tilt is geometric.

For equivalent measures μ and ρ, the partial-memory operator has an equivalent density form. If λ dominates both, with densities $m = d\mu/d\lambda$ and $r = d\rho/d\lambda$, then

$$\frac{dR_\alpha(\mu,\rho)}{d\lambda} = \frac{m^{1-\alpha}r^\alpha}{\int m^{1-\alpha}r^\alpha\,d\lambda}, \qquad 0 \le \alpha \le 1.$$

The normalised measure is independent of the dominating measure λ.

Theorem 13.1 (Partial retention remains a tilt). *If $\rho = \mathcal{T}_f[\mu]$, then*

$$\boxed{R_\alpha(\mu,\rho) = \mathcal{T}_{\alpha f}[\mu].}$$

Consequently a retention rate rescales the drive and introduces no new mathematical form.

optional proof *Proof.* Up to normalisation,

$$\mu^{1-\alpha}\rho^\alpha \propto \mu^{1-\alpha}(e^f\mu)^\alpha = \mu e^{\alpha f}.$$

□

Corollary 13.2 (Weighted history). *For updates*

$$\mu_{t+1} = R_{\alpha_t}(\mu_t, \mathcal{T}_{f_t}[\mu_t])\,,$$

one has

$$\mu_T = \mathcal{T}_{\sum_{t<T}\alpha_t f_t}[\mu_0].$$

optional proof *Proof.* Theorem 13.1 gives $\mu_{t+1} = \mathcal{T}_{\alpha_t f_t}\mu_t$ at each round. Apply Corollary 11.2. □

A doxic adaptation rate, evolutionary selection coefficient, learning rate, or institutional memory parameter can occupy the same mathematical slot. The interpretation differs; the algebra does not.

Remark 13.1 (Arithmetic and geometric retention). An arithmetic mixture $(1-\alpha)\mu+\alpha\rho$ generally does not equal a tilt of μ by αf. It belongs to a mixture family rather than the exponential path. The choice is substantive: geometric retention preserves additive accumulation in log space; arithmetic retention preserves convex plurality in probability space.

14 The Continuous-Time Limit

Let V_t be a bounded value and let $\eta > 0$. Over a short interval dt, take the drive to be $f_t = \eta V_t dt$.

Theorem 14.1 (Replicator limit). *If*

$$\mu_{t+dt} = \mathcal{T}_{\eta V_t dt}[\mu_t],$$

then, in the weak sense,

$$\boxed{\partial_t \mu_t(dx) = \eta\big(V_t(x) - \bar{V}_t\big)\mu_t(dx), \qquad \bar{V}_t = \mathbb{E}_{\mu_t}[V_t].}$$

On a finite state space this is the replicator equation.

optional proof *Proof.* For any bounded test function h,

$$\mathbb{E}_{\mu_{t+dt}}[h] = \frac{\mathbb{E}_{\mu_t}[he^{\eta V_t dt}]}{\mathbb{E}_{\mu_t}[e^{\eta V_t dt}]}.$$

Expanding numerator and denominator to first order gives

$$\mathbb{E}_{\mu_{t+dt}}[h] = \mathbb{E}_{\mu_t}[h] + \eta \operatorname{Cov}_{\mu_t}(h, V_t)dt + o(dt),$$

which is the weak form of the displayed equation. ☐

Corollary 14.2 (Fundamental selection identity)**.** *If V is time-independent, then*

$$\boxed{\frac{d}{dt}\mathbb{E}_{\mu_t}[V] = \eta \operatorname{Var}_{\mu_t}(V) \geq 0.}$$

Equality holds exactly when V is constant on the support of μ_t.

optional proof *Proof.* Apply the covariance identity in Theorem 14.1 with $h = V$. Variance vanishes exactly when V is constant on the current support. ☐

The equation commonly introduced as an evolutionary law is the infinitesimal form of retained exponential tilting. Evolutionary selection, Bayesian evidence accumulation, and multiplicative learning are different readings of the same update (Taylor and Jonker, 1978).

14.1 Time-dependent environments

When V_t changes with time,

$$\frac{d}{dt}\mathbb{E}_{\mu_t}[V_t] = \eta \operatorname{Var}_{\mu_t}(V_t) + \mathbb{E}_{\mu_t}[\partial_t V_t].$$

Selection raises value relative to the current landscape; environmental change can raise or lower the landscape itself.

15 Concentration, Equilibrium, and Memory

Theorem 15.1 (Repeated fixed drive)**.** *Let f be bounded above and let*

$$\mu_n = \mathcal{T}_{nf}[\mu_0].$$

For every $\varepsilon > 0$, if

$$A_\varepsilon = \{x : f(x) \geq \operatorname*{ess\,sup}_{\mu_0} f - \varepsilon\},$$

then $\mu_n(A_\varepsilon) \to 1$. If the essential maximiser set has positive μ_0-mass, then μ_n converges to μ_0 conditioned on that set.

optional proof *Proof.* Subtract the essential supremum from f. Outside A_ε, $e^{nf} \le e^{-n\varepsilon}$. Inside a positive-mass near-maximiser set the exponential decays more slowly. The ratio vanishes. On the exact maximiser set all points receive the same factor, preserving conditional reference weights. □

Theorem 15.2 (Interior fixed-point condition)**.** *Let the drive depend on the current measure and suppose*

$$\mu^* = \mathcal{T}_{f(\mu^*)}[\mu^*].$$

Then

$$\boxed{f(\mu^*, x) = c \quad \mu^*\text{-almost everywhere}}$$

for some constant c. Conversely, if the drive is constant on the support of μ^*, then μ^* is a fixed point.

optional proof *Proof.* On the support of μ^*,

$$1 = \frac{d\mathcal{T}_{f(\mu^*)}[\mu^*]}{d\mu^*} = \frac{e^{f(\mu^*,x)}}{Z},$$

so $f = \log Z$ almost everywhere. The converse is immediate. □

An interior adaptive equilibrium equalises effective drive over every surviving state. The condition recurs as equal fitness on evolutionary support, indifference on mixed-strategy support, and equal adjusted return across active alternatives.

15.1 Path dependence

History compression does not eliminate history. It identifies where history lives: in the cumulative drive and inherited reference. Two systems facing the same present f_t can choose differently because their μ_t differ. Arthur's lock-in, institutional inertia, and learned priors are instances of that retained asymmetry. The present law is Markovian in the enriched state μ_t and path-dependent in any description that omits it.

Time-closure result. Finite histories compress into a cumulative drive; geometric partial memory stays in the same family; the continuous limit is replicator dynamics; fixed value rises at its variance rate; interior fixed points equalise drive on support; sustained exposure concentrates on supported maximisers. Each statement follows from the tilt algebra.

Part IV
Closure Through Interaction

The central theme that runs through my remarks is that complexity frequently takes the form of hierarchy.

Herbert A. Simon, *The Architecture of Complexity*, 1962

16 Independent Systems

Let $(\mathcal{X}_i, \mathcal{F}_i, \mu_i)$ be two represented systems. Their independent reference is $\mu_1 \otimes \mu_2$. For potentials f_i on the two spaces, write

$$(f_1 \oplus f_2)(x_1, x_2) = f_1(x_1) + f_2(x_2).$$

Theorem 16.1 (Independent composition)**.** *If f_1 and f_2 are admissible for μ_1 and μ_2, then*

$$\boxed{\mathcal{T}_{f_1 \oplus f_2}(\mu_1 \otimes \mu_2) = \mathcal{T}_{f_1}\mu_1 \otimes \mathcal{T}_{f_2}\mu_2.}$$

optional proof *Proof.* The unnormalised joint measure factorises:

$$e^{f_1(x_1)+f_2(x_2)}\mu_1(dx_1)\mu_2(dx_2) = \big(e^{f_1(x_1)}\mu_1(dx_1)\big)\big(e^{f_2(x_2)}\mu_2(dx_2)\big).$$

Fubini's theorem gives

$$Z_{\mu_1 \otimes \mu_2}(f_1 \oplus f_2) = Z_{\mu_1}(f_1) Z_{\mu_2}(f_2),$$

so normalisation preserves the product. □

Two independent references subject to additive potentials remain independent after adaptation. Dependence in the selected law must therefore be paid for by dependence already present in the reference or by an interaction term in the potential.

17 Interaction as Information

17.1 Two sources of coupling

A joint adaptive system on $\mathcal{X} \times \mathcal{Y}$ may fail to factor for two distinct reasons:

(i) the reference μ_{XY} already carries dependence;

(ii) the potential includes an interaction $w(x, y)$ that is not additively separable.

Reference dependence is inherited coordination. Potential dependence is newly purchased coordination.

Suppose first that the reference factorises and write

$$f(x, y) = f_X(x) + f_Y(y) + w(x, y).$$

When $w = 0$, Theorem 16.1 gives a product result. The interaction potential is the entire departure from independent adaptation.

17.2 The exact cost of coordination

Theorem 17.1 (Coordination decomposition)**.** *Let ρ be a probability measure on $\mathcal{X} \times \mathcal{Y}$ with marginals ρ_X, ρ_Y, and let the reference factorise as $\mu_X \otimes \mu_Y$. Whenever the divergences are finite,*

$$\boxed{D_{\mathrm{KL}}(\rho \| \mu_X \otimes \mu_Y) = D_{\mathrm{KL}}(\rho_X \| \mu_X) + D_{\mathrm{KL}}(\rho_Y \| \mu_Y) + I_\rho(X; Y),}$$

where

$$I_\rho(X; Y) = D_{\mathrm{KL}}(\rho \| \rho_X \otimes \rho_Y)$$

is mutual information.

optional proof *Proof.* Insert the product of the marginals into the log-density ratio:

$$\log \frac{d\rho}{d(\mu_X \otimes \mu_Y)} = \log \frac{d\rho}{d(\rho_X \otimes \rho_Y)} + \log \frac{d\rho_X}{d\mu_X} + \log \frac{d\rho_Y}{d\mu_Y}.$$

Integration with respect to ρ gives the result. □

Relative to independent references, a joint state pays three bills: changing the first marginal, changing the second marginal, and coordinating them. The irreducible information price of coordination is mutual information.

Theorem 17.2 (Coordination must earn its price)**.** *Let*

$$\rho^* = \arg\max_\rho \left\{ \mathbb{E}_\rho[V] - \tau D_{\mathrm{KL}}(\rho \| \mu_X \otimes \mu_Y) \right\},$$

and let ρ^*_X, ρ^*_Y be its marginals. Whenever the displayed terms are finite,

$$\boxed{\mathbb{E}_{\rho^*}[V] - \mathbb{E}_{\rho^*_X \otimes \rho^*_Y}[V] \geq \tau I_{\rho^*}(X; Y).}$$

The value gained by selecting dependence rather than the product of the selected marginals is at least its mutual-information price.

optional proof *Proof.* The product $q = \rho^*_X \otimes \rho^*_Y$ is an admissible competitor with the same marginals. Optimality gives

$$\mathbb{E}_{\rho^*}[V] - \tau D_{\mathrm{KL}}(\rho^* \| \mu_X \otimes \mu_Y) \geq \mathbb{E}_q[V] - \tau D_{\mathrm{KL}}(q \| \mu_X \otimes \mu_Y).$$

By Theorem 17.1, the first divergence equals the two marginal divergences plus $I_{\rho^*}(X; Y)$, while the divergence of q contains only the same two marginal terms. Cancelling them gives the result. □

The inequality supplies a direct test: at the optimum, dependence must earn joint value unavailable from the product of the same marginals.

In organisations, ecosystems, teams, and multi-agent systems, coordination is not free. Shared state, communication, synchronisation, and contractual alignment all instantiate dependence. The theorem identifies their common informational component without claiming that it exhausts physical or institutional costs.

18 Near Decomposability as a Bound

Simon described systems in which internal interactions dominate cross-component interactions. The tilt form permits an exact perturbative statement.

Let

$$\rho_0 = \mathcal{T}_{f_X}\mu_X \otimes \mathcal{T}_{f_Y}\mu_Y$$

be the independently adapted law, let w be a bounded interaction potential, and define

$$\rho_\varepsilon = \mathcal{T}_{\varepsilon w}\rho_0 = \mathcal{T}_{f_X \oplus f_Y + \varepsilon w}(\mu_X \otimes \mu_Y).$$

Theorem 18.1 (Weak-coupling bound)**.** *Let $R = \operatorname{osc}(w) = \operatorname{ess\,sup} w - \operatorname{ess\,inf} w < \infty$. Then*

$$\boxed{D_{\mathrm{KL}}(\rho_\varepsilon \| \rho_0) \le \frac{\varepsilon^2 R^2}{8},}$$

$$\boxed{d_{\mathrm{TV}}(\rho_\varepsilon, \rho_0) \le \frac{|\varepsilon| R}{4},}$$

and

$$\boxed{I_{\rho_\varepsilon}(X;Y) \le \frac{\varepsilon^2 R^2}{8}.}$$

Thus weak bounded interaction produces second-order information coupling and first-order total-variation departure from the decomposed law.

optional proof *Proof.* Let

$$\psi(\varepsilon) = \log \mathbb{E}_{\rho_0}[e^{\varepsilon w}].$$

Then

$$D_{\mathrm{KL}}(\rho_\varepsilon \| \rho_0) = \varepsilon\psi'(\varepsilon) - \psi(\varepsilon).$$

Differentiating gives

$$\frac{d}{d\varepsilon}\big(\varepsilon\psi'(\varepsilon) - \psi(\varepsilon)\big) = \varepsilon\psi''(\varepsilon) = \varepsilon \operatorname{Var}_{\rho_\varepsilon}(w).$$

A random variable contained in an interval of width R has variance at most $R^2/4$. Integrating from 0 to ε gives the KL bound for either sign of ε. Pinsker's inequality gives the total-variation bound. Finally, Theorem 17.1 applied with product reference ρ_0 shows that mutual information is no greater than the total divergence. ☐

Information-theoretic near decomposability. Near decomposability has an information-theoretic component: a bounded cross-module interaction potential produces a controlled departure from the product law and only second-order mutual information. The theorem therefore bounds the static price of ignoring cross-module dependence. Simon–Ando near decomposability additionally concerns fast within-module and slow between-module dynamics; that spectral timescale statement requires a domain generator and is not implied by this bound alone.

