UCLA/Arisaka-GG-S1-EP
v23.1 — September 2, 2026 — *Preprint*

One Functional, Three Principles

Maximum Entropy Production, Minimum Entropy Production, and the Free Energy Principle as Boundary Gauges of GDOS

Katsushi Arisaka

Department of Physics and Astronomy, Electrical and Computer Engineering, University of California, Los Angeles, Los Angeles, CA 90095, USA

Contact: arisaka@physics.ucla.edu

Abstract

Three sciences carry a variational principle at their foundations, and the three do not appear to be the same principle. Nonequilibrium thermodynamics near equilibrium has Prigogine's theorem that stationary states minimize entropy production. Climate science, plasticity theory and parts of biology use the opposite rule — that driven systems maximize it. Theoretical neuroscience holds that anything which persists minimizes a variational free energy. Each principle works somewhere; none has an accepted derivation from more fundamental physics; and the literatures that use them rarely explain why a living cell, which falls under all three at once, is not pulled in three directions.

This supplement does three things. First, it reviews each principle on its own terms — the history, the exact mathematical statement, the complete proof where one exists, and the precise point at which every attempted general derivation has failed. Second, it collects what the critical literature actually established: each principle is exact only in a restricted regime, the restrictions are of the same three kinds, and no entropy-production extremum can be exact beyond linear order.

Third, it derives all three from one object. The program's open-system boundary law, projected from the six-dimensional bulk, carries a single constrained functional; read with the path, the currents, or the beliefs held primary, it returns the least-time law, the flux–force law with the minimum and maximum principles as dual conventions, and the natural-gradient inference rule. The domains of validity are derived rather than assumed, each classical counterexample lands where the derivation says it must, and every step is graded — derived, conditional, or open.

Executive Summary

Three principles, one suspicion. Ask which way an open system goes, and three mature sciences answer with three different extremal principles. Near equilibrium, Prigogine proved in 1945 that a stationary state minimizes the rate at which entropy is produced — a genuine theorem, with a proof a page long, and with assumptions so restrictive that its own author spent the rest of his career mapping what happens beyond them. Far from equilibrium, a tradition running from Ziegler's plasticity theory to Paltridge's climate model asserts the reverse: that driven systems arrange themselves to produce entropy as fast as possible — a rule with striking successes on Earth, Mars and Titan, and no accepted derivation. And in the brain sciences, Friston's free energy principle holds that anything which persists acts so as to minimize a variational bound on the improbability of what it senses — a framework whose working mathematics is standard inference, and whose claim to be a theorem about generic matter has not survived technical scrutiny. The suspicion that these are three faces of one law is old and widely shared. It has never been written down. This supplement reviews all three — histories, proofs, failures — and then writes it down.

What the review establishes. The three literatures fail in the same three ways, and the pattern is the finding. First, each principle is exact only in a bounded regime: Prigogine's theorem holds in the strictly linear regime with constant symmetric coefficients and even variables; the one uncontested maximum theorem (Kohler–Ziman) is likewise linear; the free energy principle's inference reading holds under fine-tuned coupling conditions that generic systems violate. Second, every attempted derivation beyond the home regime has a named, published point of failure — the non-integrability of the Glansdorff–Prigogine criterion, the Gaussian step in the path-information route to maximum entropy production, the silent no-solenoidal-coupling assumption beneath the blanket construction. Third, the failures share a mechanism: what selects a nonequilibrium state depends on time-symmetric dynamical activity that no entropy-production functional contains, and on circulating currents that do no work only near equilibrium. The community's verdict, quoted and attributed throughout, is that no pure extremum of entropy production can be exact beyond linear order — and that the free energy principle, as its own defenders now frame it, is a modeling scheme rather than a physical law.

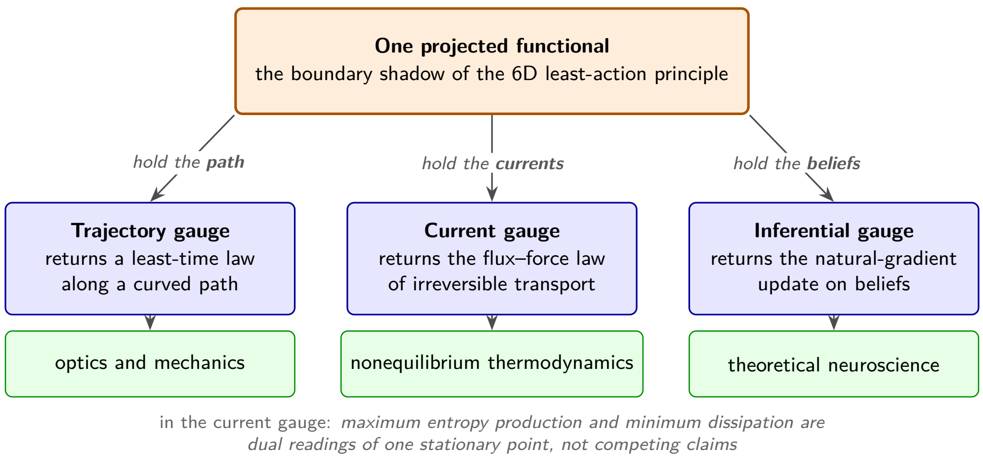
The derivation. The program's boundary law supplies what the three literatures lack: a functional that is not postulated. The six-dimensional bulk action, projected through the open-system channel established upstream, yields one constrained functional on boundary histories — a transport cost plus a scalar budget, subject to continuity. Asking for stationarity with the *path* held primary returns a least-time law: geometric optics, and mechanics in its optical form. Asking with the *currents* held primary returns the flux–force law of irreversible thermodynamics, and the minimum and maximum entropy-production principles emerge as the force-side and flux-side readings of one stationary point — dual conventions, never rivals, with Prigogine's theorem recovered exactly where its assumptions hold and the classical counterexamples located exactly where the circulating sector begins to do work. Asking with the *beliefs* held primary — coordinates on a statistical manifold, with the metric supplied by the budget's own relative-entropy term — returns the natural-gradient inference rule of the free energy principle, with the Markov blanket replaced by the declared ports of the system packet and the contested synchronization map replaced by an operational reconstruction condition. One functional, three gauges; the derivations are given in full, and each is checked against the failure cases of the principle it recovers.

What is claimed, and what is not. The derivation chain inherits the conditional status of its imports, and the boundary of the claim is drawn explicitly. What is derived: the three gauge theorems, the duality that dissolves the minimum-maximum quarrel, the location of the classical counterexamples, and the restoration of falsifiability — because a functional fixed upstream, with no adjustable landscape, can be wrong. What is conditional: the cascade closure on which the runtime stands carries one residue left open in its companion paper — a monodromy statement, since sharpened to an index-theoretic form; the budget bound and the two scope refinements once listed beside it have since been met and closed by theorems proved elsewhere in the program. What is argued rather than proved: that the three gauges exhaust the boundary readings. What is not claimed: that maximum entropy production becomes universally true (it does not, and the derivation says why), or that the expected-free-energy extension of the inference gauge follows from the functional (it does not follow in the original program either). Real observations — thermodiffusion, resistor networks, Bénard convection, planetary heat transport, ATP synthesis, chemotaxis, neural prediction — are then read through the derived gauges, each with the measurement that could refute the reading.

Why it matters. A century of extremal principles for open systems has produced working rules without foundations — each valid somewhere, none derivable, and their common source unknown. If the derivation given here is right, the three are not competing laws about nature but three coordinate choices on one projected functional, their domains of validity are theorems rather than observations, and the question that each literature could not answer from inside — *when does my principle fail?* — has an answer from outside.

Keywords: maximum entropy production; minimum entropy production; free energy principle; variational principles; Onsager reciprocity; Prigogine; Ziegler; Paltridge; Dewar; Friston; GDOS; GRIP; Qi-4; Qi-EP; GFF; open systems; falsifiability

The Three Rules Are One Thing Seen Three Ways



The whole supplement in one picture. Three sciences ask the same question — which way does an open system go? — and answer it with three different rules. Each rule works where it works: Prigogine's minimum principle is a theorem in a bounded regime, maximum entropy production is a heuristic with striking successes and no derivation, and the free energy principle is a modeling language whose own defenders have withdrawn the claim that it is a law. The suspicion that these are three faces of one law is old and weakly shared, and it had never been written down. The single functional at the top is not a new postulate — it is the boundary image of a six-dimensional least-action principle the companion papers build for reasons that have nothing to do with any of these three literatures. Each arrow is the same stationarity condition asked with a different quantity held primary. Hold the *path* and it returns the least-time law of optics and mechanics; hold the *currents* and it returns the flux–force law of irreversible transport, with maximum entropy production and minimum dissipation falling out as two readings of one stationary point rather than as rivals; hold the *beliefs* and it returns the natural-gradient update of the brain sciences. The bottom row names the community that has been using each form without knowing the other two were the same object. What is derived and what is not is marked throughout: the three gauge theorems are derived, the cascade closure they rest on carries one open residue, and maximum entropy production does not become universally true — it does not, and the derivation says why. In full: Figure 7, p. 71.

Overview of Contents

List of Figures

List of Tables

Contents

Part I

Entrance: Three Principles in Search of a Derivation

Why the derivation question is the right question, and the plan of the review

This supplement is about three rules that work without anyone knowing why.

Each of them answers the same question — *which way does an open system go?* — and each answers it with an extremum. Nonequilibrium thermodynamics, near equilibrium, says the system settles where entropy production is *least*: Prigogine's theorem of 1945, one of the few statements in this subject that is actually proved. A second tradition, running from the theory of plastic materials through the modeling of planetary climates, says driven systems arrange themselves to produce entropy *fastest* — a rule that, fed nothing but sunlight and an energy balance, reproduces the temperature contrast between the equator and the poles on three different worlds. And the brain sciences say that anything which persists behaves so as to make what it senses unsurprising — Friston's free energy principle, the most ambitious of the three, claimed at its strongest as a property of any thing that manages to keep existing.

A physicist meeting these three claims side by side notices two things at once. The first is that they cannot all be fundamental as stated: a living cell is a near-equilibrium chemical network, a driven thermodynamic engine, and a predictive controller simultaneously, and if the three principles were independent laws the cell would be receiving three sets of instructions with no account of why they do not conflict. The second is that none of the three has ever been derived. Each is either a theorem confined to a regime far narrower than its use, or a postulate whose attempted derivations have failed at published, identifiable steps. That is not a slur on the fields involved. It is the present state of knowledge, documented in their own critical literatures, and Parts II–IV of this review assemble that documentation in one place — the histories, the exact statements, the complete proofs where they exist, and the exact location of every known failure.

1 Why the derivation question is the right question

A rule that works without a derivation is not worthless — most of thermodynamics was exactly that for half a century. But it has three costs, and all three are visible in these literatures.

First, *nobody knows the domain of validity from inside.* A principle adopted as a postulate announces neither where it holds nor where it stops; its users discover the boundary by hitting it. Prigogine's minimum principle was applied far outside the linear regime for years before its own school established that nothing of it survives there. Maximum entropy production has striking successes and equally striking failures, and no user can say in advance which a new application will

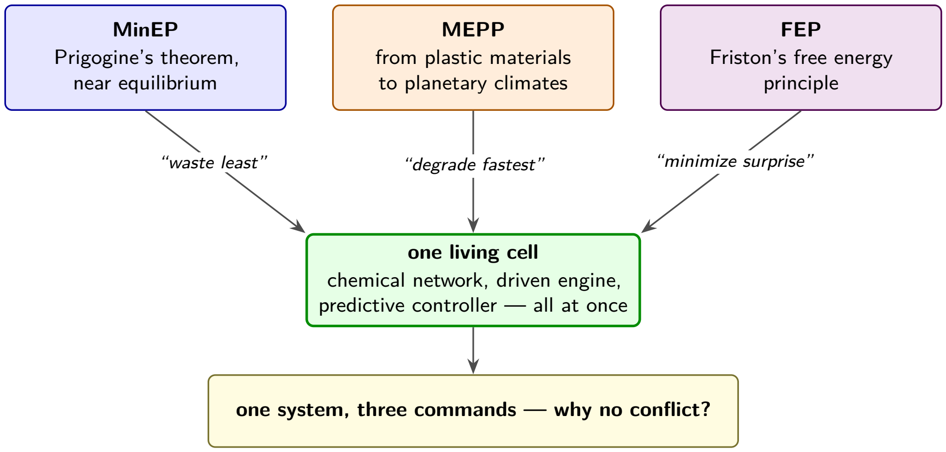


Figure 1: The question this supplement opens with, drawn. A living cell is simultaneously a near-equilibrium chemical network, a driven thermodynamic engine, and a predictive controller, and the three principles reviewed in Parts II–IV hand it three instructions at once: waste least (MinEP), degrade fastest (MEPP), and minimize surprise (FEP). Read the figure top to bottom: each principle box issues its command along its arrow, and the three commands land on one system with no account, within the three literatures, of why they do not conflict. The figure poses the supplement's question only; the resolution — one derived functional whose restricted readings the three principles are — is the business of Part VII.

be. The free energy principle's boundary is contested in the most literal sense: its critics and its defenders disagree about what would even count as crossing it.

Second, *an underived extremum invites overreading.* Each of these principles has been promoted, at times, from a working rule to a law of nature — entropy production as the objective function of the universe, surprise minimization as the imperative of everything that exists. The critical literatures reviewed below are largely the record of such promotions being withdrawn.

Third, and most practically, *underived principles cannot settle their own quarrels.* The minimum and maximum principles have been presented as contradicting each other for seventy years. The resolution — that they extremize different variables with different quantities held fixed — is known, but it is a bookkeeping observation, not a derivation; it explains why the quarrel was empty without explaining why either rule holds at all.