Near decomposability depends on effective interaction amplitude, not sparsity alone: a few strong links can dominate the system.

19 Effective Subsystem Drives

Interaction does not take a subsystem outside the class. It changes the potential that the subsystem experiences after the other variables have been integrated out.

Theorem 19.1 (Effective marginal drive). *Let the reference factorise as $\mu_X \otimes \mu_Y$ and let*

$$f(x,y) = f_X(x) + f_Y(y) + w(x,y)$$

be admissible. If

$$\rho = \mathcal{T}_f(\mu_X \otimes \mu_Y),$$

then the first marginal has the exact form

$$\boxed{\rho_X = \mathcal{T}_{f_X+\Phi_X}\mu_X, \qquad \Phi_X(x) = \log \int e^{f_Y(y)+w(x,y)}\,\mu_Y(dy).}$$

The corresponding statement holds for Y.

optional proof *Proof.* Integrating the joint density over y gives

$$\rho_X(dx) \propto \mu_X(dx)e^{f_X(x)} \int e^{f_Y(y)+w(x,y)}\mu_Y(dy),$$

which is the stated tilt after normalisation. □

The interaction enters a subsystem as a conditional log-partition. A component does not in general respond to the arithmetic mean of its environment. It responds to the logarithm of the environment's exponentially weighted possibilities. The same conditional-log-moment transformation governs coarse-graining.

20 Reference Coupling and Value Coupling

Definition 20.1 (Two couplings). A joint adaptive system carries:

(i) **reference coupling** when its inherited measure μ_{XY} is not a product;

(ii) **value coupling** when its present potential is not additively separable.

Shared history, common training data, institutions, kinship, and correlated environments enter through reference coupling. Games, complementarities, externalities, predation, and active coordination enter through value coupling.

Proposition 20.1 (Reference interaction relative to a product baseline). *Let $\mu_{XY} \sim \mu_X \otimes \mu_Y$ and define*

$$w_\mu(x,y) = \log \frac{d\mu_{XY}}{d(\mu_X \otimes \mu_Y)}(x,y).$$

Then for every admissible f,

$$\boxed{\mathcal{T}_f\mu_{XY} = \mathcal{T}_{f+w_\mu}(\mu_X \otimes \mu_Y).}$$

Technical proof — optional on a first reading. Substitute $d\mu_{XY} = e^{w_\mu} d(\mu_X \otimes \mu_Y)$ into the definition and renormalise. □

The proposition changes the baseline while preserving history. Relative to an independent reference, inherited dependence appears as an interaction potential. Relative to the inherited joint reference, that coordination is already carried by the baseline. The application must decide which interpretation is causal.

In plain language. One observed joint law cannot reveal whether coordination was inherited or created by the current interaction. That distinction requires temporal or experimental information.

Remark 20.1 (Extended and singular reference interaction)*.* With only absolute continuity, the density may vanish and $w_\mu = \log(d\mu_{XY}/d(\mu_X \otimes \mu_Y))$ may equal $-\infty$; a finite interaction potential therefore requires equivalence. If the joint law is singular, no density relative to the product baseline exists. Hard coordination belongs in the reference support or in an extended potential, outside the finite weak-coupling analysis.

The causal story fixes the coupling gauge. One observed joint law does not identify whether dependence was inherited in μ_{XY} or created by the current interaction. The distinction becomes empirical when one observes the pre-intervention reference, manipulates current payoffs while holding history fixed, changes training or institutional inheritance while holding the current objective fixed, or otherwise supplies temporal and causal structure. Without such grounds, only the combined joint potential is identified.

21 Many-System Coordination

For N components with factorised reference $\bigotimes_i \mu_i$, an additive potential produces independent adaptation. Pairwise or higher-order interactions produce a Gibbs field:

$$\rho(dx_1 \cdots dx_N) \propto \left(\prod_i \mu_i(dx_i) e^{f_i(x_i)} \right) \exp\left(\sum_{i<j} w_{ij}(x_i, x_j) + \cdots \right).$$

Proposition 21.1 (Many-system coordination cost)**.** *Let ρ be a joint law on $\prod_{i=1}^N \mathcal{X}_i$ and let $\mu = \bigotimes_i \mu_i$. Then*

$$\boxed{D_{\mathrm{KL}}(\rho\|\mu) = \sum_{i=1}^N D_{\mathrm{KL}}(\rho_i\|\mu_i) + D_{\mathrm{KL}}\left(\rho \middle\| \bigotimes_{i=1}^N \rho_i\right).}$$

The final term is the total correlation of the joint state.

Technical proof — optional on a first reading. Insert $\bigotimes_i \rho_i$ into the log-density ratio and separate the marginal terms. □

Total correlation is the information in the joint law that is absent from every marginal alone. It is the informational cost of maintaining coordinated dependence across many components.

In plain language. The marginals tell us what each component does separately. Total correlation measures what can be known only from the components together.

22 Reference Families and Aggregation

A population can carry several reference cases in three mathematically different ways. They answer different questions and must remain distinct.

Definition 22.1 (Reference family, portfolio, and latent-case law)**.** Let $\{\mu_i : i \in I\}$ be reference measures. A reference family preserves unresolved cases; a weighted portfolio adds declared weights; a latent-case law retains the case label in the represented state.

CORE

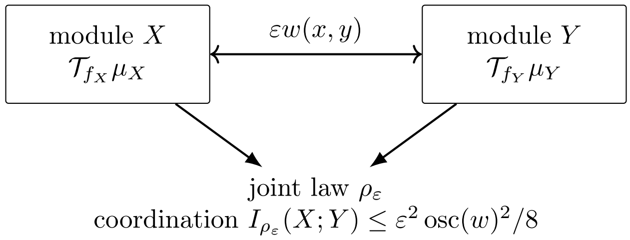

Figure 3: Near decomposability. Independent modules compose exactly. A bounded cross-module potential creates a joint law whose departure from the product and whose mutual information are controlled by interaction amplitude.

Whenever weights are used below, $w_i \ge 0$ and $\sum_i w_i = 1$.

(i) The **reference family** $\mathcal{K} = \{\mu_i : i \in I\}$ is the full answer when no weights or selector among the cases are supplied.

(ii) A **reference portfolio** is the single law $\bar{\mu} = \sum_i w_i\mu_i$, where the weights w_i are additional grounds.

(iii) A **latent-case reference** is the joint law $\widehat{\mu}(di, dx) = w_i\mu_i(dx)$, which keeps the case identity i in the represented state.

A portfolio is not the same answer as an unweighted family. It turns several cases into one predictive law. A latent-case representation preserves their provenance and permits evidence or value to update both the state within a case and the standing of the cases.

Theorem 22.1 (Tilting a reference portfolio)**.** *Suppose f is admissible for every μ_i and $\bar{\mu} = \sum_i w_i\mu_i$. Then*

$$\boxed{\mathcal{T}_f\bar{\mu} = \sum_i \widetilde{w}_i\,\mathcal{T}_f\mu_i, \qquad \widetilde{w}_i = \frac{w_i Z_{\mu_i}(f)}{\sum_j w_j Z_{\mu_j}(f)}.}$$

optional proof *Proof.* The unnormalised selected measure is

$$e^f\bar{\mu} = \sum_i w_i e^f\mu_i = \sum_i w_i Z_{\mu_i}(f)\mathcal{T}_f\mu_i.$$

Normalising gives the result. □

Corollary 22.2 (Case selection in the latent representation)**.** *Under $\widehat{\mu}(di, dx) = w_i\mu_i(dx)$ and a drive $f(i,x)$, the selected posterior weight of case i is proportional to*

$$w_i Z_{\mu_i}(f_i), \qquad f_i(x) = f(i,x),$$

and the selected conditional law inside that case is $\mathcal{T}_{f_i}\mu_i$.

optional proof *Proof.* Apply the portfolio theorem on the joint space (i,x) with reference $\bar{\mu}(di, dx) = w_i\mu_i(dx)$ and case-dependent drive $f(i,x)$. Marginalising the selected joint law over x gives the stated case weights, and conditioning on i gives $\mathcal{T}_{f_i}\mu_i$. □

Once case identity and standing are represented, ordinary reference-relative selection updates models, cultures, institutions, or strategies on the enlarged state (i,x).

CORE

22.1 Logarithmic pooling constructs a consensus point

When the references are equivalent to a common measure λ, define

$$\mu_{\text{geo}}(dx) \propto \prod_i \left(\frac{d\mu_i}{d\lambda}(x)\right)^{w_i} \lambda(dx).$$

Proposition 22.3 (Common tilt commutes with logarithmic pooling)**.** *Whenever the pool is well defined,*

$$\mathrm{Geo}_w\big(\mathcal{T}_f\mu_1, \dots, \mathcal{T}_f\mu_n\big) = \mathcal{T}_f\, \mathrm{Geo}_w(\mu_1, \dots, \mu_n).$$

optional proof *Proof.* The weighted product of the tilted densities is proportional to

$$e^{f\sum_i w_i} \prod_i \left(\frac{d\mu_i}{d\lambda}\right)^{w_i} = e^f \prod_i \left(\frac{d\mu_i}{d\lambda}\right)^{w_i}.$$

□

Mixture, latent representation, and logarithmic pooling answer different questions. A family preserves unresolved plurality. A portfolio prices cases with supplied weights. Logarithmic pooling creates a consensus reference. Each compression is rational only when its weights or pooling rule are part of the problem.

23 Hierarchy as Retained Partial Success

Let the state be partitioned into blocks $x = (x_1, \dots, x_B)$ and let

$$f(x) = \sum_{b=1}^{B} f_b(x_b) + \varepsilon w(x).$$

At $\varepsilon = 0$, the selected law factorises exactly. At small ε, Theorem 18.1 bounds the information error of treating the blocks independently. If within-block potentials update on a faster timescale than w, each module tracks a conditional equilibrium while the arrangement among modules moves on the slower clock. The domain-specific generator determines the precise singular-perturbation theorem; the common information bound determines when such a theorem can be useful (Simon and Ando, 1961).

What hierarchy buys. A retained module stores a partial solution. Disturbance elsewhere need not erase its internal adaptation, and search over arrangements need not reopen every internal degree of freedom. Hierarchy therefore lowers the effective informational and kinetic cost of adaptive search. Simon's stable intermediate form is a reference that can survive while the next level changes.

Interaction-closure result. Independent systems compose exactly. Dependence costs mutual information or total correlation; weak coupling bounds that cost and leaves subsystem marginals in the tilt family with derived drives. Inherited dependence, present interaction, and unresolved reference families remain separately typed. Hierarchy becomes useful when retained within-module adaptation dominates cross-module coupling. These results specify when components remain components of one adaptive whole.

Part V
Channels, Scale, and Emergence

At each level of complexity entirely new properties appear.
P. W. Anderson, *More Is Different*, 1972

24 Deterministic Coarse-Graining

Let $g : \mathcal{X} \to \mathcal{Y}$ be a measurable map. It may record a macrostate, sufficient statistic, aggregate, phenotype, institutional category, or any other declared resolution. Write $g_{\#}\mu$ for the pushforward of μ.

Theorem 24.1 (Exact coarse-graining closure)**.** *Let $\rho = \mathcal{T}_f[\mu]$ and let $\bar{\mu} = g_{\#}\mu$. Define the effective drive*

$$\boxed{\bar{f}(y) = \log \mathbb{E}_\mu\left[e^{f(X)} \mid g(X) = y\right]}$$

for $\bar{\mu}$-almost every y. Then

$$\boxed{g_{\#}\rho = \mathcal{T}_{\bar{f}}[\bar{\mu}].}$$

Thus the reference–tilt form survives arbitrary measurable coarse-graining exactly.

optional proof *Proof.* For every measurable $B \subseteq \mathcal{Y}$,

$$(g_{\#}\rho)(B) = \frac{1}{Z_\mu(f)} \int_{g^{-1}(B)} e^f \, d\mu.$$

Disintegrating with respect to $Y = g(X)$ gives

$$(g_{\#}\rho)(B) = \frac{1}{Z_\mu(f)} \int_B \mathbb{E}_\mu[e^f \mid Y = y] \, \bar{\mu}(dy).$$

The conditional expectation is $e^{\bar{f}(y)}$, and its integral under $\bar{\mu}$ is $Z_\mu(f)$. □

The theorem prescribes both transformed objects: the pushforward reference and the conditional-log-moment drive.