The question this supplement puts is therefore not *which of the three principles is right.* It is: *is there a single object, fixed by physics upstream of all three, whose different readings the three principles are?* If there is, then the domains of validity become theorems, the quarrels become coordinate disputes, and — the part that matters for a working scientist — the principles become falsifiable again, because a functional that is fixed by something else can be wrong.

2 What the program supplies

The GG-Theory program enters this story with one asset the three literatures do not have: a boundary law that is derived rather than postulated. The chain is built in the companion papers and is only used here. The constitutional charter — four axioms and seven postulates on what any physical theory must satisfy — is set out in Arisaka (2026, A1-Constitution). The six-dimensional bulk it forces, whose two extra directions are a scale and a phase rather than further places to go, is constructed in Arisaka (2026, A2-GG6) and proved free of anomaly in Arisaka (2026, A3-BI6). The law governing how that bulk looks to an observer who is part of the system — the open-system boundary law GDOS, descending from the Nakajima–Zwanzig equation through the GKLS semigroup to a mesoscopic continuity law — is derived in Arisaka (2026, A5-GDOS). The runtime that drives a system downhill and stops it, with its single scalar budget GFF, is established in Arisaka (2026, A6-GRIP). And the quantum-information pillars that make the budget's informational column exact — including the pillar Qi-EP, whose home paper states the three-gauge theorem this review deploys — are the subject of Arisaka (2026, A7-Qi4).

Part VII takes the projected functional those papers deliver and derives, in full, the three principles from it: the least-time law when the path is held primary; the flux–force law, with the minimum and maximum principles as its two dual readings, when the currents are; the natural-gradient inference rule when the beliefs are. Part VIII then reads real observations — from thermodiffusion cells to planetary heat transport to ATP synthesis to neural prediction — through the derived gauges, each with the measurement that could refute the reading. Part IX states what is new, and Part V, before any of that, states what the critical community has established — because a derivation is only worth having if it is a derivation of the right thing, failures included.

3 The register of the review, and the reader

Two disciplines govern what follows. First, the review Parts (II–IV) are written to be independently checkable and are deliberately free of the program's own vocabulary: a reader from nonequilibrium thermodynamics or computational neuroscience should be able to audit them against the primary literature without accepting, or even knowing, anything about GG-Theory. Attributions are to named papers with their venues; where a bibliographic detail could not be verified against a primary source, it is omitted rather than asserted; and no author is quoted verbatim whose text could not be read.

Second, the derivation Part carries the program's own honesty grammar. Every result is graded — derived, conditional, or open — and the grade travels with the statement. The reader will find the conditional status of the imported cascade stated where it is used, the non-exhaustiveness caveat on the gauge count stated where the count is made, and a closing section listing what this supplement does *not* establish. A review of three literatures whose central defect is overclaiming has a particular obligation not to add a fourth example.

4 Closing remarks: how to read this supplement

A word on navigation before the descent. A reader from the life sciences can take three routes through what follows. The fastest is the plain-English rail: every Part opens and closes in ordinary language, and the closings are written to carry the essentials on their own. The middle route adds Part V (the pattern behind the three failures), Part VI (the framework in brief), and the tables — especially the four-way comparison in Part IX — which are built to be read without the derivations around them. The full route is everything, and it assumes first-year graduate mathematics but nothing more: every derivation in this supplement displays its intermediate steps, and the one Part that needs tools beyond calculus teaches them on the way in.

One promise, kept throughout: nothing important lives only in an equation. Where a symbol carries the argument, a sentence nearby carries the same argument in words — so a reader who skips every display will lose precision but never the plot. And one request: read the failures as carefully as the successes. The three principles reviewed here went wrong, historically, in the hands of readers who remembered what they promised and forgot what they assumed. This supplement is arranged so that the assumptions are the hardest thing to forget.

Part II
Prigogine's Minimum Entropy Production

The one proved theorem, its complete proof, and the career-long map of its limits

Everything alive runs at a cost. A cell that swims, divides, or merely keeps its interior different from its surroundings must burn fuel to do so, and the burning shows up as waste heat given off to the world. Physicists keep the books on this waste with a quantity called entropy production: the rate at which useful, organized energy is degraded into disorganized heat that can never be fully called back. Zero entropy production means a system at perfect rest with its surroundings — equilibrium, which for an organism means death. Anything actually doing something produces entropy; the only question is how much.

This Part is about a celebrated answer to that question, proved by Ilya Prigogine in 1945 for systems that are only gently pushed. Gently pushed — the physicist says near equilibrium — means held just slightly out of rest: a cell sitting in a mild nutrient gradient, not a cell in crisis; a metal bar whose two ends differ by a fraction of a degree, not a bar in a furnace. Prigogine proved that such a gently driven system does something remarkably tidy: among all the steady ways it could arrange its internal flows, it settles into the laziest one available — the steady state that wastes the least, given the push it cannot escape.

A biological picture, offered with a warning: basal metabolism — the resting state an organism relaxes into when nothing is demanded of it, where upkeep continues but expenditure is trimmed toward the floor — has the flavor of such a minimum-dissipation state. The analogy is loose. A resting organism is still far busier, in the physicist's sense, than the theorem's assumptions allow, and nothing in this Part licenses reading the theorem as a law of biology.

That warning is, in fact, the Part's real subject. The theorem is true, and its proof takes a page. What takes the remaining pages is the boundary: where the gentle-pushing assumption ends, the theorem ends with it, sharply and provably — and most of the living world sits on the far side of that line. Knowing exactly where the line runs is worth more than the theorem itself.

5 Historical development: from least heat to least entropy production

The principle of minimum entropy production (MinEP) is often presented as a single theorem with a single author. In fact it is the last and sharpest member of a family of minimum-dissipation theorems running through a century of linear physics, and every member of the family owes its minimum property to the same mathematical circumstance — a quadratic dissipation functional generated by a symmetric, positive-definite linear response. Every historical failure of the principle can be traced to the loss of one of those ingredients. This section fixes the history with dated, verified sources; the theorem and its complete proof follow in Section 6.

5.1 Nineteenth-century precursors

Kirchhoff (1848). The earliest rigorous minimum theorem for a driven dissipative steady state belongs to Gustav Kirchhoff. In a paper on the distribution of galvanic currents in conductor networks (Kirchhoff, 1848) he established the least-heat theorem: in a network of linear (ohmic) conductors with prescribed sources, the physically realized steady currents distribute themselves so as to minimize the total Joule heat production among all distributions compatible with the constraints (charge conservation at the nodes and the imposed sources). The theorem, also expounded in Maxwell's *Treatise* (1873), is a statement about a *linear* constitutive law, and its one-line proof — a positive-definite quadratic form minimized on an affine constraint set exactly where the voltage law holds — is the ancestor of every argument in this Part.

Helmholtz (1868). Helmholtz proved the corresponding theorem for slow viscous flow: among all incompressible velocity fields taking the given boundary values, the steady Stokes flow has the least rate of viscous dissipation (Helmholtz, 1868). The hypothesis of negligible inertia is essential — the theorem lives entirely in the linear (creeping-flow) regime — and this is the pattern that recurs throughout: the minimum property is a property of *linearity*, not of nature at large. Korteweg (1883) and Rayleigh (1913) examined the theorem further.

Rayleigh (1873). For vibrating systems with linear damping, Rayleigh introduced the dissipation function — a quadratic form in the generalized velocities — and with it a principle of least dissipation for damped motion (Rayleigh, 1873), developed further in *The Theory of Sound* (1877). Onsager's variational function is modeled on it and named for it.

5.2 Onsager's principle of the least dissipation of energy (1931)

Onsager's papers on reciprocal relations (Onsager, 1931a,b) contain, alongside the symmetry $L_{ik} = L_{ki}$ derived from microscopic reversibility, a variational principle systematically confused with Prigogine's later theorem; it must be stated precisely to be distinguished from it. Let $X_1, \ldots, X_n$ be thermodynamic forces and $J_1, \ldots, J_n$ their conjugate fluxes, with entropy production $\sigma(J, X) = \sum_i J_i X_i$, and introduce the dissipation function

$$\Phi(J, J) \;=\; \tfrac{1}{2}\sum_{i,k} R_{ik}\, J_i J_k, \qquad R = L^{-1}, \quad R_{ik} = R_{ki}, \quad R > 0, \tag{5.1}$$

a symmetric, positive-definite quadratic form in the *fluxes*.

Proposition 5.1 (Onsager's least-dissipation principle, 1931)**.** *At given forces X, the physically realized fluxes maximize*

$$\mathcal{O}(J) \;=\; \sigma(J, X) - \Phi(J, J) \;=\; \sum_i J_i X_i - \tfrac{1}{2}\sum_{i,k} R_{ik} J_i J_k, \tag{5.2}$$

and the stationarity condition of $\mathcal{O}$ is precisely the linear phenomenological law $J = LX$.

Proof. Vary $\mathcal{O}$ with respect to J_j at fixed X. Using $R_{ik} = R_{ki}$,

$$\frac{\partial \mathcal{O}}{\partial J_j} \;=\; X_j - \tfrac{1}{2}\sum_k (R_{jk} + R_{kj}) J_k \;=\; X_j - \sum_k R_{jk} J_k \;=\; 0, \tag{5.3}$$

that is, $X = RJ$, equivalently $J = LX$: the Euler–Lagrange equation of Onsager's functional *is* the linear constitutive law. The Hessian is $-R$, negative definite, so the stationary point is a strict maximum; at the maximum $\sigma = X^{\mathsf{T}} LX = 2\Phi$, so $\mathcal{O}_{\max} = \Phi = \frac{1}{2}\sigma$. ■

Two remarks are essential. First, Onsager's principle varies the *fluxes at fixed forces* and returns the constitutive relation itself — it is exactly equivalent to the linear law plus reciprocity, and says nothing about which *state* a driven system selects. Prigogine's theorem, by contrast, varies the free forces at fixed clamped forces and characterizes the steady state. Second, in the path formulation of Onsager and Machlup (1953) the same functional reappears as the exponent governing fluctuation paths — the seed of the large-deviation viewpoint of Section 8.3.

5.3 Prigogine and the Brussels school (1945–1947)

Ilya Prigogine, working in Théophile De Donder's Brussels school of the thermodynamics of affinity at the Université Libre de Bruxelles, announced the theorem in a 1945 communication to the Belgian academy, "Modération et transformations irréversibles des systèmes ouverts" (Prigogine, 1945): for linear phenomenological laws with constant, symmetric coefficients, the stationary state compatible with the external constraints is the state of least entropy production, and P decreases monotonically during the approach to it. The full exposition, including the canonical thermodiffusion example (Section 6.2), is his doctoral thesis, published as *Étude thermodynamique des phénomènes irréversibles* (Prigogine, 1947) and reviewed in *Nature* in 1949 (vol. 163, p. 384); the theorem reached the English-language literature through Prigogine's textbook (*Introduction to Thermodynamics of Irreversible Processes*, 1955; second edition 1961) and the treatise of de Groot and Mazur (1962), whose Chapter V remains the careful neutral statement of the hypotheses.

The first microscopic test came quickly. Klein and Meijer (1954) analyzed the flow of matter and energy through a narrow tube connecting two vessels of ideal gas at slightly different temperatures by master-equation methods, and confirmed that the steady state coincides with the MinEP state — *to first order in the temperature difference.* Already in 1954 the verification carried the message, made precise half a century later (Section 8.3), that MinEP is a first-order-in-driving statement.

5.4 Glansdorff and Prigogine: the attempted generalization (1954–1971)

The attempt to carry MinEP beyond the linear regime occupied Prigogine and Paul Glansdorff for nearly two decades and ended in one of the more instructive negative results of twentieth-century physics. The opening move, "Sur les propriétés différentielles de la production d'entropie" (Glansdorff and Prigogine, 1954), split dP/dt into force and flux parts; the decisive result (Glansdorff and Prigogine, 1964) is that only the force part retains a definite sign, $d_X P/dt \leq 0$, and that this quantity is not the differential of any state functional. The mature statement is the monograph *Thermodynamic Theory of Structure, Stability and Fluctuations* (Glansdorff and Prigogine, 1971). The derivation and the precise sense of the failure are given in Section 7.

5.5 The Nobel arc

Out of the failed generalization came the positive program: if far-from-equilibrium steady states obey no entropy-production extremum, their stability must be studied directly — and it can be *lost*. The trimolecular scheme now called the Brusselator (Prigogine and Lefever, 1968) exhibits this cleanly: beyond a finite distance from equilibrium the thermodynamic branch destabilizes and the system selects a symmetry-broken regime — a Turing pattern or a limit cycle — that is emphatically *not* the branch of least entropy production. On the hydrodynamic side, Bénard convection sets in only at a finite critical Rayleigh number, where the MinEP characterization has already lapsed. Order sustained *by* dissipation is what Prigogine named a dissipative structure. The 1977 Nobel Prize in Chemistry was awarded to him "for his contributions to non-equilibrium thermodynamics, particularly the theory of dissipative structures"; the Nobel lecture, "Time, Structure and Fluctuations" (Prigogine, 1978), presents the linear-regime theorem and then, explicitly, the breakdown of its variational structure as the gateway to bifurcation and self-organization. It is worth stating plainly, because the point is often garbled: the prize honors, in essence, Prigogine's demonstration of the limits of his own 1945 theorem.

6 The linear-regime theorem and its complete proof

6.1 Setting

We work in the standard framework of linear irreversible thermodynamics (de Groot and Mazur, 1962). The system is in local equilibrium, so the Gibbs relation holds locally and the entropy production per unit time is a well-defined bilinear expression in conjugate flux–force pairs,

$$P \;=\; \sum_{i=1}^{n} J_i X_i \;\geq\; 0, \tag{6.1}$$

for a discrete system; for a continuous system $P = \int_V \sigma \, dV$ with $\sigma = \sum_i J_i X_i$, and the discrete argument below applies pointwise or after spatial discretization. The forces are differences or gradients of intensive parameters (for energy transport $X_q = \Delta(1/T)$ or $\nabla(1/T)$; for matter transport $X_m = -\Delta(\mu/T)$ or $-\nabla(\mu/T)$), and the constraints are imposed by reservoirs that clamp a subset of the forces.

Theorem 6.1 (Prigogine, 1945)**.** *Consider a thermodynamic system described by forces $X_1, \ldots, X_n$ and conjugate fluxes $J_1, \ldots, J_n$ with entropy production (6.1), and suppose:*

(A1) Linearity: $J_i = \sum_{k=1}^{n} L_{ik} X_k$*;*

(A2) Constancy: *the coefficients L_{ik} are constants, independent of the state variables and of time;*

(A3) Onsager reciprocity: $L_{ik} = L_{ki}$;

(A4) Positive definiteness: *the matrix L is positive definite (strict form of the second law in the linear regime);*

(A5) Constraints: *the forces $X_1, \dots, X_m$ are held fixed by external reservoirs under time-independent boundary conditions, while $X_{m+1}, \dots, X_n$ are free.*

Then the state minimizing P with respect to the free forces is exactly the stationary state, characterized by the vanishing of the fluxes conjugate to the free forces,

$$J_j \;=\; 0, \qquad j = m+1, \dots, n, \tag{6.2}$$

this extremum is a strict minimum, and $P_{\min} > 0$ whenever some clamped force is nonzero.

Proof. Under (A1) and (A2) the entropy production is the quadratic form

$$P(X) \;=\; \sum_{i,k=1}^{n} L_{ik}\, X_i X_k. \tag{6.3}$$

(i) Extremum $\Leftrightarrow$ stationarity. Differentiate with respect to a *free* force X_j ($j > m$), the clamped forces held constant. Because the L_{ik} are constants,

$$\frac{\partial P}{\partial X_j} \;=\; \sum_{k=1}^{n} (L_{jk} + L_{kj}) X_k \;\stackrel{\text{(A3)}}{=}\; 2 \sum_{k=1}^{n} L_{jk} X_k \;=\; 2\, J_j. \tag{6.4}$$

Onsager symmetry is *essential* at this step: without (A3) the gradient of P is $(L + L^{\mathsf{T}})X$, which is not proportional to the physical flux vector $J = LX$, and the extremum of P does not coincide with the stationary state. (This is the door through which counterexamples with antisymmetric — Casimir — couplings, magnetic fields, or variables of odd time-reversal parity enter; see Section 8.)

Setting $\partial P/\partial X_j = 0$ for all $j = m+1, \dots, n$ gives $J_j = 0$ for every free index. But the vanishing of the fluxes conjugate to the unconstrained forces is precisely the definition of the stationary state under (A5): the conserved quantities transported by $J_{m+1}, \dots, J_n$ (matter, charge, $\dots$) are fed by no reservoir, so in a steady state their net flows must vanish, while the flows $J_1, \dots, J_m$ conjugate to the clamped forces persist. Conversely, in the stationary state all free-force derivatives of P vanish, so the stationary state is an extremum of P on the constraint set.

(ii) The extremum is a minimum. The Hessian with respect to the free forces is

$$\frac{\partial^2 P}{\partial X_j\, \partial X_l} \;=\; 2\, L_{jl}, \qquad j, l = m+1, \dots, n, \tag{6.5}$$

twice the principal submatrix of L on the free indices. A principal submatrix of a positive-definite matrix is positive definite, so by (A4) the stationary point is a strict minimum of P on the affine subspace of fixed $X_1, \dots, X_m$. Since P is a positive-definite quadratic form in all the forces, $P_{\min} > 0$ unless every clamped force vanishes, in which case the minimizer is the equilibrium state $X = 0$

with $P = 0$. The stationary state thus appears as the natural extension of equilibrium: it minimizes, rather than annuls, the entropy production.

(iii) Dynamical (Lyapunov) version. Suppose in addition that each free force relaxes through the flux it drives,

$$\dot{X}_j \;=\; -\,c_j\,J_j, \qquad c_j > 0, \quad j = m+1, \dots, n, \tag{6.6}$$

as holds for discrete systems in which X_j is a difference of intensive parameters that decays when its conjugate transfer flux flows (the c_j are set by the capacities of the subsystems). Then, along the motion,

$$\frac{dP}{dt} \;=\; \sum_{j>m} \frac{\partial P}{\partial X_j}\,\dot{X}_j \;=\; -\,2\sum_{j>m} c_j\,J_j^2 \;\le\; 0, \tag{6.7}$$

with equality if and only if $J_j = 0$ for all free j. Hence P decreases monotonically to its minimum, the stationary state is asymptotically stable, and P is a Lyapunov function of the linear regime. ■

Remark 6.2*.* The comparison class matters: the theorem minimizes P over virtual states compatible with the constraints — free forces vary, clamped forces do not — and not over all conceivable states. Note also the contrast with Proposition 5.1: Onsager varies fluxes at fixed forces and recovers the constitutive law; Prigogine varies free forces at fixed clamped forces and recovers the steady state. The two principles are logically independent statements sharing one quadratic form.

Remark 6.3 (The severity of (A2) for continuous systems)*.* For a continuous conductor, Fourier's law in entropy-production form reads $J_q = L_{qq}\nabla(1/T)$ with $L_{qq} = \lambda T^2$, where λ is the usual thermal conductivity. Assumption (A2) therefore requires $\lambda T^2 = \text{const}$ across the sample — a physically non-generic material property. For real materials the stationary temperature field minimizes $\int \lambda\,(\nabla T)^2/T^2\,dV$ only approximately, in a neighborhood of equilibrium that shrinks as the imposed gradient grows. This apparently technical point carries much of the weight of the critiques in Section 8.

6.2 The classic worked example: thermodiffusion

Two vessels (or the two ends of a continuous bar) exchange energy and matter. The forces are the thermal force $X_q = \Delta(1/T)$ and the diffusive force $X_m = -\Delta(\mu/T)$, with conjugate fluxes J_q (energy) and J_m (matter):

$$J_q = L_{qq}X_q + L_{qm}X_m, \qquad J_m = L_{mq}X_q + L_{mm}X_m, \qquad L_{qm} = L_{mq}. \tag{6.8}$$

The thermal force X_q is clamped by two heat reservoirs; the diffusive force X_m is free. The entropy production is

$$P \;=\; L_{qq}X_q^2 + 2L_{qm}X_qX_m + L_{mm}X_m^2, \tag{6.9}$$

and, exactly as in (6.4),

$$\frac{\partial P}{\partial X_m} \;=\; 2\big(L_{mq}X_q + L_{mm}X_m\big) \;=\; 2\,J_m, \tag{6.10}$$

so $\partial P/\partial X_m = 0$ if and only if $J_m = 0$: the stationary state carries heat but *no net matter flow*, sustained by the induced composition difference

$$X_m^{\mathrm{st}} \;=\; -\frac{L_{mq}}{L_{mm}}\,X_q, \tag{6.11}$$

which is the Soret effect (thermal diffusion) in a mixture or, for Knudsen coupling through a narrow orifice, the thermomolecular pressure difference. The minimum entropy production is

$$P_{\min} \;=\; \Big(L_{qq} - \frac{L_{qm}^2}{L_{mm}}\Big)\,X_q^2 \;>\; 0, \tag{6.12}$$

positive by the positive definiteness of L ($L_{qq}L_{mm} > L_{qm}^2$): the free process arranges itself so as to shave the cross term off the dissipation.

6.3 The Klein–Meijer verification and its moral

Klein and Meijer (1954) supplied the first microscopic test of Theorem 6.1: two vessels of ideal gas at slightly different temperatures, coupled by a narrow tube, treated by a master equation for the exchange of particles and energy — the Knudsen realization of Section 6.2. The computed steady state agrees with the MinEP state, but only to leading (linear) order in the temperature difference. The verification is thus simultaneously the earliest quantitative delimitation of the theorem: what is confirmed microscopically is the first-order shadow of the underlying kinetics, a formulation the stochastic analysis of Section 8.3 makes exact.

7 The failed generalization: the evolution criterion and the birth of dissipative structures

7.1 The split of dP/dt

For a continuous system with $P = \int_V \sum_i J_i X_i\, dV$, differentiate under the integral sign at fixed volume:

$$\frac{dP}{dt} \;=\; \underbrace{\int_V \sum_i J_i\,\frac{\partial X_i}{\partial t}\,dV}_{d_X P/dt} \;+\; \underbrace{\int_V \sum_i X_i\,\frac{\partial J_i}{\partial t}\,dV}_{d_J P/dt}, \tag{7.1}$$

the decomposition of Glansdorff and Prigogine (1954). In the strict linear regime — (A1)–(A3) with constant symmetric L — the two parts are *equal*:

$$\frac{d_J P}{dt} = \int_V \sum_{i,k} X_i\, L_{ik}\frac{\partial X_k}{\partial t}\,dV \overset{L_{ik}=L_{ki}}{=} \int_V \sum_k \Big(\sum_i L_{ki}X_i\Big)\frac{\partial X_k}{\partial t}\,dV = \int_V \sum_k J_k\,\frac{\partial X_k}{\partial t}\,dV = \frac{d_X P}{dt}, \tag{7.2}$$

so that $dP/dt = 2\,d_X P/dt \le 0$ and the 1945 theorem is recovered, each half of the derivative separately non-positive.

7.2 The general evolution criterion, and why it is not a variational principle

Beyond the linear regime the two halves of (7.1) part company. The result of Glansdorff and Prigogine (1964) is that, under time-independent boundary conditions, local equilibrium, and mechanical equilibrium (no convection), but *without* linearity, constancy of the coefficients, or reciprocity, the force part alone retains a definite sign:

$$\frac{d_X P}{dt} \;=\; \int_V \sum_i J_i \, \frac{\partial X_i}{\partial t} \, dV \;\leq\; 0, \tag{7.3}$$

with equality at the stationary state — the *general (universal) evolution criterion*. The flux part $d_J P/dt$ has no definite sign, and therefore neither does dP/dt itself: far from equilibrium the total entropy production may perfectly well *increase* along the approach to the steady state.

The deeper disappointment — fully acknowledged by the authors (Glansdorff and Prigogine, 1964, 1971) — is that $d_X P$ is *not an exact differential.* In the linear regime, (7.2) gives $d_X P = \frac{1}{2}\, dP$: the criterion is the differential of the state functional $P/2$ and integrates to the MinEP theorem. Outside the linear regime there exists, in general, *no* functional Ψ of the macroscopic state with $\delta\Psi = d_X P$; the inequality (7.3) is a genuine evolution inequality but not the descent statement of any potential. Nothing is minimized. The "local potential" methods of Glansdorff and Prigogine (1971) are an attempt to construct approximate functionals around the criterion, succeeding only case by case.

7.3 From the missing potential to dissipative structures

The absence of a governing potential converts the study of far-from-equilibrium steady states from variational calculus into stability theory, and stability, unlike a minimum principle, can fail. In Bénard convection the conducting branch loses stability only at a finite critical Rayleigh number, and the convecting pattern that replaces it is not selected by any least-dissipation rule. In chemistry the Brusselator (Prigogine and Lefever, 1968) shows the same structure: beyond a critical affinity the homogeneous steady state destabilizes to Turing patterns or limit cycles, and the selected regime is not the branch of least entropy production. These are the dissipative structures of Glansdorff and Prigogine (1971) and of the Nobel lecture (Prigogine, 1978): macroscopic order maintained by the very dissipation that the near-equilibrium theorem minimized.

8 Critiques, counterexamples, and the exact scope

8.1 Landauer 1975: the blowtorch theorem and the RL circuit

The sharpest structural critique is due to Rolf Landauer. In Landauer (1975a) he considered a bistable system — a particle in a double-well potential, or any system with two locally stable macroscopic states — whose surroundings are heated *inhomogeneously*: a "blowtorch" warms a stretch of the transition path between the wells, in a region where the stationary occupation is vanishingly small. The relative occupation of the two stable states in the resulting steady state depends on this local heating, that is, on kinetic details along the path far from either state. Landauer's conclusion, stated carefully: no functional constructed from the entropy and its derivatives — entropy itself, entropy production, or the state-derivatives of either — can characterize the steady state of such a system, because two systems identical in every such functional can differ in their steady states through kinetic properties invisible to the entropy family. This is not a large-driving effect; it is a demonstration that the steady state is a property of the full kinetics, not of thermodynamic bookkeeping.

In the companion paper Landauer (1975b) the failure is exhibited even *arbitrarily close to equilibrium*: for a driven circuit containing an inductor, the fluctuating steady current does not minimize the entropy production. The modern diagnosis (Bruers et al., 2007) is time-reversal parity: the current through an inductor is *odd* under time reversal, whereas Theorem 6.1 tacitly assumes all relevant variables even (voltages, charges, concentrations, temperatures). For odd variables the roles of the entropic and kinetic terms interchange, and one obtains, within the same linear regime, a *maximum*-type characterization for RL circuits, a minimum for RC circuits, and coupled mixed principles in general — so that neither "MinEP" nor "MaxEP," as a universal slogan, survives even in linear electrical networks.

8.2 Jaynes 1980: the scope critique and the hot-and-cold network

The most influential assessment of scope is the critical review of Jaynes (1980). Jaynes reconstructs the genealogy of Section 5 and insists on the distinction of Remark 6.2: Onsager's least-dissipation principle is a flux-side encoding of the constitutive law, Prigogine's theorem a force-side characterization of the steady state, and conflating them has generated much confusion.