Corollary 24.2 (Effective value)**.** *If $f = V/\tau$ with common $\tau > 0$, then the macrostate has effective value*

$$\boxed{V_{\text{eff}}(y) = \tau \log \mathbb{E}_\mu\left[e^{V(X)/\tau} \mid g(X) = y\right].}$$

At low temperature this approaches the maximum supported micro-value within the fibre $g^{-1}(y)$; at high temperature it approaches the conditional mean plus fluctuation corrections.

optional proof *Proof.* Multiply the effective-drive identity in Theorem 24.1 by the common information price τ. The limiting statements are the standard log-sum-exp limits. □

24.1 The first fluctuation correction

When the conditional cumulant expansion exists,

$$V_{\text{eff}}(y) = \mathbb{E}[V \mid y] + \frac{1}{2\tau}\,\mathrm{Var}(V \mid y) + \frac{\kappa_3(V \mid y)}{6\tau^2} + O(\tau^{-3}).$$

A macrostate can receive high effective value because its conditional mean is high or because its fibre contains a high-valued tail. Emergence enters through conditional fluctuation structure.

Proposition 24.3 (Jensen and fibre bounds)**.** *For $f = V/\tau$,*

$$\boxed{\mathbb{E}[V \mid y] \le V_{\text{eff}}(y) \le \operatorname{ess\,sup}(V \mid y).}$$

On a finite fibre with conditional reference masses p_j,

$$\max_j \{V_j + \tau \log p_j\} \le V_{\text{eff}} \le \max_j V_j.$$

optional proof *Proof.* The lower bound follows from Jensen. The upper bound follows because $e^{V/\tau}$ is bounded by the exponential of the essential supremum. On a finite fibre, the conditional exponential sum is at least its largest term. □

25 Stochastic Channels

A deterministic map is a special channel. Let $K(dy \mid x)$ be a Markov kernel from $\mathcal{X}$ to $\mathcal{Y}$. Under the reference joint law, draw $X \sim \mu$ and then $Y \sim K(\cdot \mid X)$. Write μK for the output law.

Theorem 25.1 (Channel closure)**.** *Let $\rho = \mathcal{T}_f[\mu]$ and $\nu = \mu K$. Define*

$$\boxed{f_K(y) = \log \mathbb{E}_{\mu,K}\left[e^{f(X)} \mid Y = y\right].}$$

Then

$$\boxed{(\mathcal{T}_f[\mu])K = \mathcal{T}_{f_K}[\mu K].}$$

The adaptive form therefore survives any declared stochastic observation, actuation, communication, transition, or aggregation channel applied to both reference and selected law. A support-expanding transition is a generation step rather than a closed tilt.

optional proof *Proof.* For measurable $B \subseteq \mathcal{Y}$,

$$((\mathcal{T}_f\mu)K)(B) = \frac{1}{Z_\mu(f)} \int e^{f(x)} K(B \mid x)\mu(dx).$$

Under the reference joint law this is

$$\frac{1}{Z_\mu(f)} \int_B \mathbb{E}[e^{f(X)} \mid Y = y]\,\nu(dy),$$

which is the stated tilt because $\int e^{f_K}\, d\nu = Z_\mu(f)$. □

Epistemic evidence routes and economic action or diffusion channels can both be represented as stochastic kernels. The same conditional log-moment carries the tilt through either kind of channel.

Corollary 25.2 (Functorial channel composition)**.** *Let K map $\mathcal{X}$ to $\mathcal{Y}$ and L map $\mathcal{Y}$ to $\mathcal{Z}$. Then, relative to the induced reference chain,*

$$\boxed{(f_K)_L = f_{KL} \quad \textit{almost everywhere.}}$$

Equivalently, passing a tilt through K and then L gives the same effective drive as passing it through the composite channel KL.

optional proof

Proof. By the tower property,

$$\mathbb{E}[e^{f(X)} \mid Z] = \mathbb{E}\Big[\mathbb{E}[e^{f(X)} \mid Y] \mid Z\Big] = \mathbb{E}[e^{f_K(Y)} \mid Z].$$

Taking logarithms gives the result. □

26 Emergence as Effective Drive

The coarse-grained law has the same form but not the same content. Its effective drive is

$$\bar{f}(y) = \log \mathbb{E}[e^f \mid y].$$

Three consequences follow.

26.1 Emergence is reference-relative

The conditional expectation is taken under μ. Change the inherited micro-reference and the macro-drive changes even if pointwise micro-value does not. A macro-property therefore depends on microscopic laws and on which microstates the inherited reference makes available and typical.

26.2 Emergence contains hidden multiplicity

If many microstates support one macrostate, their reference mass contributes to $\bar{f}$. A macrostate can be favoured by one exceptionally high-valued microstate or by a vast family of moderately valued ones. Quality and multiplicity enter one log-partition.

26.3 Emergence can create effective interaction

Suppose the micro-drive is additive across components, but the coarse-graining couples them by mapping many joint configurations to one macrostate. The conditional log-moment need not remain additive. Effective interactions can therefore appear at the macro-level even when the microscopic drive separates. That is one precise sense in which more is different (Anderson, 1972).

27 Exact Information Accounting Across Scale

Channel closure determines the output law. The KL chain rule identifies what the channel hides.

Let $P_0(dx, dy) = \mu(dx)K(dy \mid x)$ be the reference joint law and let

$$P_f(dx, dy) = \mathcal{T}_f\mu(dx)K(dy \mid x)$$

be the selected joint law. Write $\nu_0 = \mu K$ and $\nu_f = (\mathcal{T}_f\mu)K$ for their output marginals.

Theorem 27.1 (Adaptive information decomposition)**.** *Whenever the terms are finite,*

$$\boxed{D_{\mathrm{KL}}(\mathcal{T}_f\mu\|\mu) = D_{\mathrm{KL}}(\nu_f\|\nu_0) + \mathbb{E}_{Y\sim\nu_f} D_{\mathrm{KL}}\big(P_f(dx \mid Y)\|P_0(dx \mid Y)\big).}$$

optional proof *Proof.* Because the selected and reference joint laws share the same channel,

$$D_{\mathrm{KL}}(P_f\|P_0) = D_{\mathrm{KL}}(\mathcal{T}_f\mu\|\mu).$$

Applying the marginal–conditional chain rule through Y gives the displayed decomposition. □

The first term is adaptive change visible at the output resolution. The second is adaptive structure hidden inside the output fibres. Coarse-graining does not destroy the accounting; it moves information into a conditional remainder.

Corollary 27.2 (Data processing)**.**

$$D_{\mathrm{KL}}((\mathcal{T}_f\mu)K\|\mu K) \le D_{\mathrm{KL}}(\mathcal{T}_f\mu\|\mu).$$

A channel cannot increase distinguishability between the selected and reference laws.

optional proof *Proof.* The adaptive information decomposition writes the input divergence as the output divergence plus a nonnegative conditional remainder. Dropping that remainder gives the inequality. □

Proposition 27.3 (Equality and sufficiency)**.** *Equality in data processing holds exactly when the conditional laws of X given Y agree under P_f and P_0. Equivalently, under standard regularity, the likelihood ratio $d\mathcal{T}_f\mu/d\mu = e^f/Z$ is recoverable from the output.*

optional proof *Proof.* Equality holds exactly when the conditional KL remainder in Theorem 27.1 vanishes, which is equivalent to equality of the two conditional laws almost everywhere. Under standard regularity this holds exactly when the likelihood ratio is measurable with respect to the output. □

A sufficient macrostate loses no information about the adaptive tilt. An insufficient one preserves the form but hides part of the drive in unresolved conditional structure.

28 Autonomous Macro-Dynamics

One-step closure is exact for every channel. Autonomous dynamics at the macro-level requires more: the next effective drive must be computable from the macrostate without reconstructing unresolved micro-details.

Definition 28.1 (Adaptive lumpability)**.** A coarse-graining $g : \mathcal{X} \to \mathcal{Y}$ is **adaptively lumpable** for a family of references and drives when there exists a macro-update depending only on $(g_\#\mu, \bar{f})$ such that the pushforward of every admissible micro-update equals that macro-update, and the next effective drive is determined by the retained macrostate and declared macro-environment.

Proposition 28.1 (Sufficient condition for autonomous closure)**.** *Suppose:*

(i) *the drive is g-measurable,* $f = \tilde{f} \circ g$*;*
(ii) *transition and action channels are lumpable through* g*;*
(iii) *retention depends only on the pushed-forward selected law.*

Then the macro-system evolves autonomously by the same reference–tilt update with drive $\tilde{f}$*.*

optional proof *Proof.* If $f = \tilde{f} \circ g$, then

$$\mathbb{E}[e^f \mid g = y] = e^{\tilde{f}(y)},$$

so the effective drive is exactly $\tilde{f}$. Channel lumpability and macro-retention carry the remaining update through g. □

When the condition fails, the macro-law still exists at each step, but its effective drive carries memory of the conditional micro-distribution. The macrostate is then non-Markovian unless that hidden state is added. The omitted conditional state is the memory introduced by coarse-graining.

29 Coarse-Graining Flow and the Threshold for Renormalisation

A chain of coarse-grainings

$$\mathcal{X}_0 \xrightarrow{K_1} \mathcal{X}_1 \xrightarrow{K_2} \cdots \xrightarrow{K_m} \mathcal{X}_m$$

produces references

$$\mu_{r+1} = \mu_r K_{r+1}$$

and effective drives

$$f_{r+1} = (f_r)_{K_{r+1}}.$$

Functoriality gives

$$f_m = f_{K_1 K_2 \cdots K_m}.$$

Thus the unrestricted pair (μ, f) has an exact coarse-graining flow. Every level asks the same probability-change question at a different resolution, and every lost distinction appears in the conditional information remainder.

Definition 29.1 (Restricted adaptive family)**.** Let $\{(\mu_\theta, f_\theta) : \theta \in \Theta\}$ be a parameterised family. It is **closed under a channel** K when there exists a parameter map $R_K : \Theta \to \Theta'$ and additive gauges c_θ such that

$$\mu_\theta K = \mu'_{R_K(\theta)}, \qquad (f_\theta)_K = f'_{R_K(\theta)} + c_\theta.$$

Only at this restricted level does the flow become a renormalisation map on parameters. Fixed points, relevant and irrelevant directions, critical exponents, and universality classes require additional structure: a chosen family, rescaling, regularity, and a study of the iterated map R_K. The exact channel theorem supplies the substrate for that analysis; it does not by itself prove a full renormalisation-group theory.

A resolution is part of the problem. The macro-description is not uniquely privileged by the micro-law alone. A channel or partition declares which distinctions are retained. Different coarse-grainings produce different effective references and drives. Once the channel is stated, the output law is fixed; whether the resulting effective object is simple, autonomous, or universal is a further theorem.

Full-answer preservation also governs scale: a resolution may compress only the distinctions its map declares irrelevant. Every surviving macro-distinction inherits its conditional weight.

30 The Scale Closure Theorem

Theorem 30.1 (State-space scale closure). *The class of reference–tilt laws is closed under:*

(i) *deterministic coarse-graining;*

(ii) *arbitrary stochastic channels;*

(iii) *composition of channels;*

(iv) *repeated changes of resolution.*

The transformed reference is the pushforward of the original reference. The transformed drive is the logarithm of the conditional exponential moment of the original drive. Autonomous macro-dynamics obtains when the declared macrostate is sufficient or lumpable for the future update.

optional proof *Proof.* Items (i)–(iii) are the preceding theorem and corollary. Item (iv) follows by induction and functoriality. The final statement is the adaptive-lumpability condition. □

The theorem supplies exact closure of the unrestricted state-space form under scale. It does not guarantee a low-dimensional macro-theory; that stronger result begins at adaptive lumpability or restricted-family closure.

Part VI
Full Answers and Path-Space Adaptation

Information is information, not matter or energy.

Norbert Wiener, *Cybernetics*, 1948

Full-answer closure and path-space closure concern orthogonal distinctions. The first preserves how determinate the problem is; the second changes the represented object from states to trajectories. A full answer may contain state laws or path laws, and neither type is a refinement of the other.

31 Full Answers Survive Adaptation

A single reference measure is not always warranted. When the grounds leave several references, representations, or minimisers standing, *Intelligent Epistemology* retains the family itself as the answer. Closure must preserve that shape.

Definition 31.1 (Full reference answer)**.** Let $\mathcal{K}$ be a family of attained reference measures, let $\approx$ record the equivalences the problem declares among them, and let $\mathcal{U}$ record permitted cases for which a requested normalised tilt is empty or unattained. The triple

$$\mathbf{K} = (\mathcal{K}, \approx, \mathcal{U})$$

is a **full reference answer**. A point reference is the singleton case.

For each $\mu \in \mathcal{K}$, let f_μ be the potential supplied in that case. Define

$$\mathrm{TiltAns}(\mathbf{K}, f) = \left(\{(\mu, \mathcal{T}_{f_\mu}\mu) : \mu \in \mathcal{K},\ 0 < Z_\mu(f_\mu) < \infty\}, \approx_f, \mathcal{U}_f\right),$$

where $\approx_f$ is the transported equivalence and $\mathcal{U}_f$ contains the inherited boundary cases together with any newly nonnormalisable case.

Theorem 31.1 (Full-answer closure)**.** *Suppose the declared point operations are defined on every attained case and respect the equivalence relation $\approx$. Then the time, product, channel, and coarse-graining closures lift casewise to full reference answers. In particular:*

(i) *sequential tilts carry every case to $\mathcal{T}_{f_\mu+g_\mu}\mu$;*

(ii) *product families carry every permitted product case to the product of its selected factors under additive drives;*

(iii) *a channel K carries every case to*

$$\mathcal{T}_{(f_\mu)_K}(\mu K), \qquad (f_\mu)_K(y) = \log \mathbb{E}_{\mu,K}[e^{f_\mu(X)} \mid Y = y];$$

(iv) *empty, nonnormalisable, inequivalent, and unresolved cases remain visible rather than being replaced by a selected point.*

optional proof *Proof.* Items (i)–(iii) apply the corresponding point theorem to every member of the family. Because the operations respect equivalent descriptions, the equivalence relation transports with the cases. Omitting an attained image would arbitrarily reduce the answer; adding a point selected from several images would require a selector not supplied by the problem. Boundary cases remain boundary cases until further structure changes them. □

Corollary 31.2 (No mixture without weights). *A family $\{\mu_i\}$ cannot be replaced by $\sum_i w_i\mu_i$ unless the weights are included among the grounds. A portfolio, a latent-case model, and a credal family are distinct full answers.*

optional proof *Proof.* Without supplied weights, replacing the family by a mixture adds a selector not contained in the grounds. Full-answer preservation therefore retains the family. □

State-space closure does not force epistemic precision; it preserves whatever precision the grounds purchased.

32 Path Reference and Exponential Change of Measure

A system acts through trajectories as well as terminal distributions. Let $(\Omega, \mathcal{G})$ be a path space and let Q be a reference law over trajectories. The law may encode passive dynamics, inherited motion, environmental noise, or the uncontrolled process against which intervention is priced.

Definition 32.1 (Admissible path potential). A measurable functional $F : \Omega \to \mathbb{R}$ is **Q-admissible** when

$$0 < Z_Q(F) = \mathbb{E}_Q[e^F] < \infty.$$

Define the tilted path law

$$\boxed{\frac{d\mathbb{T}_F Q}{dQ}(\omega) = \frac{e^{F(\omega)}}{Z_Q(F)}.}$$

Theorem 32.1 (Path-space variational identity). *For every path law $P \ll Q$ with finite terms,*

$$\boxed{\mathbb{E}_P[F] - D_{\mathrm{KL}}(P\|Q) = \log Z_Q(F) - D_{\mathrm{KL}}(P\|\mathbb{T}_F Q).}$$

Hence $\mathbb{T}_F Q$ is the unique maximiser of

$$P \longmapsto \mathbb{E}_P[F] - D_{\mathrm{KL}}(P\|Q).$$

In value units, $F = \mathcal{V}/\tau$ gives

$$P^* = \arg\max_P \{\mathbb{E}_P[\mathcal{V}] - \tau D_{\mathrm{KL}}(P\|Q)\}.$$

optional proof *Proof.* The proof is the state-space reference–tilt identity applied on $(\Omega, \mathcal{G})$. □

Proposition 32.2 (Path potentials compose). *For admissible F and G,*

$$\boxed{\mathbb{T}_G(\mathbb{T}_F Q) = \mathbb{T}_{F+G} Q.}$$

optional proof *Proof.* Apply the tilt algebra on path space: the two Radon–Nikodym factors multiply to e^{F+G} and the normalising constants cancel after renormalisation. □

The same change-of-measure algebra applies on any represented space; path space makes the kinetic branch explicit.

33 Feynman–Kac Recursion

Let Q be the Markov reference law

$$Q(dx_{0:T}) = \mu_0(dx_0) \prod_{t=0}^{T-1} M_t(dx_{t+1} \mid x_t),$$

and let

$$F(x_{0:T}) = \sum_{t=0}^{T-1} f_t(x_{t+1}).$$

Theorem 33.1 (Normalised Feynman–Kac flow)**.** *Define $\eta_0 = \mu_0$ and recursively*

$$\boxed{\eta_{t+1} = \mathcal{T}_{f_t}(\eta_t M_t).}$$

Then η_T is the terminal marginal of the path law $\mathbb{T}_F Q$. Equivalently, transition through M_t followed by selection through f_t is the normalised Feynman–Kac recursion.

optional proof *Proof.* Define the unnormalised measures

$$\gamma_0 = \mu_0, \qquad \gamma_{t+1}(dy) = e^{f_t(y)} \int \gamma_t(dx) M_t(dy \mid x).$$

Iterating and integrating the path law over $x_0, \ldots, x_{T-1}$ gives γ_T as the unnormalised terminal marginal of $e^F Q$. Normalising at every step yields $\eta_t = \gamma_t/\gamma_t(1)$ and the displayed recursion. □

The recursion is standard in filtering, rare-event simulation, genetic algorithms, and interacting particle systems (Kac, 1949; Del Moral, 2004). It is also the marginal flow of a path-space change of measure. The kernel may preserve represented support or carry mass into states absent from the preceding marginal; the latter case is generation.

Remark 33.1 (Transition and history-dependent drives). A potential depending on (x_t, x_{t+1}) or the full history is handled directly on path space. It can also be made state-local by enlarging the state to carry the required history. The cost of Markovian closure is therefore visible in the representation.

34 Path Information and Kinetic Cost

Theorem 34.1 (Sequential path-KL accounting)**.** *Let P and Q be path laws with regular conditional transitions. Whenever finite,*

$$\boxed{D_{\mathrm{KL}}(P\|Q) = D_{\mathrm{KL}}(P_0\|Q_0) + \sum_{t=0}^{T-1} \mathbb{E}_P D_{\mathrm{KL}}\big(P(dx_{t+1} \mid X_{0:t})\|Q(dx_{t+1} \mid X_{0:t})\big).}$$

For Markov laws, the conditional kernels depend only on X_t.

optional proof *Proof.* Factor the Radon–Nikodym derivative into the initial density ratio and the product of conditional density ratios, take logarithms, and integrate under P. □

The total information price of a trajectory is therefore its initial departure plus every conditional departure paid along the route. A terminal distribution does not record how that endpoint was reached.

Theorem 34.2 (Endpoint decomposition). *Let X_T have laws ρ_T under P and μ_T under Q. Then*

$$\boxed{D_{\mathrm{KL}}(P\|Q) = D_{\mathrm{KL}}(\rho_T\|\mu_T) + \mathbb{E}_{x\sim\rho_T} D_{\mathrm{KL}}\big(P(d\omega \mid X_T = x)\|Q(d\omega \mid X_T = x)\big).}$$

optional proof *Proof.* Apply the marginal–conditional chain rule to the map $\omega \mapsto X_T(\omega)$. □

Corollary 34.3 (An exact $H + C + K$ decomposition). *For a terminal value $V(X_T)$ and price $\tau > 0$,*

$$-\mathbb{E}_P[V(X_T)] + \tau D_{\mathrm{KL}}(P\|Q) = H(\rho_T) + C(\rho_T\|\mu_T) + K(P\|Q; X_T),$$

where

$$H(\rho_T) = -\mathbb{E}_{\rho_T}[V], \qquad C = \tau D_{\mathrm{KL}}(\rho_T\|\mu_T),$$

and

$$K = \tau \mathbb{E}_{\rho_T} D_{\mathrm{KL}}\big(P(d\omega \mid X_T)\|Q(d\omega \mid X_T)\big).$$

The first term prices the endpoint, the second its departure from the reference terminal law, and the third the additional route information not determined by the endpoint.

optional proof *Proof.* Substitute the endpoint decomposition of Theorem 34.2 into $-\mathbb{E}_P[V(X_T)] + \tau D_{\mathrm{KL}}(P\|Q)$ and group the three terms. □

On path space the economic action separates into three derived terms. H prices the endpoint, C its departure from the reference terminal law, and K the remaining route information; all three arise from terminal value and one path relative entropy.

35 Girsanov and Controlled Diffusions

Theorem 35.1 (Kinetic KL for drift control). *Let Q be the law of*

$$dX_t = b_t(X_t)\,dt + \sqrt{2}\,dW_t$$

and let P^u be the law with the same initial distribution and controlled drift

$$dX_t = (b_t(X_t) + u_t)\,dt + \sqrt{2}\,dW_t,$$

under standard existence and Novikov conditions. Then

$$\boxed{D_{\mathrm{KL}}(P^u\|Q) = \frac{1}{4}\mathbb{E}_{P^u}\int_0^T \|u_t\|^2 dt.}$$

optional proof *Proof.* Girsanov's theorem gives the likelihood ratio between the two path laws. Taking expectation under P^u cancels the stochastic integral and leaves one quarter of the control energy because the diffusion covariance is $2I$ (Girsanov, 1960; Föllmer, 1985). ☐

Corollary 35.2 (Path-space bounded choice)**.** *For a terminal reward $V(X_T)$,*

$$\sup_u \left\{ \mathbb{E}_{P^u}[V(X_T)] - \frac{\tau}{4}\mathbb{E}_{P^u} \int_0^T \|u_t\|^2 dt \right\}$$

is the controlled-diffusion restriction of the path variational problem

$$\sup_P \{\mathbb{E}_P[V(X_T)] - \tau D_{\mathrm{KL}}(P\|Q)\}.$$

optional proof *Proof.* Restrict the path variational problem to laws generated by admissible controls u and substitute the Girsanov identity from Theorem 35.1. ☐

The kinetic cost in *Intelligent Economics* is therefore the physical-control representation of the same path information price. Different dynamics can realise the same selected endpoint; the path law records their differing costs.

36 Doob Transformation and the Driven Process

Let the Markov reference Q be as above and let the path drive be a terminal potential $g(X_T)$. Define

$$h_T(x) = e^{g(x)}, \qquad h_t(x) = \int M_t(dy \mid x) h_{t+1}(y).$$

Theorem 36.1 (Finite-horizon Doob transform)**.** *If h_t is positive and finite, the path tilt $\mathbb{T}_{g(X_T)}Q$ is Markov with initial law*

$$\mu_0^h(dx) = \frac{h_0(x)}{\mathbb{E}_{\mu_0}[h_0]}\mu_0(dx)$$

and transitions

$$\boxed{M_t^h(dy \mid x) = M_t(dy \mid x)\frac{h_{t+1}(y)}{h_t(x)}.}$$

optional proof *Proof.* The conditional weight of a continuation from state x is $h_t(x)$. Multiplying the reference transition by the next continuation weight and dividing by the current one produces a normalised kernel. The product of the transformed initial density and transformed kernels telescopes to $e^{g(X_T)}dQ/Z$. ☐

The theorem turns a global path preference into local driven dynamics. In continuous time the corresponding h-transform modifies the generator; under diffusion regularity it adds the gradient of the log continuation value to the drift. Conditioned processes, rare-event driven processes, Schrödinger bridges, and linearly solvable control occupy this established branch (Doob, 1957; Léonard, 2014; Chetrite and Touchette, 2015; Todorov, 2006).

37 Path Channels and Full Path Answers

A sensor, report, intervention record, or macro-trajectory is a channel $C(dy \mid \omega)$ from path space to an output space. The state-space channel theorem applies unchanged.

Theorem 37.1 (Path-channel closure)**.** *Let $P^* = \mathbb{T}_F Q$. Define*

$$F_C(y) = \log \mathbb{E}_{Q,C}[e^{F(\omega)} \mid Y = y].$$

Then

$$\boxed{P^* C = \mathcal{T}_{F_C}(QC).}$$

For a family of path references, the result holds casewise and returns the full family of output path laws and effective drives.

optional proof *Proof.* Apply Theorem 25.1 on path space with reference Q, drive F, and channel C. The family statement follows casewise from Theorem 31.1. □

Epistemic channels can transport evidence about trajectories, while economic and physical channels can transport controlled motion. In either case, coarse-grained histories inherit a conditional log-moment drive and an exact hidden-path information remainder.

The two lifts. Closure requires neither an unjustified point reference nor a state-only description. Every surviving reference case carries its own transformed drive, while path selection uses the same relative-entropy variational law on trajectory space. Normalised Feynman–Kac recursion gives the Markov marginal; path KL separates sequential, endpoint, and route information; Girsanov converts path divergence into kinetic control cost; and Doob transformation converts global path value into local driven dynamics. Full-answer precision and path dynamics belong to the same architecture without being collapsed into one type.

Part VII
Openness, Novelty, and Resilience

Natural selection can do nothing until favourable individual differences or variations occur.

Charles Darwin, *On the Origin of Species*, 1859

38 The Boundary of Closed Adaptation

A finite tilt changes the relative standing of represented states while preserving the reference measure class. Repeating the operation does not alter that fact. The resulting boundary separates closed reweighting from the admission of new possibilities.