His central counterexample is disarmingly elementary. Take a network of linear resistors in which different resistors are held at *different* temperatures T_k. The physical steady state distributes the currents according to Kirchhoff's laws, which — by the 1848 theorem — minimize the total *heat* production

$$Q \;=\; \sum_k i_k^2 R_k \tag{8.1}$$

subject to the constraints. A minimum *entropy* production principle would instead minimize

$$P \;=\; \sum_k \frac{i_k^2 R_k}{T_k}, \tag{8.2}$$

which weights each resistor by $1/T_k$ and therefore predicts a *different* current distribution — one in which the currents partially avoid the cold resistors, where a unit of Joule heat buys more entropy. The two prescriptions coincide only when all T_k are equal, that is, in an infinitesimal neighborhood of global thermal equilibrium; since the actual currents obey Kirchhoff, MinEP gives the wrong steady state whenever the temperature is nonuniform. In the terms of Theorem 6.1, nonuniform temperature makes the coefficients of the entropy-production quadratic form state-dependent across the system, violating (A2) exactly as in Remark 6.3. Jaynes's verdict on scope, paraphrased: where the principle is exact it adds nothing beyond the linear constitutive laws that already imply it, and where it would add something, it is wrong. (The original article is consulted here through its bibliographic record and through the later literature that reanalyzes its example (Bruers et al., 2007, Maes and Netočný, 2013); we accordingly paraphrase and do not quote.)

8.3 Stochastic thermodynamics: MinEP as the first order of a fluctuation functional

The master-equation setting makes the approximate character of MinEP a theorem rather than a grievance. For a Markov jump process with states x, rates $k(x, y)$ obeying local detailed balance, and distribution ρ, the mean entropy production rate is the Schnakenberg functional (Schnakenberg, 1976)

$$\sigma(\rho) \;=\; \sum_{x,y} \rho(x)\, k(x,y)\, \log \frac{\rho(x)\, k(x,y)}{\rho(y)\, k(y,x)} \;\geq\; 0. \tag{8.3}$$

Prigogine's question becomes: does the stationary distribution ρ_s minimize $\sigma(\rho)$? In general it does not. The precise positive statement was proved by Maes and Netočný (2007): consider the Donsker–Varadhan large-deviation rate functional $I(\mu)$ governing the fluctuations of the *empirical occupation times* — the exponential cost for the fraction of time spent in each state to look like μ instead of ρ_s. For dynamics obtained by a small deformation of a detailed-balance dynamics, to leading order in the deformation,

$$I(\mu) \;\simeq\; \tfrac{1}{4}\,[\sigma(\mu) - \sigma(\rho_s)], \tag{8.4}$$

so minimizing the entropy production over distributions reproduces the true stationary distribution *up to first order* in the driving. MinEP is thereby derived — and simultaneously bounded — as a consequence of the structure of dynamical fluctuations. The obstruction at higher order is structural: the rate functional always decomposes into a time-antisymmetric (entropic) part and a time-symmetric part — the *frenetic* contribution, measuring traffic and dynamical activity — and only near detailed balance is the frenetic part enslaved to the entropic one. Away from equilibrium the frenetic contribution enters independently, and no functional built from entropy production alone, minimized or maximized, determines the steady state (Maes and Netočný, 2007, 2013). Landauer's blowtorch is the canonical illustration: the blowtorch alters the frenetic factors along the path while the entropic bookkeeping is blind to them. The parity analysis of Bruers et al. (2007) completes the linear-regime picture, deriving MinEP for even variables and its maximum-type dual for odd variables from one and the same fluctuation functional.

8.4 Macroscopic fluctuation theory: the verdict from hydrodynamics

Macroscopic fluctuation theory (MFT) reaches the same verdict from the hydrodynamic side. For driven diffusive systems the central object is the quasi-potential $V(\rho)$ — the minimal large-deviation cost to drive the system from the stationary profile to the profile ρ — which solves an infinite-dimensional Hamilton–Jacobi equation and whose minimizer *is* the stationary state. Bertini et al. (2004) generalized the Onsager–Machlup path picture (Onsager and Machlup, 1953) to genuinely nonequilibrium stationary states, with the entropy production appearing as the cost function of an optimal-control problem for the macroscopic dynamics; the comprehensive account is Bertini et al. (2015). Within MFT, MinEP is not an axiom and is explicitly noted to be of limited validity: near equilibrium the quasi-potential reduces to the excess-dissipation functional and Theorem 6.1 reappears as its quadratic shadow, while far from equilibrium the quasi-potential is a nonlocal functional bearing no relation to the entropy production. The Markov-chain and hydrodynamic analyses thus converge on one statement of scope: Prigogine's principle is the universal *first order* of nonequilibrium variational structure, and nothing more.

8.5 Assumptions against counterexamples

Every classical counterexample violates an identifiable hypothesis of Theorem 6.1; Table 1 performs the mapping. Two hypotheses do the bulk of the work: constancy of the coefficients (A2), broken by any nonuniform-temperature or state-dependent-conductivity configuration; and the tacit evenness of the variables under time reversal, listed as (A6) since Prigogine's own formulation never states it — its necessity was discovered through Landauer's circuits.

8.6 The exact scope, stated once

Positively: under (A1)–(A5) of Theorem 6.1, with all variables even under time reversal, the minimum entropy production principle is an exact theorem, equivalent to the statement that the fluxes conjugate to the unconstrained forces vanish in the steady state, with the entropy production a Lyapunov function of the relaxation. *Negatively:* outside those hypotheses it holds at most to first order in the distance from equilibrium — as the leading term of the occupation-time rate functional (8.4) and of the MFT quasi-potential — and each classical counterexample marks the loss of one enumerated hypothesis, as cataloged in Table 1. MinEP is thus neither a fourth law of thermodynamics nor a discredited slogan: it is the universal quadratic shadow, near equilibrium, of variational structures that exist in full generality only at the level of fluctuations. How this scope boundary is read within the present program is the subject of Part VII.

9 Closing remarks: what a life scientist should take from this Part

Four things are worth carrying away.

First, the theorem is real. Within its stated hypotheses — gentle driving, linear responses with fixed coefficients, the reciprocity symmetry, variables indifferent to the direction of time's arrow — minimum entropy production is not a heuristic or a tendency but an exact mathematical fact, as solid as anything in physics. But its passport is stamped for a small territory. Each hypothesis is doing load-bearing work, and where any one of them fails, the theorem simply stops. It is not weakened there, or approximately true there; it has nothing to say there.

Second, the counterexamples that fill the later sections of this Part are boundary markers, not scandals. A circuit containing a coil, a network of warm and cold resistors, a chemical mixture pushed hard enough to organize itself into spirals and stripes: each fails the principle by violating one identifiable assumption, and the table in this Part maps every known failure onto the assumption it breaks. A theorem whose exceptions can all be cataloged this way is not discredited; it is delimited, which is the healthiest condition a scientific statement can be in. The people who found

Table 1: Mapping of the classical counterexamples and failure modes of the minimum entropy production principle onto the specific hypothesis of Theorem 6.1 that each one violates. The assumptions are labeled as in the theorem, with (A6) denoting the tacit requirement that all relevant state variables be even under time reversal. Read each row as: the named system fails MinEP, the failure occurs at the indicated distance from equilibrium, and the listed hypothesis is the one whose removal explains the failure. The takeaway is that the principle has no anomalous exceptions — every known breakdown is the loss of an enumerated hypothesis, and conversely, within (A1)–(A6) the theorem is exact.

Counterexample / failure mode	Source	Regime of failure	Hypothesis violated
Network of resistors at different temperatures: currents minimize heat $\sum i^2 R$, not entropy production $\sum i^2 R/T$	Jaynes (1980)	Any nonuniform T	(A2) constancy of L across the system
Fourier conduction with realistic $\lambda(T)$: stationary field not the P-minimizer	Jaynes (1980); Remark 6.3	Any finite gradient	(A2) constancy of L (λT^2 not constant)
Driven RL circuit: steady current fails to minimize entropy production	Landauer (1975b)	Arbitrarily close to equilibrium	(A6) evenness under time reversal (current is odd)
Blowtorch: bistable system with local heating on the transition path; steady occupations depend on kinetics invisible to the entropy family	Landauer (1975a)	Any driving; structural	Outside the framework: steady state not a functional of entropy production at all (frenetic/kinetic data required)
Magnetic field or Coriolis coupling: antisymmetric (Casimir) part of the response	standard; see (6.4)	Linear regime	(A3) Onsager symmetry
Nonlinear chemical kinetics; Brusselator beyond critical affinity	Glansdorff and Prigogine (1971), Prigogine and Lefever (1968)	Finite distance from equilibrium	(A1) linearity of the phenomenological laws
Bénard convection beyond the critical Rayleigh number	Glansdorff and Prigogine (1971)	Finite distance from equilibrium	(A1) linearity; mechanical-equilibrium hypothesis of (7.3)
Master-equation steady states beyond first order in the driving	Klein and Meijer (1954), Maes and Netočný (2007)	Second order in the driving	(A1)/(A2) jointly: MinEP is the first-order truncation of (8.4)

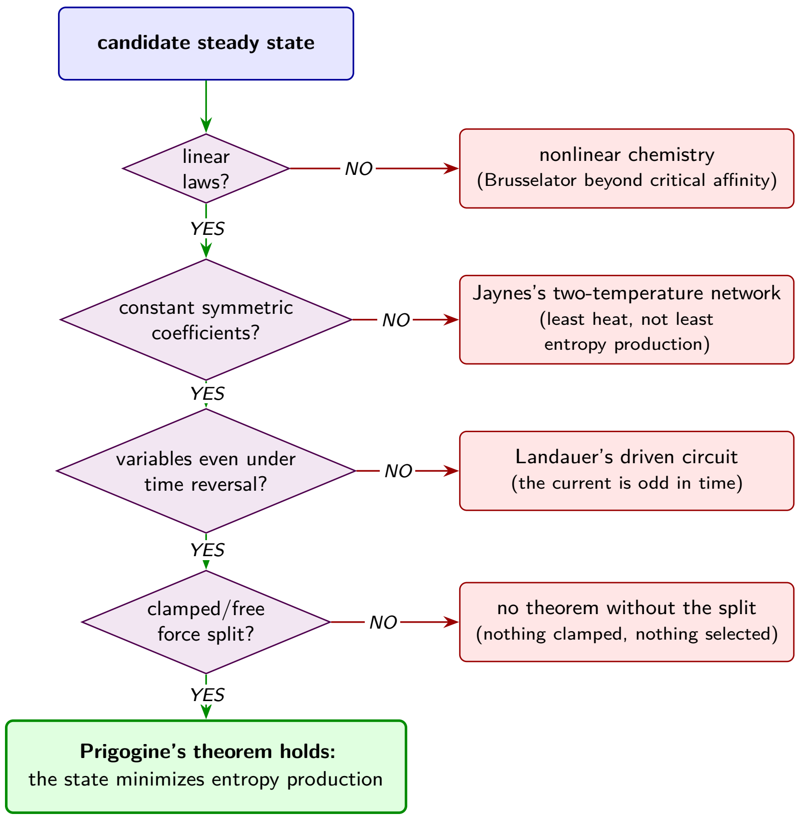


Figure 2: The exact scope of Prigogine's theorem, drawn as a decision chart. A candidate steady state descends the spine through the four load-bearing hypotheses of Theorem 6.1: strictly linear phenomenological laws, constant coefficients with Onsager symmetry, state variables even under time reversal, and the clamped/free split of the forces. Only a state that answers yes four times reaches the conclusion at the bottom, where minimum entropy production is an exact theorem. Each no exits right to the classical counterexample that lives at that door: nonlinear chemistry, Jaynes's two-temperature network, Landauer's driven circuit, and the absence of any theorem when nothing is clamped. The theorem is exact inside its passport and silent outside it; the exits are boundary markers, not scandals, and Table 1 performs the same mapping in full.

the counterexamples — including, notably, Prigogine himself — were surveying the border, not demolishing the house.

Third, and most directly for the life sciences: “living systems minimize entropy production” is wrong as a slogan. A growing cell is not gently pushed. It is far from equilibrium in exactly the sense the theorem’s hypotheses exclude — strongly driven, strongly nonlinear, with its response coefficients shifting as its own state changes. About such a system the theorem is silent, and silence is not endorsement. Appeals to Prigogine on behalf of a general biological economy of dissipation claim support from a theorem that stops at the door. Indeed the mature lesson of the field runs the other way: far from equilibrium, sustained order is often bought by dissipating *more*, not less, and no single bookkeeping quantity — neither entropy production minimized nor entropy production maximized — selects the state.

Fourth, the theorem still matters, and not merely as history. Near equilibrium it is exact, which means that any deeper theory of driven systems, whatever its machinery, must reproduce minimum entropy production in the gentle-driving limit — just as any theory of motion must reproduce Newton’s at low speed. The modern fluctuation theories described above pass exactly this test: they recover Prigogine’s theorem as their first-order shadow and then say, precisely, what replaces it beyond. A statement that every successor is obliged to contain is not a relic; it is a checkpoint, and it has already caught overreaching proposals from both the minimizing and the maximizing camps.

Prigogine’s theorem is a small truth with sharp edges, and the edges are the useful part.

Part III
Maximum Entropy Production

From Ziegler's orthogonality and Paltridge's climate to the failed path-information derivation

Picture a full stadium emptying after a game. There are many exits, and the crowd could, in principle, leave in countless different patterns — everyone funneling through one gate, an even spread across all of them, anything in between. What actually happens is one particular pattern, set by the pushing from behind, the bottlenecks, and the layout of the corridors. Much of nature is in the crowd's position. A planet heated at the equator and cooled at the poles, a metal bar bent past its yield point, a microbe living off a chemical gradient: each is a driven system with many conceivable ways of letting its driving gradient run down, and each in fact settles into one. The maximum entropy production principle — MEPP for short — is a bold guess about which one: of all the flow patterns the constraints allow, the system adopts the one that degrades the driving gradient, and so produces entropy, the fastest. It is the "widest-gates-first, pushed as hard as they can bear" theory of the crowd.

The guess has two historical faces. In materials engineering it became a quiet, workmanlike rule about how metals yield and flow, where it has done honest service for decades. In climate science it became a spectacular shortcut: a formula that fits on a postcard, with no fluid dynamics in it at all, roughly predicts how much heat the atmosphere carries from equator to pole — on Earth, on Mars, and on Titan.

The honest headline, and the theme of this Part, is the tension between those successes and the absence of any derivation. Nobody has shown, from the underlying physics, that driven systems must behave this way; when a general proof was finally claimed, specialists took the argument apart, located the precise step that fails, and the original author conceded the point in print. What survives is exact near equilibrium, empirically striking on planetary scales, and unproven everywhere else. The sections that follow display all three faces in full mathematical detail; readers who prefer words to equations can pass from this introduction directly to the closing remarks at the end of the Part.

10 The maximum entropy production principle: a dated history

10.1 Least dissipation: Rayleigh and Onsager

Every modern statement of a maximum entropy production principle (MEPP, or MaxEP) descends from a nineteenth-century *minimum* principle. In 1873 Rayleigh introduced the dissipation function $F = \frac{1}{2}\sum_{ij} b_{ij}\dot{q}_i\dot{q}_j$ for damped mechanical systems and observed that the balance of applied and dissipative generalized forces can be cast as an extremum condition on F (Rayleigh, 1873). Kirchhoff had already proved, a quarter century earlier, that the currents in a network of linear conductors distribute themselves so as to minimize the total heat generated, given the injected currents (Kirchhoff, 1848); we return to that theorem, with proof, in Section 13.

The decisive step is Onsager's. In the two 1931 papers that established the reciprocal relations $L_{ij} = L_{ji}$ from microscopic reversibility (Onsager, 1931a,b), Onsager also formulated what he called, after Rayleigh, the *principle of the least dissipation of energy*: among all conceivable fluxes at given thermodynamic state, the actual fluxes render $\dot{S}(\mathbf{J}) - \Phi(\mathbf{J}, \mathbf{J})$ a maximum, where $\dot{S}$ is the entropy production and Φ a quadratic dissipation function (Section 11.1). In 1953 Onsager and Machlup promoted the same functional to the exponent of a Gaussian path measure (Onsager and Machlup, 1953), the seed of every later path-probability approach, including Dewar's. Prigogine's theorem of minimum entropy production (Prigogine, 1947) belongs to the same near-equilibrium family: in the strictly linear regime with constant symmetric coefficients and a subset of forces clamped, the stationary state minimizes the entropy production over the free forces. Jaynes later gave a sharp assessment of how narrow the valid scope of that theorem is (Jaynes, 1980).

10.2 Ziegler: orthogonality in thermomechanics and plasticity

The first explicitly *maximum* formulation is due to Hans Ziegler, who developed it between the late 1950s and early 1960s in the context of continuum thermomechanics and plasticity (Ziegler, 1958, 1961), and gave it textbook form in *An Introduction to Thermomechanics* (Ziegler, 1983). Ziegler's *orthogonality principle* posits that the dissipative force realized by a material is normal to the level surface of a prescribed dissipation function of the fluxes; equivalently, for prescribed forces the actual fluxes maximize the entropy production subject to the single constraint that the entropy production equal the power expended by the forces (Section 11.2). In rate-independent plasticity the principle reduces to the classical normality rule—plastic strain rates normal to the yield surface—which is why it is native to that field and why it long circulated there under the name "principle of maximum plastic dissipation" rather than as a claim about nonequilibrium thermodynamics at large.

10.3 Kohler and Ziman: the one uncontested theorem

In parallel, transport theory produced the single MaxEP statement that is a theorem in the full sense. Kohler (1948) and, in general form, Ziman (Ziman, 1956) showed that the stationary solution of the *linearized* Boltzmann equation is characterized variationally: of all distributions consistent with a self-consistency constraint equating two expressions for the entropy production, the true one maximizes the entropy production. Because the statement is short and its proof is three lines, we record it here.

Theorem 10.1 (Kohler–Ziman variational principle). *Let P be the linearized collision operator, symmetric and positive semidefinite on the space of deviation functions ϕ with respect to the equilibrium inner product $\langle \cdot, \cdot \rangle$, and let X be the driving (inhomogeneous) term, so that the steady transport equation reads $P\phi = X$. Then among all trial functions ψ satisfying*

$$\langle \psi, P\psi \rangle = \langle \psi, X \rangle, \tag{10.1}$$

the solution ϕ of $P\phi = X$ maximizes the entropy production $\sigma[\psi] = \langle\psi, P\psi\rangle$.

Proof. For admissible ψ, the constraint and $P\phi = X$ give $\langle\psi, P\psi\rangle = \langle\psi, P\phi\rangle$. By the Cauchy–Schwarz inequality in the P-inner product, $\langle\psi, P\phi\rangle^2 \leq \langle\psi, P\psi\rangle\langle\phi, P\phi\rangle$, hence $\langle\psi, P\psi\rangle^2 \leq \langle\psi, P\psi\rangle\langle\phi, P\phi\rangle$ and therefore $\sigma[\psi] \leq \sigma[\phi]$, with equality only when ψ is P-proportional to ϕ. ■

The content of Theorem 10.1 is strictly linear-response, and the principle is equivalent to Onsager's least dissipation; it licenses variational solution methods for transport coefficients, not any claim about states far from equilibrium. Ziman himself noted that the principle, though essentially known to Onsager, was absent from the standard monographs of the time.

10.4 Paltridge and the climate wave

The empirical career of MEPP begins with Garth Paltridge. In 1975 he closed a zonally resolved energy-balance model of the earth–atmosphere system—boxes exchanging shortwave input, longwave output, and meridional heat transport—not with fluid dynamics but with an entropy extremum, phrased in that paper as *minimum entropy exchange* with space (Paltridge, 1975). In 1978 he restated the selection rule in its now standard form: the meridional transports arrange themselves to *maximize* the entropy production associated with the internal (turbulent) heat transport (Paltridge, 1978), with a further development in 1981 (Paltridge, 1981). With no rotation rate, no diffusivity, and no equation of motion, the extremum reproduced the observed zonal-mean surface temperature, cloud cover, and meridional heat flux to a degree that even the principle's critics call impressive (Goody, 2007, Ozawa et al., 2003). Sawada, motivated by convection experiments, proposed at about the same time that maximum entropy production acts as a selection principle among nonlinear nonequilibrium steady states (Sawada, 1981).

The comparative-planetology test came with Lorenz et al. (2001): a two-box MEP model, fed only each body's absorbed solar radiation, predicts the observed equator-to-pole temperature contrasts of Earth, Mars, and Titan—three atmospheres spanning orders of magnitude in pressure and forcing—where any single diffusive scaling law tuned to Earth fails off-planet. The geophysical review of Ozawa et al. (2003) consolidated the climate work and extended the discussion to turbulent heat transport generally; the edited volume of Kleidon and Lorenz (2005) collected the field; and the survey of Martyushev and Seleznev (2006) assembled applications from physics, chemistry, and biology into the standard advocacy review. The theoretical centerpiece of this 2003–2006 wave was Dewar's pair of papers (Dewar, 2003, 2005), which claimed to derive MaxEP from Jaynes-style information theory applied to trajectories. Sections 11.5 and 12 treat the claim and its collapse in detail.

11 Exact statements, with derivations

11.1 Onsager's least dissipation

Proposition 11.1 (Least dissipation)**.** *Let the entropy production be the bilinear form $\sigma(\mathbf{J}) = \mathbf{X}\cdot\mathbf{J}$ at prescribed forces $\mathbf{X}$, and let $\Phi(\mathbf{J}) = \frac{1}{2}\sum_{ij} R_{ij}J_iJ_j$ with R symmetric and positive definite. Then* *the functional*

$$\Lambda(\mathbf{J}) \;=\; \sigma(\mathbf{J}) - \Phi(\mathbf{J}) \;=\; \mathbf{X}\cdot\mathbf{J} - \tfrac{1}{2}\,\mathbf{J}\cdot R\,\mathbf{J} \tag{11.1}$$

attains a unique maximum at the fluxes obeying the linear phenomenological law, $\mathbf{J}^* = L\mathbf{X}$ *with* $L = R^{-1}$.

Proof. Stationarity of (11.1) gives the Euler equation

$$\frac{\partial\Lambda}{\partial J_i} = X_i - \sum_j R_{ij} J_j = 0 \qquad \Longrightarrow \qquad \mathbf{J}^* = R^{-1}\mathbf{X} = L\mathbf{X}, \tag{11.2}$$

and the second variation is $-R < 0$, so the stationary point is the global maximum of a strictly concave function. ■

This is the exact content of the principle in Onsager (1931a,b), with the dissipation function inherited from Rayleigh (1873). Note what is varied and what is held fixed: the *fluxes* vary at *fixed forces*, and the quadratic form Φ—the constitutive information—is part of the data. Every dispute below about "MaxEP versus MinEP" dissolves once this bookkeeping is made explicit.

11.2 Ziegler's orthogonality principle

Definition 11.2 (Orthogonality condition)**.** Let a dissipation function $\sigma(\mathbf{J}) \geq 0$ be prescribed as a function of the fluxes alone, with the power balance $\mathbf{X}\cdot\mathbf{J} = \sigma(\mathbf{J})$ required of the actual state. Ziegler's orthogonality condition states that the force realized by the medium is normal to the level surface of σ at the actual flux,

$$X_i \;=\; \lambda\,\frac{\partial\sigma(\mathbf{J})}{\partial J_i}, \qquad \lambda \;=\; \frac{\sigma(\mathbf{J})}{\sum_k J_k\,\partial\sigma(\mathbf{J})/\partial J_k}, \tag{11.3}$$

the multiplier being fixed by contracting the first relation with $\mathbf{J}$ and imposing the power balance (Ziegler, 1958, 1961, 1983).

Proposition 11.3 (Orthogonality equals constrained maximization for homogeneous dissipation)**.** *Let σ be positively homogeneous of degree k in $\mathbf{J}$, so that Euler's relation $\sum_k J_k\,\partial\sigma/\partial J_k = k\,\sigma$ holds. Then $\mathbf{J}$ satisfies the orthogonality condition (11.3), with $\lambda = 1/k$, if and only if $\mathbf{J}$ is a stationary point of the constrained problem*

$$\max_{\mathbf{J}}\ \sigma(\mathbf{J}) \quad \textit{subject to} \quad \sigma(\mathbf{J}) = \mathbf{X}\cdot\mathbf{J}. \tag{11.4}$$

If in addition σ is strictly convex on the constraint surface, the stationary point is the maximum.

Proof. Introduce a multiplier μ for the constraint $g(\mathbf{J}) = \sigma(\mathbf{J}) - \mathbf{X}\cdot\mathbf{J} = 0$. Stationarity of $\sigma - \mu g$ gives $\nabla\sigma = \mu(\nabla\sigma - \mathbf{X})$, i.e. $\mathbf{X} = \frac{\mu-1}{\mu}\,\nabla\sigma$, which is (11.3) with $\lambda = (\mu-1)/\mu$. Contracting with $\mathbf{J}$ and using the constraint together with Euler's relation, $\sigma = \mathbf{X}\cdot\mathbf{J} = \lambda k\,\sigma$, whence $\lambda = 1/k$, which is exactly the value dictated in (11.3) by the power balance. Conversely, (11.3) plus the power balance reproduces the stationarity system. The convexity statement is the standard second-order condition for equality-constrained extrema. ■

The quadratic case. For $\sigma(\mathbf{J}) = \sum_{ij} R_{ij} J_i J_j$ with R symmetric positive definite ($k = 2$, $\lambda = \frac{1}{2}$), the orthogonality condition (11.3) reads

$$X_i = \tfrac{1}{2}\,\frac{\partial \sigma}{\partial J_i} = \sum_j R_{ij} J_j \qquad \Longrightarrow \qquad \mathbf{J} = L\mathbf{X}, \quad L = R^{-1}, \tag{11.5}$$

recovering Onsager's linear law, Proposition 11.1. The symmetry $L_{ij} = L_{ji}$ appears automatically, but one should be precise about the logical direction: it is the *assumption* that a dissipation potential $\sigma(\mathbf{J})$ exists—so that the response is a gradient, with symmetric Hessian—that encodes reciprocity, not an independent derivation of it.

Remark 11.4 (Polettini's caveat). The equivalence in Proposition 11.3 uses homogeneity twice (once in Euler's relation, once to fix λ). Polettini (2013) emphasizes that for non-homogeneous dissipation functions the orthogonality condition and the constrained maximization (11.4) are *inequivalent* statements, so that "Ziegler's principle" is not one principle but a convention-dependent family; and that beyond the linear regime no analytic dissipation function of the currents need exist at all (Section 12.4).