Theorem 38.1 (Support inheritance through time)**.** *Let*

$$\mu_{t+1} = \mathcal{T}_{f_t}\mu_t,$$

where every f_t is finite μ_t-almost everywhere. Then for every $s > t$,

$$\mu_s \sim \mu_t.$$

In particular, every event assigned zero probability at time t remains null at every later time.

optional proof *Proof.* Proposition 7.5 gives $\mu_{t+1} \sim \mu_t$ at each step. Equivalence is transitive. □

Corollary 38.2 (No endogenous rediscovery)**.** *If $\mu_t(A) = 0$, no finite sequence of evidence, value, payoff, or fitness tilts internal to the represented state space can produce $\mu_s(A) > 0$ for $s > t$.*

optional proof *Proof.* Theorem 38.1 preserves every null event under each later finite tilt. □

The corollary identifies a boundary shared by Bayes, selection, and soft optimisation. A hypothesis outside the model class, a phenotype outside the mutation-accessible population, an action absent from policy support, or a social possibility excluded by the institutional reference cannot be recovered by reweighting alone.

Hard exclusions and limits can contract support. If a drive takes the value $-\infty$ on a set, or if a sequence of finite tilts converges to a boundary measure, possibilities can disappear from the effective state. The closed dynamics is therefore asymmetric: it can preserve or lose standing, but it cannot create standing from zero.

39 Generation as a Separate Operator

Enlarging the represented possibilities requires an operation separate from selection. The canonical order considered first is generation followed by selection.

Definition 39.1 (Generation channel)**.** Let $M_t(dy \mid x)$ be a Markov kernel from the inherited state space to the state space admitted at time $t+1$. The generated reference is

$$\nu_t = \mu_t M_t.$$

An **open adaptive step** is

$$\boxed{\mu_{t+1} = \mathcal{T}_{f_t}(\mu_t M_t).}$$

The identity kernel gives closed adaptation. A nontrivial kernel can represent mutation, recombination, exploration, invention, model enlargement, migration, institutional reform, or another declared source of new standing.

Theorem 39.1 (The exact source of new support)**.** *Let $\mu_{t+1} = \mathcal{T}_{f_t}(\mu_t M_t)$ with f_t finite $(\mu_t M_t)$-almost everywhere. Then*

$$\mu_{t+1} \sim \mu_t M_t.$$

Consequently an event B receives positive mass after the step exactly when the generation channel gives it positive generated mass:

$$\boxed{\mu_{t+1}(B) > 0 \quad \Longleftrightarrow \quad (\mu_t M_t)(B) > 0.}$$

optional proof *Proof.* Apply Proposition 7.5 to the generated reference $\mu_t M_t$. □

Selection determines which generated possibilities expand. Generation determines which possibilities enter comparison at all. The two operations are complementary and irreducible.

Corollary 39.2 (Generation and selection generally do not commute)**.** *Let $M(dy \mid x)$ be a Markov kernel from $\mathcal{X}$ to $\mathcal{Y}$, let h be a finite admissible drive on $\mathcal{X}$, and let f be a finite admissible drive on $\mathcal{Y}$. Define the channel-transformed drive*

$$h_M(y) = \log \mathbb{E}_{\mu,M}[e^{h(X)} \mid Y = y].$$

Then

$$\boxed{\mathcal{T}_f(\mu M) = (\mathcal{T}_h \mu)M}$$

if and only if $f - h_M$ is constant (μM)-almost everywhere.

optional proof *Proof.* The channel-closure theorem gives

$$(\mathcal{T}_h \mu)M = \mathcal{T}_{h_M}(\mu M).$$

Two finite tilts of the same reference agree exactly when their potentials differ by an additive constant, by Proposition 7.2. □

The order of the two operations is therefore substantive. Selection before mutation, transport, recombination, or institutional entry is equivalent to selection after it only when the post-generation drive is the conditional log-moment induced by the earlier drive. Treating the two orders as interchangeable silently changes the model.

Definition 39.2 (Mixture generation)**.** Given a reference μ, an exploratory law ν, and $\varepsilon \in [0, 1]$, define

$$\mathcal{G}_{\varepsilon,\nu}[\mu] = (1 - \varepsilon)\mu + \varepsilon\nu.$$

Proposition 39.3 (Support of mixture generation)**.** *For* $0 < \varepsilon < 1$,

$$\operatorname{supp} \mathcal{G}_{\varepsilon,\nu}[\mu] = \overline{\operatorname{supp} \mu \cup \operatorname{supp} \nu}$$

under the usual regularity on a topological state space. A subsequent finite tilt preserves that generated support.

optional proof *Proof.* Every open set meeting either support has positive mass under the corresponding component and therefore under the mixture. Outside the closure of their union both components vanish. The final claim follows from Proposition 7.5. □

A small exploratory weight can therefore reopen a possibility without forcing it to survive. The tilt still prices it against the incumbent reference and the current drive.

40 The Replicator–Mutator Limit

Let $p(t)$ be a row probability vector on a finite type space. Let Q be a conservative mutation generator, so $Q_{ij} \geq 0$ for $i \neq j$ and $\sum_j Q_{ij} = 0$. Over a short interval dt, generate through

$$M_{dt} = I + Q\,dt + o(dt),$$

then tilt by the fitness potential $\eta F_j(t)dt$.

Theorem 40.1 (Replicator–mutator limit)**.** *If*

$$p(t + dt) = \mathcal{T}_{\eta F(t)dt}\big(p(t)M_{dt}\big),$$

then

$$\boxed{\dot{p}_j = \sum_i p_i Q_{ij} + \eta p_j\big(F_j - \bar{F}\big), \qquad \bar{F} = \sum_j p_j F_j.}$$

optional proof *Proof.* Generation gives

$$q_j = p_j + dt \sum_i p_i Q_{ij} + o(dt).$$

The subsequent tilt gives

$$p_j(t + dt) = \frac{q_j(1 + \eta F_j dt + o(dt))}{1 + \eta \bar{F} dt + o(dt)}.$$

Expanding the denominator and retaining first-order terms yields the equation. □

The first term moves mass through the possibility graph. The second changes relative abundance within the represented population. In evolutionary biology they are mutation and selection. In learning they are exploration and reward. In inquiry they are hypothesis generation and evidential revision. The contents differ; the decomposition is exact.

Remark 40.1 (Recombination and structured generation). A Markov kernel is the simplest generation object. Recombination, grammar-based invention, program synthesis, and endogenous creation of a new state space may require kernels on populations, product spaces, or higher-order specifications. The theorem does not claim to derive those mechanisms. It identifies the slot they occupy before selection can act on their output.

41 The Higher-Order Problem of Invention

Intelligent Epistemology distinguishes evaluating candidates from creating the candidate class. A language, representation, or hypothesis generator enlarges the problem; inference then evaluates what the enlarged problem contains (Mostaque, 2026b). The same distinction is now dynamic.

Let G_t construct a state space, representation, and generation channel. Let $\Pi_t(G_t)$ be the adaptive problem thereby specified. Then the full loop is

$$\boxed{G_t \longrightarrow \mu_t M_t \longrightarrow \mathcal{T}_{f_t}(\mu_t M_t) \longrightarrow \text{action and retention} \longrightarrow G_{t+1}.}$$

The first arrow is creative or evolutionary. The second is inferential or selective. The third is causal in the environment. A theory of adaptation that omits the first can explain optimisation inside a world but not the appearance of a new world of options.

Novelty concerns support or representation expansion. Randomness can implement a generation channel, while unpredictability alone leaves the represented possibilities unchanged. A random draw within existing support is exploration; a deterministic construction outside it is invention. The operative question is whether the operation gives standing to a state absent from the previous problem.

42 Reference Portfolios and Resilience

A concentrated reference can be highly efficient in familiar conditions and incur infinite KL loss on an excluded case. A mixture of references pays a finite insurance premium for retaining alternatives.

Theorem 42.1 (Reference-portfolio bound)**.** *Let*

$$\bar{\mu} = \sum_{i=1}^{m} w_i \mu_i, \qquad w_i > 0, \quad \sum_i w_i = 1.$$

For every probability law $q \ll \mu_i$,

$$\boxed{D_{\mathrm{KL}}(q\|\bar{\mu}) \le D_{\mathrm{KL}}(q\|\mu_i) - \log w_i.}$$

optional proof

Proof. As measures, $\bar{\mu} \ge w_i \mu_i$. Hence on the support of q,

$$\log \frac{dq}{d\bar{\mu}} \le \log \frac{dq}{d\mu_i} - \log w_i.$$

Integrating with respect to q gives the bound. □

Corollary 42.2 (Finite excess information loss for every retained case)**.** *If the realised law is well represented by at least one component, the mixture's information loss is no more than that component's loss plus the logarithmic insurance price $-\log w_i$. If a monoculture reference gives zero mass to a realised event while one retained component supports it, the monoculture incurs infinite KL loss while the mixture remains finite.*

optional proof

Proof. Apply Theorem 42.1 to the component that represents the realised law. If the monoculture assigns zero mass to a positive-q event, its KL divergence is infinite; the retained mixture remains absolutely continuous wherever that component does. □

The bound prices diversity without appealing to an independent preference for variety. A reference portfolio caps the additional surprise paid against every retained model, strategy, culture, or ecological type.

43 Collapse and Recovery

Definition 43.1 (Three collapses). (i) **support collapse**: possibilities become null through hard exclusion, boundary projection, extinction, or model truncation;

(ii) **concentration collapse**: support formally remains, but mass becomes so concentrated that finite sampling or bounded search no longer reaches alternatives;

(iii) **representation collapse**: the language or state space ceases to express a relevant alternative.

Support collapse is irreversible under closed tilting by Theorem 38.1. Concentration collapse may be reversible under a counter-tilt if hidden mass remains nonzero, but practical recovery cost can be extreme. Representation collapse requires higher-order enlargement; a new weight inside the old model is insufficient.

Recovery requires the matching operation. A collapsed system should not be treated with one generic remedy. Counter-selection can reverse concentration. Mutation, exploration, migration, or institutional entry can reopen support. New concepts, sensors, or models can repair representation. The diagnosis is the location of the lost possibility.

The boundary result. Finite tilts cannot create support. New standing enters through generation, whose order relative to selection is constrained by conditional-log-moment compatibility and whose weak finite-state limit is replicator–mutator dynamics. Candidate creation remains distinct from evaluation; reference portfolios bound the cost of retained alternatives; recovery must match the kind of possibility lost. Closed adaptation ends where support expansion begins.

TECHNICAL

Part VIII
Positive Operators and Adaptive Closure

Only variety can destroy variety.
W. Ross Ashby, *An Introduction to Cybernetics*, 1956

In plain language. Open adaptation can be written as transition followed by reweighting. For a fixed transition and potential, the unnormalised step is a positive linear operator, and sequences of steps compose by operator multiplication. The spectral and projective results are technical; their statements carry the main line, while the proofs may be skipped on a first reading.

$$\Phi_{L_2}(\Phi_{L_1}(\gamma)) = \Phi_{L_1 L_2}(\gamma).$$

44 Positive Operators and Projective Dynamics

FIRST READING Read the definition and the normalised-composition theorem. The eigenvalue and projective-metric results deepen the same idea.

Let $M(dy \mid x)$ be a fixed Markov kernel and let f be a measurable potential on its output space. Define the positive kernel

$$L_{M,f}(x, dy) = e^{f(y)} M(dy \mid x).$$

For a finite positive measure γ with $0 < \gamma L\mathbf{1} < \infty$, define

$$\boxed{\Phi_L(\gamma) = \frac{\gamma L}{\gamma L\mathbf{1}}.}$$

Then

$$\Phi_{L_{M,f}}(\mu) = \mathcal{T}_f(\mu M).$$

For fixed M and f, one open adaptive step is linear before normalisation.

Theorem 44.1 (Normalised composition)**.** *Let L_1 and L_2 be composable positive kernels. Whenever all normalising constants are finite and nonzero,*

$$\boxed{\Phi_{L_2}(\Phi_{L_1}(\gamma)) = \Phi_{L_1 L_2}(\gamma).}$$

A finite sequence of open steps is therefore the normalised action of one positive-operator product.

Technical proof — optional on a first reading. The scalar introduced by the first normalisation cancels in the second:

$$\Phi_{L_2}\left(\frac{\gamma L_1}{\gamma L_1\mathbf{1}}\right) = \frac{\gamma L_1 L_2}{\gamma L_1 L_2\mathbf{1}}.$$

□

Closed tilts are diagonal multiplication kernels and commute. General positive kernels retain the order of transitions and potentials, so open histories are generally noncommutative.

In plain language. A long open history can be compressed into one product of positive operators followed by one final normalisation. Order matters because transition and reweighting need not commute.

Proposition 44.2 (Fixed points and eigenmeasures)**.** *A probability measure π is a fixed point of Φ_L if and only if*

$$\boxed{\pi L = \lambda\pi, \qquad \lambda = \pi L\mathbf{1} \in (0, \infty).}$$

Thus a fixed point of a time-homogeneous open step is a normalised left eigenmeasure.

Technical proof — optional on a first reading. If $\Phi_L(\pi) = \pi$, multiply by $\pi L\mathbf{1}$. The converse follows by normalising the eigenmeasure relation. □

Theorem 44.3 (Primitive finite-state limit)**.** *Let L be a primitive nonnegative matrix on a finite state space. Let $\lambda > 0$ be its Perron root and let π be its positive left eigenvector normalised to a probability vector. For every nonzero nonnegative initial vector p_0,*

$$\boxed{\frac{p_0 L^t}{p_0 L^t\mathbf{1}} \longrightarrow \pi, \qquad \frac{1}{t}\log(p_0 L^t\mathbf{1}) \longrightarrow \log\lambda.}$$

Technical proof — optional on a first reading. Primitivity gives a simple dominant eigenvalue, positive left and right eigenvectors, and a spectral gap. The Perron term dominates $p_0 L^t$; normalisation removes its scalar coefficient. The same expansion gives the logarithmic growth rate. □

General state spaces require their own irreducibility, compactness, spectral, or contraction conditions. Fixed-point equivalence alone gives neither uniqueness nor convergence.