11.3 Paltridge's variational problem, and the two-box model in full

The n-box problem. Partition the planet into n zonal boxes. Box i absorbs shortwave radiation R_i^{in} (a function of the local cloud fraction θ_i), emits longwave radiation $R_i^{\text{out}}(T_i, \theta_i)$, and receives the convergence F_i of the meridional heat transport. The steady energy balances and the global transport constraint are

$$R_i^{\text{in}}(\theta_i) - R_i^{\text{out}}(T_i, \theta_i) + F_i = 0 \quad (i = 1, \dots, n), \qquad \sum_{i=1}^{n} F_i = 0. \tag{11.6}$$

The system (11.6) is underdetermined; Paltridge's selection rule closes it by demanding that the *material* entropy production of the horizontal transport,

$$\sigma_{\text{transp}} = \sum_{i=1}^{n} \frac{F_i}{T_i} \;\longrightarrow\; \max, \tag{11.7}$$

be maximal over the transports compatible with (11.6) (Ozawa et al., 2003, Paltridge, 1975, 1978). The 1975 paper phrased the same extremum as minimum entropy exchange with space; the 1978 paper states the maximum form.

The two-box model: setup. Let box 1 (low latitudes) absorb shortwave power I_1 per unit area and box 2 (high latitudes) absorb $I_2 < I_1$; let $F \geq 0$ denote the heat flux carried poleward from box 1 to box 2 by the atmosphere and ocean; and linearize the outgoing longwave emission of each box as $R^{\text{out}}(T) = A + BT$, with empirical constants A and $B > 0$ taken from radiation data. The steady balances (11.6) specialize to

$$I_1 - (A + BT_1) - F = 0, \tag{11.8}$$

$$I_2 - (A + BT_2) + F = 0, \tag{11.9}$$

so that each box temperature is an affine function of the single unknown F:

$$T_1(F) = \frac{I_1 - A - F}{B}, \qquad T_2(F) = \frac{I_2 - A + F}{B}. \tag{11.10}$$

Adding (11.8) and (11.9) eliminates F entirely:

$$T_1 + T_2 = \frac{I_1 + I_2 - 2A}{B} \;\equiv\; 2\bar{T}, \tag{11.11}$$

so the mean temperature $\bar{T}$ is fixed by the global radiation budget and is untouched by the transport. Subtracting the balances gives the contrast, linear in F:

$$\Delta T(F) \;\equiv\; T_1 - T_2 \;=\; \frac{\Delta I - 2F}{B}, \qquad \Delta I \equiv I_1 - I_2. \tag{11.12}$$

Two flux values are distinguished. At $F = 0$ each box sits in local radiative equilibrium with the largest possible contrast, $\Delta T_{\mathrm{rad}} = \Delta I/B$. At $F_{\mathrm{eq}} = \Delta I/2$ the contrast (11.12) closes, $T_1 = T_2 = \bar{T}$: the transport has erased the very temperature difference that drives it. The physically admissible range is $0 \le F \le F_{\mathrm{eq}}$.

Entropy production as a function of the flux alone. The heat F leaves box 1 at temperature T_1 and is delivered to box 2 at the lower temperature T_2; the entropy production of the transfer is therefore

$$\sigma(F) \;=\; F\left(\frac{1}{T_2} - \frac{1}{T_1}\right) \;=\; \frac{F\,\Delta T(F)}{T_1(F)\,T_2(F)}, \tag{11.13}$$

which is (11.7) for $n = 2$. Substituting (11.10) and using $T_1T_2 = \bar{T}^2 - \Delta T^2/4$ from (11.11) and (11.12),

$$\sigma(F) \;=\; \frac{F\,(\Delta I - 2F)}{B\left[\bar{T}^2 - (\Delta I - 2F)^2/(4B^2)\right]}. \tag{11.14}$$

The two zeros are manifest: $\sigma(0) = 0$ (finite force, no flux) and $\sigma(F_{\mathrm{eq}}) = 0$ (finite flux, no force), with $\sigma > 0$ in between. A smooth nonnegative function vanishing at both ends of the interval has an interior maximum; the question is only where.

Differentiation, step by step. From (11.10) the temperatures respond to the flux with opposite signs,

$$\frac{dT_1}{dF} = -\frac{1}{B}, \qquad \frac{dT_2}{dF} = +\frac{1}{B}, \tag{11.15}$$

so the thermodynamic force weakens monotonically as the flux grows:

$$\frac{d}{dF}\left(\frac{1}{T_2} - \frac{1}{T_1}\right) = -\frac{1}{T_2^2}\frac{dT_2}{dF} + \frac{1}{T_1^2}\frac{dT_1}{dF} = -\frac{1}{B}\left(\frac{1}{T_1^2} + \frac{1}{T_2^2}\right) \;<\; 0. \tag{11.16}$$

The product rule applied to (11.13) then gives the stationarity condition

$$\frac{d\sigma}{dF} = \underbrace{\left(\frac{1}{T_2} - \frac{1}{T_1}\right)}_{\text{marginal gain}} - \underbrace{\frac{F}{B}\left(\frac{1}{T_1^2} + \frac{1}{T_2^2}\right)}_{\text{marginal loss}} = 0. \tag{11.17}$$

The first term is the entropy gained by pushing one more unit of heat across the existing temperature difference; the second is the entropy foregone because that same unit of heat erodes the difference. At $F = 0$ the derivative equals the bare force $1/T_2 - 1/T_1 > 0$; at $F = F_{\text{eq}}$, where $T_1 = T_2 = \bar{T}$, it equals $-(F_{\text{eq}}/B)(2/\bar{T}^2) < 0$. The stationary point lies strictly inside the interval.

Explicit solution in the small-contrast regime. On a planet whose equator-to-pole contrast is small compared with its mean absolute temperature, $\Delta T \ll \bar{T}$, one may replace $T_1 T_2 \to \bar{T}^2$ and $1/T_1^2 + 1/T_2^2 \to 2/\bar{T}^2$. The entropy production (11.14) becomes an exact parabola in F,

$$\sigma(F) \simeq \frac{F\,(\Delta I - 2F)}{B\,\bar{T}^2}, \tag{11.18}$$

whose derivative can be taken directly,

$$\frac{d\sigma}{dF} \simeq \frac{\Delta I - 4F}{B\,\bar{T}^2} = 0; \tag{11.19}$$

equivalently, inserting the same approximations into (11.17) and using $1/T_2 - 1/T_1 = \Delta T/(T_1 T_2) \simeq (\Delta I - 2F)/(B\bar{T}^2)$ reduces the stationarity condition to $\Delta I - 2F - 2F = 0$. Either route yields, with every quantity now explicit,

$$F^* = \frac{\Delta I}{4}, \qquad \Delta T^* = \Delta T(F^*) = \frac{\Delta I - 2F^*}{B} = \frac{\Delta I}{2B}, \qquad \sigma_{\max} = \sigma(F^*) = \frac{(\Delta I)^2}{8B\bar{T}^2}. \tag{11.20}$$

Second-derivative check. That F^* is a maximum and not a minimum is immediate from (11.18):

$$\frac{d^2\sigma}{dF^2} \simeq -\frac{4}{B\,\bar{T}^2} < 0, \tag{11.21}$$

and for the exact expression (11.14) the boundary argument already given—σ smooth, positive inside, zero at both endpoints—forces the unique interior stationary point to be the maximum. The MEP state thus transports exactly half of the flux that would equalize the two boxes, $F^* = F_{\text{eq}}/2$, and sustains exactly half of the radiative-equilibrium temperature contrast, $\Delta T^* = \Delta T_{\text{rad}}/2$. The trade-off exhibited in (11.17)—flux grows, force collapses—is the entire mathematical content of the climate MEPP in its simplest form: a driven system whose force is consumed by its own flux has an interior optimum, and the principle bets that nature sits there.

The Earth–Mars–Titan test. Lorenz et al. (2001) applied precisely this construction to Earth, Mars, and Titan, feeding it nothing but each body's absorbed solar radiation and radiative response. The predicted heat fluxes and equator-to-pole temperature contrasts came out consistent with the

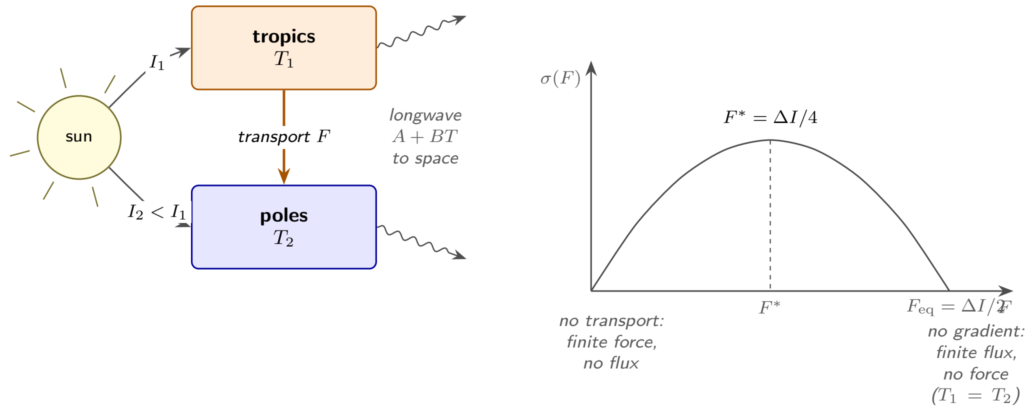


Figure 3: The two-box climate of Section 11.3, drawn. Left: shortwave input $I_1 > I_2$ heats the tropical and polar boxes, longwave emission $A + BT$ returns energy to space, and a single unknown flux F carries heat poleward. Right: the entropy production $\sigma(F)$ of that transport vanishes at both ends of the admissible interval — at $F = 0$ (finite force, no flux) and at $F_{\rm eq} = \Delta I/2$ (finite flux, no force, $T_1 = T_2$) — so it has an interior maximum, at $F^* = \Delta I/4$ in the small-contrast regime (11.20). The closure so drawn reproduces the observed contrasts of Earth, Mars, and Titan with no adjustable transport coefficient, and it remains underived; what it ignores — rotation, momentum, moist dynamics — is stated in Section 11.4.

observed ones on all three bodies—most strikingly on Titan, whose observed contrast is small despite feeble driving, exactly as a strong MEP-selected transport demands—whereas a diffusive closure with its exchange coefficient tuned to Earth fails when exported to the other two atmospheres. Three planets, one postcard-sized formula, no adjustable transport coefficient: this is the empirical core of the climate MEPP literature, and any assessment of the principle has to reckon with it.

11.4 What the closure ignores, and why the success still constrains theory

The two-box closure is silent about almost everything a dynamicist holds essential. It contains no rotation rate: a rapidly and a slowly rotating planet with the same radiation budget receive the identical prediction, although the baroclinic eddies that actually carry the heat have scales and intensities set by rotation. It contains no momentum budget, no equations of motion, no gravitational stratification, and no moist processes; Goody (2007) pressed exactly this point against the theoretical side of the program—a derivation "unconstrained, apart from energy and mass conservation," cannot know about the constraints that real atmospheres obey, and only the chaotic component of the circulation could even in principle be governed by a statistical extremum. It does not select its own functional: the maximized quantity is the material entropy production of the transport, while the planet's dominant entropy production is radiative and is not extremized (Essex, 1984). And its predictions are not invariant under refinement: what the extremum selects depends on how the system is partitioned into boxes and on which fluxes are treated as free—the resolution and constraint-choice dependence emphasized in the consensus assessment of Singh and O'Neill (2022), and conceded from within the program when Dewar (2009) reframed MEP as an inference whose output is only as good as the constraints supplied to it.

None of this erases the result; it sharpens what the result is. A parameter-free regularity that holds across three atmospheres is a fact about atmospheres, however innocent the formula that captures it. If MEPP is wrong as a law, something else must explain why the constrained extremum lands so near the observed state on bodies with utterly different dynamics; if it is right in some yet-unproven regime, the same success delimits that regime. Either way the two-box calculation worked above is a benchmark that any candidate theory of turbulent heat transport—and any future derivation or refutation of MEPP—must reproduce or explain away. The failed derivations reviewed in Section 12 discharge the theoretical claim, not the observation; the observation stands, and it disciplines both sides: critics owe an account of the success, advocates an account of the failures.

11.5 Dewar's path-information (MaxCal) construction

Dewar's papers (Dewar, 2003, 2005) attempt to derive MaxEP from Jaynes' maximum-entropy inference applied to trajectories rather than to phase-space states. The steps, as given in Dewar (2003), are these.

Step 1: path ensemble. Let Γ denote microscopic paths on $[0,\tau]$. Assign path probabilities p_Γ by maximizing the path (Shannon) entropy

$$S_I = -\sum_{\Gamma} p_\Gamma \ln p_\Gamma \tag{11.22}$$

subject to: normalization; a prescribed initial macrostate; prescribed *time-averaged boundary fluxes* $\sum_\Gamma p_\Gamma \bar{F}^n(\mathbf{x})_\Gamma = \langle \bar{F}^n(\mathbf{x})\rangle$; and local conservation of energy and mass along each path. The stationary solution is the Gibbs-type form $p_\Gamma = Z^{-1} e^{A_\Gamma}$.

Step 2: splitting the action. Integrating the conservation constraints by parts and identifying the Lagrange multiplier field with inverse temperature and chemical potentials, Dewar brings the action to the form (his equation (14))

$$A_\Gamma = \underbrace{-\frac{1}{2}\int_V \frac{H_\Gamma(0)+H_\Gamma(\tau)}{k_B T}}_{\text{even under time reversal}} + \underbrace{\frac{\tau\,\sigma_\Gamma}{2k_B}}_{\text{odd}}, \tag{11.23}$$

an end-point part symmetric under time reversal plus one half of the time-integrated entropy production σ_Γ along the path.

Step 3: the fluctuation-theorem corollary (correct). Since $\sigma_{\Gamma_R} = -\sigma_\Gamma$ for the reversed path while the end-point part is even, (11.23) gives immediately

$$\frac{p_\Gamma}{p_{\Gamma_R}} = e^{\tau\sigma_\Gamma/k_B} \qquad \Longrightarrow \qquad p(\sigma) = e^{\tau\sigma/k_B}\, p(-\sigma), \qquad \langle e^{-\tau\sigma/k_B}\rangle = 1, \tag{11.24}$$

whence $\langle\sigma\rangle \geq 0$. This part of the construction—a transient fluctuation theorem and the second law as its Jensen corollary—is uncontested.

Step 4: the contested inference of MaxEP. Writing $S_{I,\max}(\lambda) = \ln Z(\lambda) - \langle A(\lambda)\rangle$, Dewar (2003) approximates

$$S_{I,\max} \;\approx\; \ln W(\langle A^{\mathrm{irr}}\rangle), \tag{11.25}$$

where W counts paths whose irreversible action equals its ensemble mean, *and assumes that W is an increasing function of $\langle A^{\mathrm{irr}}\rangle$*. Under that assumption, selecting the macrostate of maximal path entropy selects maximal mean entropy production: MaxEP. In Dewar (2005) the analogous move is a steepest-descent (Gaussian) evaluation of $\ln Z$ about the mean fluxes, from which Ziegler's orthogonality condition (11.3) is claimed for arbitrary, including nonlinear, constitutive behavior. Both closing steps fail under scrutiny, as the next section documents; the failure is localized exactly here, in (11.25) and in the Gaussian step, not in Steps 1–3.

12 Critiques and derivation status

12.1 Grinstein and Linsker

Grinstein and Linsker (2007) identify two failures. First, the steepest-descent step of Dewar (2005) is uncontrolled: the Gaussian evaluation requires the fluxes to fluctuate weakly about their means, while the fluctuation theorem (11.24), which the derivation also invokes, requires sign-reversed fluxes to remain probable; the two demands are compatible only when the mean fluxes are small, i.e., in the near-equilibrium linear regime. The claimed far-from-equilibrium MaxEP for nonlinear constitutive relations therefore does not follow. Second, the claim of Dewar (2003) that self-organized criticality follows from MaxEP for slowly driven systems is unjustified. Dewar (2009) conceded both points, writing that the Gaussian approximation "is only valid close to equilibrium," and retreated to the position that MEP is an inference algorithm whose outputs are only as reliable as the constraints fed into it—no longer a claimed theorem of physics.

12.2 Bruers: the same machinery yields MinEP

Bruers (2007) reconstructs both derivations. For Dewar (2003) he isolates the assumption displayed in (11.25)—"the number of paths W should be an increasing function of the averaged action"—and exhibits an explicit lattice-current model in which the path count works out to $S_{I,\max}(X) \approx \ln 2^{\tau(\iota^2-\iota)} - \frac{\tau\sigma}{2}$: the path entropy is *decreasing* in the entropy production, so the same machinery, run honestly in the linear regime, delivers Prigogine's *minimum* entropy production rather than MaxEP. For Dewar (2005) he shows that consistency of the Gaussian step with the exact constitutive relations of the model forces an extra condition, $\sum_{ij} F_{ij}\,\partial A_{ij,kl}/\partial F_{lm} = 0$, which holds only when the response matrix is constant—the linear regime—so that the derivation proves at most Ziegler's principle in linear response, where it is equivalent to Onsager's least dissipation and hence not new; the appendix argument for a "total steady state MaxEP" is left unresolved by its own author's remarks. Bruers' taxonomy separates at least five inequivalent statements circulating under the one name: (i) linear-response Ziegler–Onsager MaxEP (a theorem, Section 10.3); (ii) Paltridge's *partial steady-state* MaxEP over unknown transports; (iii) a *total* steady-state MaxEP; (iv) Sawada-type non-variational selection among discrete nonlinear steady states (Sawada, 1981);

and (v) the MinEP that Dewar's 2003 setup actually implies. His verdict is that the climate successes are *not* underwritten by the path-information argument—"partial steady state MaxEP is unrelated with the principles derived" in the Dewar papers—and that MaxEP "remains an unproven hypothesis."

12.3 Climate-side critiques

Essex (1984) raised the structural objection that still stands: the overwhelmingly dominant entropy production of a planet is radiative—the degradation of solar photons—and Paltridge's functional (11.7) neither contains it nor extremizes it; the climate MEPP maximizes only the small material part, and nothing in the principle says why that part and not another. Nicolis and Nicolis (1980) questioned whether the extremum is recoverable from the underlying dynamics at all. Goody (2007) observed that Dewar's construction is "unconstrained, apart from energy and mass conservation," whereas real atmospheres obey fluid equations, rotation, and gravitational constraints that the extremum ignores; that only the chaotic component of the circulation could even in principle obey such a principle; and that comparative planetary data do not quantitatively support the unconstrained statement. The modern consensus assessment of Singh and O'Neill (2022) is that theoretical and modeling issues—among them the dependence of the selected state on the chosen constraints and on the coarse-graining, and the insensitivity of the predictions to rotation—preclude broad acceptance of MEPP in climate science. On the advocacy side, Martyushev (2010) concedes that a principle of this kind "cannot be proved" but only corroborated, any derivation from the second law requiring assumptions less obvious than the principle itself, and Martyushev and Seleznev (2014) codify the restrictions under which the authors still defend it: local formulation, fixed forces, and well-separated scales.

12.4 Polettini: constitutive versus phenomenological, and the MinEP/MaxEP resolution

Polettini (2013) supplies the cleanest available resolution of the apparent MinEP/MaxEP contradiction, together with counterexamples beyond the linear regime. His findings: (i) beyond linear response, Ziegler's principle fails—in a two-channel (laminar plus turbulent) example no analytic dissipation function of the currents satisfies orthogonality for both channels; (ii) if the forces are Taylor-expanded in the currents, MaxEP imposes generalized reciprocal relations on the higher-order response coefficients, and an explicit three-state Markov model of stochastic thermodynamics violates them; (iii) on Kirchhoff networks, maximizing the entropy production fixes coarse-grained *constitutive* relations but does not produce the steady state, whose selection is governed instead by a minimum principle. Hence the dichotomy, which deserves to be stated as the standard resolution:

Remark 12.1 (What is held fixed). *MaxEP is a constitutive principle*: the fluxes are varied at fixed forces, subject only to the power balance $\sigma(\mathbf{J}) = \mathbf{X} \cdot \mathbf{J}$; the maximization selects the response law (the transport coefficients), as in Proposition 11.3. *MinEP is a phenomenological principle*: the constitutive law is fixed (constant symmetric L), a subset of forces is clamped, and the stationary state minimizes $\sigma(\mathbf{X})$ over the free forces, whose conjugate fluxes vanish at the minimum (Jaynes, 1980, Prigogine, 1947). The two principles extremize different functionals over different variation spaces; in the linear regime both are theorems and both are re-expressions of Onsager's least dissipation, Proposition 11.1. Neither is a theorem beyond it.

13 Rigorous partial results

13.1 Kirchhoff's minimum-heat theorem and its duals

Theorem 13.1 (Minimum heat at fixed injected currents)**.** *Consider a finite network of branches k with resistances $R_k > 0$, and prescribe the external currents injected at the boundary nodes. Among all branch-current vectors $\mathbf{i}$ satisfying Kirchhoff's current law (KCL) with those injections, the physical distribution $\mathbf{i}^{\mathrm{ph}}$—the one that in addition satisfies Ohm's law and Kirchhoff's voltage law—is the unique minimizer of the dissipated power*

$$P(\mathbf{i}) = \sum_k R_k\, i_k^2. \tag{13.1}$$

Proof. The admissible set is the affine space $\mathbf{i}^{\mathrm{ph}} + \mathcal{C}$, where $\mathcal{C}$ is the cycle space: current vectors satisfying KCL with zero injections. For $\mathbf{c} \in \mathcal{C}$,

$$P(\mathbf{i}^{\mathrm{ph}} + \mathbf{c}) = P(\mathbf{i}^{\mathrm{ph}}) + 2\sum_k R_k\, i_k^{\mathrm{ph}} c_k + \sum_k R_k c_k^2. \tag{13.2}$$

By Ohm's law $R_k i_k^{\mathrm{ph}} = v_k = \phi_{a(k)} - \phi_{b(k)}$ is a potential-difference vector, so the cross term telescopes over nodes, $\sum_k v_k c_k = \sum_n \phi_n$ (net outflow of $\mathbf{c}$ at n) $= 0$, since $\mathbf{c}$ has zero injections. Hence $P(\mathbf{i}) = P(\mathbf{i}^{\mathrm{ph}}) + \sum_k R_k c_k^2 \geq P(\mathbf{i}^{\mathrm{ph}})$, with equality iff $\mathbf{c} = 0$. ■

This is the theorem of Kirchhoff (1848); in the probabilistic literature it is Thomson's principle (Doyle and Snell, 1984). The dual statement exchanges the roles of the two Kirchhoff laws.

Proposition 13.2 (Dirichlet dual: fixed boundary potentials)**.** *Prescribe instead the potentials on the boundary nodes. Among all node potential vectors $\boldsymbol{\phi}$ with those boundary values, the physical one minimizes*

$$P(\boldsymbol{\phi}) = \sum_k \frac{(\Delta\phi_k)^2}{R_k}, \tag{13.3}$$

where $\Delta\phi_k$ is the potential drop across branch k. The proof is identical in structure to that of Theorem 13.1: perturb by $\mathbf{h}$ vanishing on the boundary; the cross term $\sum_k(\Delta\phi_k^{\mathrm{ph}}/R_k)\,\Delta h_k = \sum_k i_k^{\mathrm{ph}}\Delta h_k$ telescopes to a sum over interior nodes of h_n times the net physical current there, which vanishes by KCL.

The complementary maximum at fixed EMF. Both principles above are minima; the "maximum dissipation at fixed voltage" statements in the MaxEP literature are the Legendre-complementary form of the same quadratic structure. At fixed source electromotive forces $\mathcal{E}_k$, over trial KCL-flows $\mathbf{i}$, define

$$G(\mathbf{i}) = 2\sum_k \mathcal{E}_k i_k - \sum_k R_k i_k^2. \tag{13.4}$$

Expanding about the physical flow as in (13.2), the linear term involves $\mathcal{E}_k - R_k i_k^{\mathrm{ph}}$, which is a potential-drop vector and hence orthogonal to the cycle space, so $G(\mathbf{i}^{\mathrm{ph}}+\mathbf{c})-G(\mathbf{i}^{\mathrm{ph}}) = -\sum_k R_k c_k^2 \leq 0$: the physical flow *maximizes* G, and at the maximum, by the power balance $\sum_k \mathcal{E}_k i_k^{\mathrm{ph}} = \sum_k R_k (i_k^{\mathrm{ph}})^2$, the value of G equals the dissipated power. The network MaxEP of Županović et al. (2004)—steady Kirchhoff currents maximize $\sigma(\mathbf{i}) = \sum_k R_k i_k^2/T$ subject to the constraint $\sigma = \sum_k \mathcal{E}_k i_k/T$—is exactly the constrained problem (11.4) with quadratic σ and linear constraint: it is Onsager's least dissipation, Proposition 11.1, in network clothing, not an independent far-from-equilibrium principle. The stochastic analysis of Bruers et al. (2007), using Langevin dynamics with Johnson–Nyquist noise, fixes the scope exactly: entropy-production extremum statements for circuits are the first-order shape of the dynamical fluctuation functional around equilibrium, they require the fluctuating observable to be even under time reversal, and the classical counterexamples (Landauer's, and those discussed by Jaynes, 1980) are precisely the cases violating these conditions.

13.2 Path weights, macroscopic fluctuation theory, and the frenetic obstruction

The Gaussian path measure of Onsager and Machlup (1953),

$$p[\dot{\boldsymbol{\alpha}}] \;\propto\; \exp\left\{\frac{1}{2k_B}\int(\dot{S} - \Phi - \Psi)\,dt\right\}, \tag{13.5}$$

with Φ and Ψ the flux- and force-space dissipation functions, makes least dissipation literal: the most probable path is the least-dissipation path. Macroscopic fluctuation theory (MFT) generalizes this rigorously to genuinely nonequilibrium stationary states of driven diffusive systems (Bertini et al., 2004, 2015). The hydrodynamic drift decomposes into a dissipative part, the gradient of the quasi-potential S (which solves a Hamilton–Jacobi equation), plus a non-dissipative transverse part $\mathcal{A}$; the functional

$$\mathcal{F}(\rho,\dot{\rho}) = -\dot{S}(\rho) + \big\langle(\dot{\rho} - \mathcal{A}(\rho)),\, K^{-1}(\dot{\rho} - \mathcal{A}(\rho))\big\rangle \tag{13.6}$$

is minimized by the hydrodynamic evolution, and along solutions twice the dissipation functional equals the entropy production rate. The exact extremum principle that survives far from equilibrium is therefore a *minimum-dissipation principle for a corrected functional*—not MaxEP, and not a functional of the entropy production alone.

The obstruction to any pure EP extremum is now understood structurally. For Markov dynamics near equilibrium, the large-deviation rate function for occupation statistics is, to leading order, an entropy-production difference,

$$I(\mu) \;\simeq\; \tfrac{1}{4}\big[\sigma(\mu) - \sigma(\mu_{\mathrm{st}})\big], \tag{13.7}$$

which *derives* Prigogine's MinEP as a first-order fluctuation statement (Maes and Netočný, 2007). At the next order, however, time-symmetric dynamical activity (the "frenetic" contribution, in later terminology) enters the rate function, so the stationary state is no longer determined by any functional of the entropy production alone—minimum or maximum. This is the modern, precise reason why every EP extremum principle is at best a near-equilibrium approximation, and why the far-from-equilibrium claims of MEPP could not have had a general proof of the advertised kind.

13.3 Scope summary

Table 2 condenses the status of the principle by domain.