In plain language. A stable long-run law is an eigenmeasure of the open update. Convergence to it follows only when the operator mixes strongly enough.

44.1 Projective distance

For equivalent positive probability measures μ and ν with finite likelihood-ratio oscillation, define Hilbert's projective distance by

$$d_H(\mu,\nu) = \operatorname{ess\,sup} \log \frac{d\mu}{d\nu} - \operatorname{ess\,inf} \log \frac{d\mu}{d\nu}.$$

Theorem 44.4 (Common tilts and common channels)**.** *For every finite potential f admissible for μ and ν,*

$$\boxed{d_H(\mathcal{T}_f\mu, \mathcal{T}_f\nu) = d_H(\mu,\nu).}$$

For one reference and two finite potentials f, g,

$$\boxed{d_H(\mathcal{T}_f\mu, \mathcal{T}_g\mu) = \operatorname{osc}(f-g).}$$

If M is a common Markov kernel, then

$$\boxed{d_H(\mu M, \nu M) \le d_H(\mu,\nu).}$$

Technical proof — optional on a first reading. A common tilt adds only a normalising constant to the log-density ratio. For the kernel inequality, write $r = d\mu/d\nu$. The output likelihood ratio is $\mathbb{E}_{\nu,M}[r(X) \mid Y]$, which lies between the essential infimum and supremum of r. Taking logarithmic oscillations gives the result. □

A common potential changes standing without changing projective disagreement. A common transition may contract that disagreement. In a finite state space, a strictly positive kernel has finite projective diameter and contracts by a coefficient below one.

In plain language. Applying the same value signal to two different references does not erase their inherited disagreement. Shared mixing or communication can reduce it.

Remark 44.1 (Standing conditions). The tilt algebra is a partial action over admissible potentials: every normalising constant must be finite and nonzero. Differentiation along a tilt path requires local exponential integrability. A finite interaction potential relative to a product reference requires equivalence of the joint and product laws; structural zeroes require an extended potential. Logarithmic pools use nonnegative weights summing to one.

45 Channel Transport and Information Across Scale

FIRST READING The core claim is simple: once the input reference, input potential, and channel are fixed, the output potential is determined. The KL identities say exactly how much information remains visible and how much is hidden.

Let $K(dy \mid x)$ be a Markov channel, let $\rho = \mathcal{T}_f\mu$, and write $\nu_0 = \mu K$ and $\nu_f = \rho K$. Channel closure gives

$$\nu_f = \mathcal{T}_{f_K}\nu_0, \qquad f_K(y) = \log \mathbb{E}_{\mu,K}[e^{f(X)} \mid Y = y].$$

Theorem 45.1 (Uniqueness of the transported potential)**.** *If a finite potential g satisfies $\nu_f = \mathcal{T}_g\nu_0$, then*

$$\boxed{g = f_K + c \qquad \nu_0\textit{-almost everywhere}}$$

for a constant c.

Technical proof — optional on a first reading. The output likelihood ratio is

$$\frac{d\nu_f}{d\nu_0}(y) = \frac{\mathbb{E}_{\mu,K}[e^{f(X)} \mid Y = y]}{\mathbb{E}_\mu[e^f]}.$$

The finite representation theorem identifies every potential representing this ratio up to an additive constant. □

In plain language. The effective potential after observation or aggregation is not a modelling choice. The channel determines it, up to the usual additive gauge.

Define the residual

$$r_K(x, y) = f(x) - f_K(y).$$

Then $\mathbb{E}_{\mu,K}[e^{r_K(X,Y)} \mid Y] = 1$. The transported potential is visible at the output; the residual is the within-fibre change hidden by the output.

Theorem 45.2 (Visible and hidden information)**.** *Under the required integrability conditions,*

$$\boxed{D_{\mathrm{KL}}(\rho\|\mu) - D_{\mathrm{KL}}(\rho K\|\mu K) = \mathbb{E}_{\rho,K}[r_K(X, Y)].}$$

The right-hand side is also the expected conditional KL divergence between the selected and reference input laws given the output.

In plain language. The output shows one part of the adaptive change. The remainder is the distinction still present inside states that the output treats as the same.

Technical proof — optional on a first reading. The conditional likelihood ratio inside a fibre is

$$\frac{d\rho(dx \mid Y = y)}{d\mu(dx \mid Y = y)} = e^{r_K(x,y)}.$$

Integrating its logarithm under the selected conditional law gives the conditional term in the KL chain rule. □

Theorem 45.3 (Sufficiency of a coarse-graining)**.** *Let $Y = g(X)$ be deterministic. The following are equivalent:*

(i) $D_{\mathrm{KL}}(\rho\|\mu) = D_{\mathrm{KL}}(g_\#\rho\|g_\#\mu)$*;*

(ii) $\rho(dx \mid Y) = \mu(dx \mid Y)$ *almost surely;*

(iii) f *is measurable with respect to* $\sigma(g)$*, up to an additive constant.*

For one fixed tilt, $\sigma(f)$ is the smallest lossless σ-algebra modulo null sets.

In plain language. A summary loses nothing about a particular update exactly when the potential can be read from the summary alone.

Technical proof — optional on a first reading. The KL chain rule makes (i) and (ii) equivalent. The conditional likelihood ratio is $e^{f-f_g\circ g}$, so (ii) holds exactly when $f = f_g \circ g$ almost everywhere. Taking $g = f$ preserves the full likelihood ratio. □

45.1 Repeated channels and event probabilities

For a chain $X_0 \to X_1 \to \cdots \to X_m$, let μ_r and ρ_r be the reference and selected laws visible at level r. Repeated application of the KL chain rule gives

$$\boxed{D_{\mathrm{KL}}(\rho_0\|\mu_0) = D_{\mathrm{KL}}(\rho_m\|\mu_m) + \sum_{r=1}^{m} \mathbb{E}_{\rho_r} D_{\mathrm{KL}}(\rho_{r-1|r}\|\mu_{r-1|r}).}$$

The first term remains visible at the coarsest level; each summand records the distinction hidden at one resolution.

Proposition 45.4 (Minimum KL cost of changing an event probability)**.** *Let A be an event with $p = \mu(A) \in (0,1)$ and let $q = \rho(A)$. Then*

$$\boxed{D_{\mathrm{KL}}(\rho\|\mu) \geq q \log \frac{q}{p} + (1-q) \log \frac{1-q}{1-p}.}$$

Equality holds for

$$\rho_q = q\,\mu(\cdot \mid A) + (1-q)\mu(\cdot \mid A^c).$$

The minimum information required to make A certain is $\log(1/p)$.

Technical proof — optional on a first reading. Apply data processing to the indicator channel $\mathbf{1}_A$. Equality requires the reference conditional law to be preserved on A and A^c. ☐

46 Response Geometry and Selection Across References

FIRST READING This section gives two readable consequences: covariance is the local response matrix, and selection across a portfolio separates into change between cases and change within each case.

Let $A = (A_1, \ldots, A_k)$ be a vector of observables and let

$$f_\theta(x) = \theta^\top A(x), \qquad \rho_\theta = \mathcal{T}_{f_\theta}\mu, \qquad \psi(\theta) = \log \mathbb{E}_\mu[e^{\theta^\top A}].$$

Where the required local exponential moments are finite,

$$\nabla\psi(\theta) = \mathbb{E}_{\rho_\theta}[A], \qquad \nabla^2\psi(\theta) = \mathrm{Cov}_{\rho_\theta}(A, A).$$

Theorem 46.1 (Reciprocal response and KL geometry)**.** *The response matrix satisfies*

$$\boxed{\frac{\partial}{\partial\theta_j}\mathbb{E}_{\rho_\theta}[A_i] = \mathrm{Cov}_{\rho_\theta}(A_i, A_j) = \frac{\partial}{\partial\theta_i}\mathbb{E}_{\rho_\theta}[A_j].}$$

It is symmetric and positive semidefinite. For parameter values θ and η,

$$\boxed{D_{\mathrm{KL}}(\rho_\theta\|\rho_\eta) = \psi(\eta) - \psi(\theta) - \nabla\psi(\theta)^\top(\eta - \theta).}$$

Technical proof — optional on a first reading. Differentiate the log-partition function. Substitution of the exponential-family densities gives the Bregman form of KL. ☐

Antisymmetric empirical response requires an intervention-dependent reference, delayed or path-dependent effects, a nonconservative interaction, an omitted state, or failure of the rest-point tilt model.

Proposition 46.2 (Loss from using an outdated selected law). *Let $\rho_f = \mathcal{T}_f\mu$ and $\rho_g = \mathcal{T}_g\mu$, and define*

$$J_g(q) = \mathbb{E}_q[g] - D_{\mathrm{KL}}(q\|\mu).$$

Then

$$\boxed{J_g(\rho_g) - J_g(\rho_f) = D_{\mathrm{KL}}(\rho_f\|\rho_g).}$$

Technical proof — optional on a first reading. Apply the reference–tilt variational identity with target potential g and evaluate it at $q = \rho_f$. □

Proposition 46.3 (Marginal value per nat). *Let $\beta > 0$, $\rho_\beta = \mathcal{T}_{\beta V}\mu$, $m(\beta) = \mathbb{E}_{\rho_\beta}[V]$, and $D(\beta) = D_{\mathrm{KL}}(\rho_\beta\|\mu)$. Then*

$$m'(\beta) = \mathrm{Var}_{\rho_\beta}(V), \qquad D'(\beta) = \beta\,\mathrm{Var}_{\rho_\beta}(V).$$

Whenever the variance is positive,

$$\boxed{\frac{dm}{dD} = \frac{1}{\beta} = \tau.}$$

Technical proof — optional on a first reading. Differentiate the log-partition representation of m and the identity $D = \beta m - \psi(\beta)$. □

46.1 Selection across reference cases

Let

$$\widehat{\mu}(i, dx) = w_i\mu_i(dx), \qquad w_i > 0, \qquad \sum_i w_i = 1,$$

be a reference mixture with observed case label i, and let each $V_i(x)$ be fixed in time. Write $\bar{V}_i = \mathbb{E}_{\mu_i}[V_i]$ and $\bar{V} = \sum_i w_i\bar{V}_i$.

Theorem 46.4 (Between-case and within-case selection). *The joint replicator equation decomposes into*

$$\boxed{\dot{w}_i = \eta w_i(\bar{V}_i - \bar{V})}$$

and

$$\boxed{\partial_t\mu_i(dx) = \eta\big(V_i(x) - \bar{V}_i\big)\mu_i(dx).}$$

Moreover,

$$\boxed{\frac{d\bar{V}}{dt} = \eta\,\mathrm{Var}_w(\bar{V}_i) + \eta\sum_i w_i\,\mathrm{Var}_{\mu_i}(V_i).}$$

Technical proof — optional on a first reading. Apply the replicator equation to the joint law on (i, x), integrate within each case to obtain the weight equation, and divide by w_i to obtain the conditional equation. The last identity is the law of total variance. □

For a binary hierarchy with leaves $X_1, \ldots, X_N$ and internal nodes v joining child blocks L_v and R_v, repeated use of

$$\mathrm{TC}(A, B) = \mathrm{TC}(A) + \mathrm{TC}(B) + I(A; B)$$

gives

$$\boxed{\mathrm{TC}(X_1, \ldots, X_N) = \sum_{v\in\mathrm{Int}(T)} I(X_{L_v}; X_{R_v}).}$$

If ρ^* maximises $\mathbb{E}_\rho[V] - \tau D_{\mathrm{KL}}(\rho\|\bigotimes_i \mu_i)$, then

$$\boxed{\mathbb{E}_{\rho^*}[V] - \mathbb{E}_{\otimes_i\rho_i^*}[V] \geq \tau\,\mathrm{TC}_{\rho^*}(X_1, \ldots, X_N).}$$

Every retained layer must earn the information required to coordinate its child modules.

CORE

47 Adaptive Closure Theorem

In plain language. The theorem assembles results already proved. It says which answer shapes, state laws, path laws, operator products, information remainders, and support boundaries survive the declared transformations. Each clause keeps the assumptions of the theorem from which it comes.

The preceding results jointly define the closure class.

Theorem 47.1 (Adaptive Closure Theorem)**.** *Consider an inferential or adaptive problem whose permitted cases supply:*

(A1) *a full reference answer* $\mathbf{K} = (\mathcal{K}, \approx, \mathcal{U})$*;*

(A2) *for each attained state case* $\mu \in \mathcal{K}$*, an admissible potential* f_μ *and, where trajectories matter, a reference path law* Q_μ *with an admissible path functional* F_μ*;*

(A3) *declared product structures, channels, coarse-grainings, retention rules, and generation kernels whenever those operations belong to the problem.*

Then, subject to the hypotheses of the corresponding results:

(i) ***Full-answer closure.*** *Every attained case is transported, declared equivalences are preserved, and empty, nonnormalisable, tied, or unattained cases remain visible. A family is replaced by a point or portfolio only when the problem supplies a selector or weights.*

(ii) ***State-space closure.*** *Sequential closed tilts add their potentials; independent systems factorise under additive potentials; relative to a product reference, dependence contributes mutual information or total correlation; deterministic and stochastic channels preserve the reference-tilt form with a conditional-log-moment potential.*

(iii) ***Path-space closure.*** *Path potentials compose additively; Markov references generate normalised Feynman–Kac recursion; path KL obeys sequential and endpoint decompositions; controlled diffusions price drift change through Girsanov's identity; path channels preserve the same exponential form.*

(iv) ***Positive-operator closure.*** *For a fixed transition kernel and potential, an open step is the normalised action of a positive kernel. Sequential open steps compose by positive-operator multiplication before normalisation. Time-homogeneous fixed points are normalised eigenmeasures; primitive finite-state systems converge to the principal one.*

(v) ***Information accounting.*** *Interaction and channel transport partition KL into marginal, coordination, visible, and conditionally hidden components.*

(vi) ***Support boundary.*** *Every finite state or path tilt is equivalent to its immediate reference. Support expansion is attributable to a declared generation step; representation change enlarges the problem itself.*

Technical proof — optional on a first reading. Full-answer closure applies the pointwise state and path results to every permitted reference case while preserving the problem's equivalence and boundary structure. State-space closure follows from temporal addition, product composition, interaction decompositions, and channel closure. Path-space closure follows from the path variational identity, Feynman–Kac recursion, the KL chain rules, Girsanov, and Doob transformation. Positive-operator closure follows from normalised composition. Information accounting follows from the KL chain rule. The support statement follows from equivalence under finite tilts and the separate generation operator. □

Remark 47.1 (Precision and object type). Full-answer closure is a precision-preserving lift of the state and path results. It records whether the supported answer is a point, family, equivalence class, empty case, or boundary. State laws, path laws, actions, and models specify the object being answered about; any such object can remain set-valued at the level of the full answer.

48 Conditions, Identification, and Intelligence

The theorem applies only after the problem specifies the objects on which each clause operates. The table below separates what the mathematics fixes from what must still be learned, measured, or justified in the domain.

Condition	What it fixes	What remains outside the clause
Declared reference	inherited weight and support	causal origin and calibration
Admissible potential	finite normalisation and selected law	support expansion
Answer shape	point, family, equivalence, empty case, or boundary	any selector not supplied by the problem
Product or joint structure	independence or coordination cost	omitted coupling
Declared channel	transported reference and potential	truth of the channel model
Coarse-graining	retained distinctions	low-dimensional macro autonomy
Reference path law	route information and kinetic cost	unmodelled path constraints
Generation kernel	support entry in a declared space	creation of a new language or ontology

A single observed law does not identify its reference and potential. For equivalent laws,

$$f = \log \frac{d\rho}{d\mu} + c.$$

Empirical content therefore requires at least one independently fixed or restricted component: the reference, intervention, information price, channel, generator, path law, or coarse-graining. A maintained application must also satisfy the transported restrictions across time, interaction, channels, and scale without redefining latent components after failure.

The closure class is broader than intelligence. It covers reference-relative inference, value, and selection even when no central predictive model exists. An adaptive system is intelligent when represented information about hidden or future consequences guides inference, generation, commitment, experiment choice, or revision.

A canonical systems tuple is

$$\boxed{(\mu, V, \tau, K, M, \alpha, \Gamma, Q),}$$

where μ is the reference, V value, τ the information price, K the evidence or observation channel, M transition or generation, α retention, Γ commitment, and Q the reference path law. The domain supplies their meaning, calibration, and legitimacy.

Falsification. The maintained class fails when independently fixed references and potentials violate the predicted log-ratio, covariance, coordination, path, or channel relations; when support appears without an identified enlargement; or when no stable typed decomposition survives intervention. Refitting the potential after failure defines a different model.

48.1 The typed cycle

A complete intelligent cycle keeps inference, selection, and action distinct:

$$\boxed{\begin{array}{c}\text{problem} \to \text{generation if required} \to \text{evidence-based answer} \to \text{selection,}\\ \text{commitment and observation} \to \text{revision, support expansion, or representation change.}\end{array}}$$

Generation changes the candidate problem. Evidence determines what is supported. Value, fitness, or policy determines what is selected or done. Commitment makes the action precise. Observation determines whether the next step is revision within the model, support expansion, or representation change. The closure results describe each transformation without making their roles interchangeable.

CORE

Part IX
Commitment and Revision

Information is information, not matter or energy.

Norbert Wiener, *Cybernetics*, 1948

In plain language. Learning and selection do not yet answer what happens when an unresolved model is compressed into one decision. Commitment may discard distinctions; retained information determines whether later review and revision remain possible.

$$\text{supported answer} \longrightarrow \text{committed action} \longrightarrow \text{retained information and revision.}$$

49 Commitment and Retained Information

FIRST READING The main idea is that action is a channel. It may discard distinctions that were present in belief. A retained record can preserve the distinctions needed for review and revision.

Let $(\mathcal{X}, \mathcal{F})$ be an epistemic or model state space and let $\Gamma(da \mid x)$ be a channel into an action, decision, code artefact, or institutional output. For probability laws P and Q on $\mathcal{X}$, define the information lost through commitment by

$$\Delta_\Gamma(P, Q) = D_{\mathrm{KL}}(P \| Q) - D_{\mathrm{KL}}(P\Gamma \| Q\Gamma).$$

Theorem 49.1 (Information lost through commitment)**.** *Under regular conditional laws and finite divergence,*

$$\boxed{\Delta_\Gamma(P, Q) = \mathbb{E}_{a \sim P\Gamma} D_{\mathrm{KL}}(P(dx \mid a) \| Q(dx \mid a)) \geq 0.}$$

Technical proof — optional on a first reading. Apply the KL chain rule to the joint laws $P(dx)\Gamma(da \mid x)$ and $Q(dx)\Gamma(da \mid x)$. Their action kernels agree, so the joint divergence equals $D_{\mathrm{KL}}(P\|Q)$. Disintegrating in the opposite direction gives the action divergence plus the conditional term. □

Corollary 49.2 (Post-processing cannot restore a lost distinction)**.** *For every channel H acting on the committed output,*

$$\boxed{\Delta_{\Gamma H}(P, Q) \geq \Delta_\Gamma(P, Q).}$$

Technical proof — optional on a first reading. Data processing gives

$$D_{\mathrm{KL}}(P\Gamma H \| Q\Gamma H) \leq D_{\mathrm{KL}}(P\Gamma \| Q\Gamma).$$

□

If $P\Gamma = Q\Gamma$ for distinct permitted laws, the action alone cannot identify which law preceded it. Recovery requires retained information or new evidence.

In plain language. A decision may be perfectly clear while the reasons behind it are lost. Reversible commitment means preserving enough of those reasons for the declared form of later review.

Definition 49.1 (Reversible commitment)**.** A commitment channel Γ is **reversible for an answer family** $\mathcal{P}$ when one recovery channel R satisfies

$$P\Gamma R = P \qquad \text{for every } P \in \mathcal{P}.$$

The definition concerns recovery of the relevant epistemic state; the physical action need not be reversible.

Proposition 49.3 (Exact recovery preserves KL)**.** *If Γ is reversible for $\mathcal{P}$, then for all $P, Q \in \mathcal{P}$,*

$$\boxed{D_{\mathrm{KL}}(P\Gamma \| Q\Gamma) = D_{\mathrm{KL}}(P \| Q).}$$

Technical proof — optional on a first reading. Data processing through Γ gives one inequality; data processing through the common recovery channel gives the reverse inequality. □

An augmented channel $\widetilde{\Gamma}(da, ds \mid x)$ can retain an action A together with a record S: sources, assumptions, answer shape, tests, provenance, the action rule, and revision conditions.

Definition 49.2 (ε-sufficient retained record)**.** The augmented output is **ε-sufficient in KL for** $\mathcal{P}$ when

$$\sup_{P,Q \in \mathcal{P}} \left[D_{\mathrm{KL}}(P \| Q) - D_{\mathrm{KL}}(P\widetilde{\Gamma} \| Q\widetilde{\Gamma}) \right] \leq \varepsilon,$$

over pairs with finite divergence. The case $\varepsilon = 0$ is exact sufficiency.

Theorem 49.4 (Information carried by the retained record)**.** *Let A be the action marginal of $\widetilde{\Gamma}$. If $\widetilde{\Gamma}$ is exactly sufficient for P and Q, then*

$$\boxed{\Delta_\Gamma(P,Q) = \mathbb{E}_{a\sim P\Gamma} D_{\mathrm{KL}}\Big((P\widetilde{\Gamma})_{S|A=a} \| (Q\widetilde{\Gamma})_{S|A=a}\Big)\,.}$$

Technical proof — optional on a first reading. Exact sufficiency identifies the joint (A,S) divergence with $D_{\mathrm{KL}}(P\|Q)$. Apply the chain rule and subtract the action-marginal divergence. □

50 Action Precision, Revision, and Correlated Evidence

Let $\ell(a,x)$ be a loss and let $\mathcal{P}$ be an epistemic family. Define

$$\mathcal{B}(\mathcal{P}) = \bigcap_{P\in\mathcal{P}} \arg\min_a \mathbb{E}_P[\ell(a,X)].$$

Proposition 50.1 (A common action under unresolved belief)**.** *If $\mathcal{B}(\mathcal{P}) = \{a^\star\}$, then $a^\star$ is the unique action common to the Bayes-action sets of all laws in $\mathcal{P}$. Individual laws may retain additional tied actions, and the epistemic answer remains the full family. Conversely, a point-valued law may leave several actions tied.*

Precise action and precise belief are different properties. A system can act through one common decision while preserving the uncertainty that supports it.

In plain language. A family of plausible models can support one common action. Acting decisively does not require pretending that the evidence selected one unique model.

World-facing mismatch has three forms:

(i) **revision:** the mismatch is represented within the current model and changes the current epistemic state;

(ii) **support expansion:** the mismatch is expressible in the ambient state space but absent from the immediate measure class, so a generator expands support;

(iii) **representation change:** the mismatch lies outside the current language or state space, so the problem itself is enlarged.

Let X be a latent state and let $Y_1,\ldots,Y_n$ be reports with positive joint likelihood density $k(y_{1:n} \mid x)$ at the realised tuple. Let $k_i(y_i \mid x)$ be the corresponding positive marginal likelihoods.

Proposition 50.2 (Conditional dependence correction)**.** *The combined evidence potential is*

$$f(x) = \log k(y_{1:n} \mid x).$$

If the reports are conditionally independent given X, then

$$f(x) = \sum_{i=1}^n \log k_i(y_i \mid x).$$

In general,

$$\boxed{f(x) = \sum_{i=1}^n \log k_i(y_i \mid x) + \log \frac{k(y_{1:n} \mid x)}{\prod_i k_i(y_i \mid x)}.}$$

The final term corrects for conditional dependence.

Reports sharing a source, reference, training corpus, tool output, or failure mode must be represented jointly. Agreement alone does not determine how much evidence they contain.

A complete operating cycle is

$$\boxed{\text{generate} \to \text{ground} \to \text{preserve the answer} \to \text{commit} \to \text{observe} \to \text{revise or enlarge.}}$$

Generation changes the candidate problem; evidence determines what is supported; value, fitness, or policy selects what is acted upon; commitment makes the action precise; observation determines whether the next step is revision, support expansion, or representation change.

Part X
Recoveries, Tests, and Open Problems

The purpose of computing is insight, not numbers.
Richard W. Hamming, 1962

In plain language. Bayesian inference, statistical mechanics, evolution, control, and economics supply different references, potentials, channels, and mechanisms. The common form is substantive only when an additional composition rule, information remainder, support boundary, or empirical restriction transfers with it.

$$\text{shared form} \quad + \quad \text{transferred restriction} \quad = \quad \text{substantive recovery.}$$

51 Recoveries and Transfer

Bayesian updating, Gibbs measures, replicator equations, exponential weights, Feynman–Kac flows, and KL-regularised control are established instances of the component mathematics. They become instances of the closure structure only when the domain also supplies the reference, potential, channel, or generator needed for an additional restriction to transfer.

Definition 51.1 (Transfer test)**.** An identification between domains is substantive when it preserves the roles of reference, potential, information price, channel, and generation closely enough for at least one further composition rule, information remainder, support boundary, or empirical restriction to transfer.

Domain	Reference and potential	Transferred structure
Bayesian inference	prior; log likelihood	temporal addition, posterior support inheritance, channel conditioning
Statistical mechanics	base measure; $-E/\tau$	Gibbs variational law, free energy, effective potential under coarse-graining
Evolution	population; relative fitness	replicator limit, generation–selection order, mutation–selection dynamics
Online learning	initial weights; cumulative negative loss	exponential weights, additive history, comparator support boundary
Soft control	reference policy or path law; action or path value	KL-regularised policy, path cost, driven dynamics
Economics	inherited social reference; utility or payoff	bounded choice, comparative statics, reference and value coupling
Institutions	admissibility and maintained incentives	retention, lock-in, portfolio resilience, entry as generation

51.1 Inference and learning

Bayes' rule is the tilt by log likelihood:

$$\mu(dh \mid s) = \mathcal{T}_{\log L_s}[\mu](dh).$$

Conditionally independent observations add log likelihoods. Zero prior support remains zero under ordinary conditioning, and a credal prior family yields the corresponding family of posteriors. Exponential weights has the same time algebra with potential equal to negative cumulative loss. Adding a previously absent hypothesis, expert, or policy is a generation event rather than another closed update.

51.2 Physics, evolution, and control

With $f = -E/\tau$, the selected law is Gibbs relative to the base measure and the coarse-grained potential is a conditional free energy. With $f = \eta F$, the weak-step limit is the replicator equation; mutation or recombination supplies the generation kernel and yields replicator–mutator dynamics. On path space, KL-regularised control tilts a passive path law, while Feynman–Kac recursion and Doob transformation convert global path value into local driven dynamics.

51.3 Economics and institutions

In economics, μ records inherited comparison structure, V utility or payoff, and τ the price of departure. At social levels the reference may be interpreted as doxa: expectations, salience, admissibility, and institutional normality. Interactions can enter through current value or through an inherited joint reference; aggregation carries an effective potential; reference portfolios provide a quantitative resilience bound. These are domain identifications, not parts of the formal core.

52 Empirical Tests

Because any law equivalent to a reference can be written as a tilt, a fit to one distribution is not enough. The tests below hold the reference, intervention, channel, generator, or path law fixed and then ask whether the predicted relation survives another time, scale, or intervention.

Test	Independently fixed	Prediction
Temporal accumulation	baseline reference and sequential potentials	pairwise log ratios add across closed updates
Local response	intervention h and pre-intervention law	$d\mathbb{E}[A]/d\theta = \mathrm{Cov}(A, h)$
Coordination	marginals, joint value, τ	joint value gain covers $\tau I(X;Y)$
Path control	passive path law and controlled drift	path KL equals the corresponding control energy
Macro transport	micro-reference, potential, and channel	conditional-log-moment effective potential
Support entry	pre-change support and observation process	no new standing without generation or representation change
Portfolio stress	component references and prior weights	excess KL loss bounded by $-\log w_i$ relative to each retained case

Generation–selection order supplies an additional intervention. Given an identified generator M and a pre-generation potential h, selected-then-generated evolution predicts $(\mathcal{T}_h\mu)M = \mathcal{T}_{h_M}(\mu M)$. A claimed post-generation potential f is compatible with that order only when $f - h_M$ is constant on generated support.

Together these tests overidentify the maintained decomposition: one independently fixed set of references, potentials, channels, generators, and path laws must satisfy the transported restrictions without being refitted after each intervention.

53 Open Problems

The closure results leave six concrete research problems.

1. **Restricted macro families.** Determine when exponential families, graphical models, policy classes, ecological models, or institutional descriptions remain closed or approximately closed under repeated channels and coarse-graining.

2. **Dynamic near decomposability.** Extend the static information bound to generator-specific results with spectral separation, rapid within-module equilibration, slow between-module motion, and controlled aggregation error.

3. **General positive operators.** Establish principal eigenmeasures, projective contraction rates, and perturbation stability in general state spaces.

4. **Singular regimes.** Characterise phase transitions, cascades, symmetry breaking, concentration, support loss, and hysteresis where bounded perturbation theory fails.

5. **Identification and causal abstraction.** Estimate references, channels, path laws, and representations when they remain families; distinguish macrostates that summarise from those that support autonomous intervention.

6. **Endogenous generation and institutional design.** Model resources allocated to search, mutation, experimentation, recombination, and representation change; design evidence channels, retained records, and recovery mechanisms that preserve resilience and coordination.

54 Conclusion

An adaptive system begins with inherited structure. Its reference determines which represented possibilities already have standing. Evidence changes what is supported. Value, fitness, or policy changes what is selected. Channels reveal some distinctions and hide others. Retention carries part of the result forward.

The same reference-relative form survives these changes. Closed updates add through time. Interaction has an information cost. Observation and coarse-graining preserve the form while separating visible change from what remains hidden. On trajectories, the same law accounts for both the destination and the path taken to reach it. Open steps compose as positive operators before normalisation.

The boundary is sharp. Reweighting can concentrate or suppress possibilities already represented; it cannot give standing to a state outside the current measure class. Mutation, search, invention, and model expansion therefore belong to generation. Revision changes a state within a model; representation change alters the model itself.

Action creates a second boundary. A commitment turns an unresolved model into one behaviour and may discard information in the process. A retained record preserves the evidence, assumptions, provenance, and answer shape needed for later review. Decisive action can therefore coexist with honest uncertainty.

The framework becomes empirical when its reference, potential, channel, generator, path law, and resolution are fixed independently of the outcome. Those inputs can then be tested through the relations the theory transports across time, interaction, paths, and scale.

Intelligence is the disciplined movement among these operations: learning from evidence, selecting under value, acting with the required precision, and reopening the problem when the world changes.

References

The Preceding Foundations

E. Mostaque. *Intelligent Economics.* Intelligent Internet, 2026.

E. Mostaque. *Intelligent Epistemology: MU and Epistemic Zero.* Intelligent Internet, 2026.

Systems, Cybernetics, Architecture, and Scale

P. W. Anderson. More is different. *Science*, 177(4047):393–396, 1972.

W. R. Ashby. *Design for a Brain: The Origin of Adaptive Behaviour.* Chapman & Hall, London, 1952.

W. R. Ashby. *An Introduction to Cybernetics.* Chapman & Hall, London, 1956.

P. Bak, C. Tang, and K. Wiesenfeld. Self-organized criticality: An explanation of $1/f$ noise. *Physical Review Letters*, 59(4):381–384, 1987.

A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

L. von Bertalanffy. *General System Theory: Foundations, Development, Applications.* George Braziller, New York, 1968.

R. C. Conant and W. R. Ashby. Every good regulator of a system must be a model of that system. *International Journal of Systems Science*, 1(2):89–97, 1970.

G. A. Cowan. The founding of the Santa Fe Institute. *Complexity*, 1(3):9–13, 1995.

H. Haken. *Synergetics: An Introduction.* Springer, Berlin, 1977.

L. P. Kadanoff. Scaling laws for Ising models near T_c. *Physics*, 2(6):263–272, 1966.

H. R. Maturana and F. J. Varela. *Autopoiesis and Cognition: The Realization of the Living.* D. Reidel, Dordrecht, 1980.

R. M. May. Will a large complex system be stable? *Nature*, 238:413–414, 1972.

G. Nicolis and I. Prigogine. *Self-Organization in Nonequilibrium Systems.* Wiley, New York, 1977.

H. A. Simon. A behavioral model of rational choice. *Quarterly Journal of Economics*, 69(1):99–118, 1955.

H. A. Simon. The architecture of complexity. *Proceedings of the American Philosophical Society*, 106(6):467–482, 1962.

H. A. Simon and A. Ando. Aggregation of variables in dynamic systems. *Econometrica*, 29(2):111–138, 1961.

A. M. Turing. The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London B*, 237:37–72, 1952.

J. von Neumann. *Theory of Self-Reproducing Automata.* Edited by A. W. Burks. University of Illinois Press, Urbana, 1966.

C. H. Waddington. *The Strategy of the Genes.* George Allen & Unwin, London, 1957.

D. J. Watts and S. H. Strogatz. Collective dynamics of small-world networks. *Nature*, 393:440–442, 1998.

W. Weaver. Science and complexity. *American Scientist*, 36(4):536–544, 1948.

N. Wiener. *Cybernetics: Or Control and Communication in the Animal and the Machine.* MIT Press, Cambridge, MA, 1948.

K. G. Wilson. Renormalization group and critical phenomena. I. Renormalization group and the Kadanoff scaling picture. *Physical Review B*, 4(9):3174–3183, 1971.

K. G. Wilson. Renormalization group and critical phenomena. II. Phase-space cell analysis of critical behavior. *Physical Review B*, 4(9):3184–3205, 1971.

Evolution, Adaptation, and Complex Adaptive Systems

C. Darwin. *On the Origin of Species by Means of Natural Selection.* John Murray, London, 1859.

M. Eigen. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, 58:465–523, 1971.

R. A. Fisher. *The Genetical Theory of Natural Selection.* Clarendon Press, Oxford, 1930.

M. Gell-Mann. *The Quark and the Jaguar: Adventures in the Simple and the Complex.* W. H. Freeman, New York, 1994.

J. Hofbauer and K. Sigmund. *Evolutionary Games and Population Dynamics.* Cambridge University Press, Cambridge, 1998.

J. H. Holland. *Adaptation in Natural and Artificial Systems.* University of Michigan Press, Ann Arbor, 1975.

J. H. Holland. Complex adaptive systems. *Daedalus*, 121(1):17–30, 1992.

J. H. Holland. *Hidden Order: How Adaptation Builds Complexity.* Addison-Wesley, Reading, MA, 1995.

S. A. Kauffman. *The Origins of Order: Self-Organization and Selection in Evolution.* Oxford University Press, New York, 1993.

C. G. Langton, editor. *Artificial Life: The Proceedings of an Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems.* Addison-Wesley, Redwood City, CA, 1989.

P. D. Taylor and L. B. Jonker. Evolutionarily stable strategies and game dynamics. *Mathematical Biosciences*, 40(1–2):145–156, 1978.

S. Wright. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. In *Proceedings of the Sixth International Congress of Genetics*, volume 1, pages 356–366, 1932.

Probability, Information, Learning, and Control

S.-i. Amari. *Information Geometry and Its Applications.* Springer, Tokyo, 2016.

N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning, and Games.* Cambridge University Press, Cambridge, 2006.

R. Chetrite and H. Touchette. Nonequilibrium Markov processes conditioned on large deviations. *Annales Henri Poincaré*, 16(9):2005–2057, 2015.

P. F. Christiano, J. Leike, T. B. Brown, M. Martic, S. Legg, and D. Amodei. Deep reinforcement learning from human preferences. In *Advances in Neural Information Processing Systems 30*, 2017.

I. Csiszár. I-divergence geometry of probability distributions and minimization problems. *The Annals of Probability*, 3(1):146–158, 1975.

P. Del Moral. *Feynman–Kac Formulae: Genealogical and Interacting Particle Systems with Applications.* Springer, New York, 2004.

M. D. Donsker and S. R. S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time, I. *Communications on Pure and Applied Mathematics*, 28(1):1–47, 1975.

J. L. Doob. Conditional Brownian motion and the boundary limits of harmonic functions. *Bulletin de la Société Mathématique de France*, 85:431–458, 1957.

H. Föllmer. An entropy approach to the time reversal of diffusion processes. In M. Métivier and E. Pardoux, editors, *Stochastic Differential Systems: Filtering and Control*, Lecture Notes in Control and Information Sciences 69, pages 156–163. Springer, Berlin, 1985.

J. W. Gibbs. *Elementary Principles in Statistical Mechanics.* Yale University Press, New Haven, 1902.

I. V. Girsanov. On transforming a certain class of stochastic processes by absolutely continuous substitution of measures. *Theory of Probability and Its Applications*, 5(3):285–301, 1960.

T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *Proceedings of the 35th International Conference on Machine Learning*, pages 1861–1870, 2018.

E. T. Jaynes. Information theory and statistical mechanics. *Physical Review*, 106(4):620–630, 1957.

M. Kac. On distributions of certain Wiener functionals. *Transactions of the American Mathematical Society*, 65(1):1–13, 1949.

J. Kivinen and M. K. Warmuth. Exponentiated gradient versus gradient descent for linear predictors. *Information and Computation*, 132(1):1–63, 1997.

S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.

C. Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete and Continuous Dynamical Systems A*, 34(4):1533–1574, 2014.

C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423 and 623–656, 1948.

J. E. Shore and R. W. Johnson. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Transactions on Information Theory*, 26(1):26–37, 1980.

E. Todorov. Linearly-solvable Markov decision problems. In *Advances in Neural Information Processing Systems 19*, 2006.

D. M. Ziegler, N. Stiennon, J. Wu, T. B. Brown, A. Radford, D. Amodei, P. Christiano, and G. Irving. Fine-tuning language models from human preferences. arXiv:1909.08593, 2019.

Economies, Computation, and Social Emergence

W. B. Arthur. Complexity and the economy. *Science*, 284(5411):107–109, 1999.

W. B. Arthur, S. N. Durlauf, and D. A. Lane, editors. *The Economy as an Evolving Complex System II*. Addison-Wesley, Reading, MA, 1997.

J. M. Epstein and R. Axtell. *Growing Artificial Societies: Social Science from the Bottom Up*. Brookings Institution Press and MIT Press, Washington, DC, 1996.

T. C. Schelling. Dynamic models of segregation. *Journal of Mathematical Sociology*, 1(2):143–186, 1971.

Positive Operators, Comparison, and Selection

G. Birkhoff. Extensions of Jentzsch's theorem. Transactions of the American Mathematical Society, 85(1):219-227, 1957.

P. J. Bushell. Hilbert's metric and positive contraction mappings in a Banach space. Archive for Rational Mechanics and Analysis, 52(4):330-338, 1973.

D. Blackwell. Equivalent comparisons of experiments. The Annals of Mathematical Statistics, 24(2):265-272, 1953.

G. R. Price. Selection and covariance. Nature, 227:520-521, 1970.

E. Seneta. Non-negative Matrices and Markov Chains. 2nd edition. Springer, New York, 2006.