Table 2: Scope and status of the maximum entropy production principle by domain of application. The "theorem" rows are exhaustive: outside the strictly linear regime no derivation of MaxEP from microscopic dynamics is accepted, the 2003–2005 path-information derivations having been localized to near-equilibrium by the critiques of 2007 and conceded by their author in 2009. The heuristic rows record empirical performance, not derivation; the failure rows record either explicit counterexamples or ill-posedness of the extremand itself.

Domain	Status	Reason
Linearized transport (Boltzmann equation; resistor networks)	Theorem	Kohler–Ziman principle (Theorem 10.1); network duals (Theorem 13.1); all equivalent to Onsager least dissipation.
Planetary meridional heat transport (Earth, Mars, Titan)	Successful heuristic	Two-box and zonal extremum (11.20) matches observed contrasts and fluxes (Lorenz et al., 2001, Paltridge, 1978); no accepted derivation; constraint choice untheorized.
Which entropy production to maximize	Ill-posed	Radiative production dominates the planetary budget and is not extremized (Essex, 1984); the principle does not select its own functional.
Nonlinear constitutive selection (Ziegler beyond linear response)	Fails	No analytic dissipation function need exist; higher-order reciprocal relations violated in explicit models (Polettini, 2013).
Steady-state selection in multistable nonlinear systems	Unproven	Sawada-type conjecture (Sawada, 1981); counterexamples and parity/activity obstructions (Bruers et al., 2007).
Exact far-from-equilibrium variational structure	Superseded	MFT minimum-dissipation principle for the corrected functional (13.6) (Bertini et al., 2004, 2015); frenetic contribution excludes pure EP extrema beyond linear order (Maes and Netočný, 2007).

The landscape, in one sentence: MEPP is one exact theorem (linear response), one durable geophysical heuristic (Paltridge's, sharpened by Lorenz et al., 2001), and one failed general derivation (Dewar's, broken at the step (11.25) and at the Gaussian saddle point), while the rigorous far-from-equilibrium variational structure that actually exists is a corrected minimum-dissipation principle. How the present program reads this landscape is the subject of Part VII.

14 Closing remarks: what a life scientist should take from this Part

First: a rule can work on three planets and still not be a law. What separates the two is not the size of the success but the presence of a derivation — an argument from established physics that says when the rule *must* hold and when it may not. MEPP has a scorecard, not a domain. It predicts planetary heat transport remarkably well, fails or becomes ambiguous in other settings, and contains nothing internal that says which case is which. A weather proverb that happens to work on three continents is still a proverb; it becomes meteorology only when someone shows why it works, and that is the step that has never been completed here.

Second: keep separate in your mind a bookkeeping identity and a selection rule, because MEPP trades on the ease of confusing them. That every real process produces entropy is bookkeeping — once you know the flows, the entropy production is computed from them the way a bank balance is computed from the transactions, and it is automatically true. MEPP claims something categorically stronger: that among all the flow patterns the constraints would allow, nature *picks* the one with the largest production. The ledger is true by construction; the pick is a physical hypothesis, and it is the hypothesis, not the ledger, that has no proof.

Third: the failed-derivation episode is worth knowing as a case study in how physics vets its principles. A general proof was claimed in 2003–2005. Specialists reconstructed the argument line by line, located the exact step where it fails, and built an explicit model in which the same machinery predicts the *opposite* extremum. The original author then conceded the point in print and reframed his claim as something weaker. No scandal occurred; this is the system working as intended. A principle earns its status by surviving precisely this treatment, and MEPP survived it demoted — from candidate law to hypothesis with a mapped boundary.

Finally, the reason this Part exists in a review that biologists may read: MEPP reaches the life sciences dressed in an attractive slogan — life exists because it accelerates the degradation of energy gradients; ecosystems mature toward states of maximal dissipation; evolution favors the better dissipator. Understand clearly that this is a *reading* of MEPP, not a result of it. Nothing in the verified record derives any biological regularity from entropy maximization, and every claim built on the slogan inherits the unproven status of the principle underneath it. Used as a lens — a generator of hypotheses that must then stand on their own evidence — the idea can be genuinely fruitful; used as a foundation, it is borrowed authority. A rule that predicts three planets has earned our attention; it has not yet earned the name of law.

Part IV
The Free Energy Principle

From Helmholtz's unconscious inference to Bayesian mechanics, and the technical critiques

Before the mathematics, the idea in plain terms. The free energy principle begins from an old thought about perception: the brain does not passively receive the world, it continuously guesses the world, and what we experience is the guess — a controlled hallucination, held in check by a stream of error signals reporting where guess and sensors disagree. Seeing, on this picture, is the ongoing correction of an internal model, not the delivery of a picture from the eye.

A homely contrast makes the ambition clear. An ordinary thermostat regulates by reaction: it waits for the room to drift, then pushes back. A weather-forecasting thermostat would regulate by prediction: it carries a model of the day ahead, anticipates the afternoon sun and the evening chill, and acts before the drift occurs. The free energy principle says that brains — and, in its strongest version, all living things — are regulators of the second kind: they persist by predicting their sensory input and by acting to make those predictions come true.

The "free energy" in the name is the single most common stumbling block for biologists, so let it be said explicitly: it is not the chemist's free energy. No Gibbs, no Helmholtz, no bond enthalpies, no calorimetry. It is an information-theoretic quantity — a computable, measurable stand-in for how surprising the current sensory data are under the organism's internal model. Surprise itself cannot be evaluated directly, because it would require knowing everything the model does not; the stand-in can be evaluated, and driving it down forces the model to track the world.

The honest headline, documented in both directions below: the working mathematics behind this idea is uncontested and now runs half of computational neuroscience, from models of visual cortex to computational psychiatry. The further claim — that free energy minimization is a law of nature governing every persisting thing — is contested, and the sharpest critiques have come from mathematical physics. This Part covers both halves, with the derivations that separate them.

15 Historical development: from unconscious inference to Bayesian mechanics

The free energy principle (FEP) did not appear from nowhere in 2005; it is the confluence of three older streams — perceptual epistemology, cybernetic regulation theory, and variational statistics — and its later "physics phase" grafted a fourth stream, nonequilibrium stochastic dynamics, onto the first three. Because the derivational disputes examined below turn on *which* of these streams a given claim belongs to, we give the history with dates and sources.

Perception as inference (1856–1867). Hermann von Helmholtz's *Handbuch der physiologischen Optik* (Helmholtz, 1867), completed with its third volume in 1867, introduced the doctrine of *unbewusster Schluss* — unconscious inference: percepts are conclusions drawn from sensory premises together with prior experience. This is the conceptual ancestor claimed by both predictive coding and the FEP. The connection is programmatic rather than mathematical; Helmholtz supplies no objective functional.

Regulation and the good-regulator theorem (1952–1970). W. Ross Ashby's *Design for a Brain* (Ashby, 1952) framed adaptive behavior as the maintenance of "essential variables" within physiological bounds — homeostasis realized by ultrastable dynamics. The line most often quoted in FEP papers is the good-regulator theorem of Conant and Ashby (1970): every good regulator of a system must be (i.e., must instantiate) a model of that system. The FEP's claim that persisting systems *are* models of their environments is Conant–Ashby restated in probabilistic dress.

Variational free energy enters neural modeling (1995–1998). The quantity that Friston would later denote F entered neuroscience through machine learning. The Helmholtz machine of Dayan et al. (1995) (*Neural Computation* 7(5), 889–904) trains a *recognition* density q to bound the log evidence of a *generative* model, with the wake–sleep algorithm as its learning rule; the bound is exactly the negative variational free energy. Neal and Hinton (1998) then showed that expectation–maximization is coordinate descent on a single free energy functional, so that inference (E-step) and learning (M-step) optimize one objective — the template for the FEP's unification of perception and learning.

Predictive coding (1999). Rao and Ballard (1999) (*Nature Neuroscience* 2(1), 79–87) gave the canonical hierarchical predictive-coding model of visual cortex: descending connections carry predictions, ascending connections carry prediction errors, and extra-classical receptive-field effects fall out of the error dynamics. Predictive coding is an *algorithm*; it neither requires nor implies the FEP, a point that matters when empirical support for the former is offered on behalf of the latter.

The brain-level FEP (2005–2010). Friston (2005) ("A theory of cortical responses," *Philosophical Transactions of the Royal Society B* 360(1456), 815–836) cast hierarchical cortical processing as empirical-Bayes inference minimizing free energy. Friston et al. (2006) (*Journal of Physiology—Paris* 100(1–3), 70–87) first used the phrase as a *principle* and introduced action as a second route to the minimum: the organism can change its predictions to match its sensations, or change its sensations to match its predictions. A widely read tutorial statement followed in Friston (2009), and the most-cited formulation is Friston (2010) ("The free-energy principle: a unified brain theory?" *Nature Reviews Neuroscience* 11(2), 127–138).

Active inference (2010–2017). Friston et al. (2010) (*Biological Cybernetics* 102(3), 227–260) is the founding continuous-state active-inference paper. The discrete-state formulation on partially observed Markov decision processes — Friston et al. (2017), "Active inference: a process theory" (*Neural Computation* 29(1), 1–49) — introduced the *expected* free energy G over policies and is the basis of essentially all empirical and computational active-inference work since, consolidated in the textbook of Parr et al. (2022).

The physics phase (2013–2019). Friston (2013) ("Life as we know it," *Journal of the Royal Society Interface* 10(86), 20130475) pivoted from brains to matter: a random dynamical system possessing a Markov blanket at nonequilibrium steady state (NESS) will "appear to" perform Bayesian inference about what lies beyond its blanket. The program culminated in the 148-page monograph Friston (2019) ("A free energy principle for a particular physics," arXiv:1906.10184, posted 24 June 2019, never published in journal form), which assembles the full chain examined in Section 16.6: Helmholtz decomposition of the NESS drift, the four-way "particular partition," the synchronization map, and the claim that expected internal states flow down a variational free energy gradient.

Bayesian mechanics and the path-integral reformulation (2021–2023). Under the pressure of the critiques reviewed in Section 18, the program produced its first mathematically careful treatment — Da Costa et al. (2021), "Bayesian mechanics for stationary processes" (*Proceedings of the Royal Society A* 477(2256), 20210518), rigorous in the linear-Gaussian setting — and then reformulated the principle itself. Friston et al. (2023a) ("The free energy principle made simpler but not too simple," *Physics Reports* 1024, 1–29) retreats from the NESS density factorization to a path-integral, paths-of-least-action formulation, extended in Friston et al. (2023b) (*Physics of Life Reviews* 47, 35–62). The programmatic statement of the resulting "Bayesian mechanics" is Ramstead et al. (2023) ("On Bayesian mechanics: a physics of and by beliefs," *Interface Focus* 13(3), 20220029), whose subtitle is candid about the deflationary turn: the FEP is there presented as a principle of least action *on belief spaces*, not a falsifiable law about matter.

16 The mathematics

We now separate, with derivations, the four mathematical strata of the FEP: the variational bound (textbook), the predictive-coding dynamics it induces in Gaussian models (elementary), the stationary Helmholtz decomposition (classical nonequilibrium statistical mechanics), and the particular-physics chain built on top (contested). Throughout, s denotes sensory data, ψ external or latent states, and μ internal states.

16.1 Variational free energy and the evidence bound

Definition 16.1 (Variational free energy)**.** Let $p(s, \psi)$ be a generative density and $q_\mu(\psi)$ a recognition density parameterized by internal states μ. The variational free energy is

$$F(s, \mu) \;=\; \mathbb{E}_{q_\mu}[\ln q_\mu(\psi) - \ln p(s, \psi)]\,. \tag{16.1}$$

Proposition 16.2 (Three decompositions and the surprisal bound)**.** *With* $\mathfrak{I}(s) = -\ln p(s)$ *the surprisal,*

$$F(s,\mu) = D_{\mathrm{KL}}[q_\mu(\psi) \,\|\, p(\psi \mid s)] + \mathfrak{I}(s) \tag{16.2}$$

$$= \underbrace{\mathbb{E}_{q_\mu}[-\ln p(s,\psi)]}_{\textit{energy}} - \underbrace{H[q_\mu]}_{\textit{entropy}} \tag{16.3}$$

$$= \underbrace{D_{\mathrm{KL}}[q_\mu(\psi) \,\|\, p(\psi)]}_{\textit{complexity}} - \underbrace{\mathbb{E}_{q_\mu}[\ln p(s \mid \psi)]}_{\textit{accuracy}}, \tag{16.4}$$

and $F(s,\mu) \geq \mathfrak{I}(s)$*, with equality if and only if* $q_\mu(\psi) = p(\psi \mid s)$ *almost everywhere.*

Proof. Factorize $p(s,\psi) = p(\psi \mid s)\, p(s)$ inside (16.1):

$$F = \mathbb{E}_{q_\mu}[\ln q_\mu(\psi) - \ln p(\psi \mid s)] - \ln p(s) = D_{\mathrm{KL}}[q_\mu \| p(\psi \mid s)] + \mathfrak{I}(s),$$

which is (16.2). Splitting (16.1) into its two terms gives (16.3). Factorizing instead $p(s,\psi) = p(s \mid \psi)\, p(\psi)$ gives (16.4). For the bound, Jensen's inequality applied to the concave logarithm yields

$$-D_{\mathrm{KL}}[q \| p(\psi \mid s)] = \mathbb{E}_q\left[\ln \frac{p(\psi \mid s)}{q(\psi)}\right] \;\leq\; \ln \mathbb{E}_q\left[\frac{p(\psi \mid s)}{q(\psi)}\right] = \ln 1 = 0,$$

so the divergence is nonnegative and (16.2) gives $F \geq \mathfrak{I}(s)$. Equality in Jensen for the strictly concave logarithm requires the ratio $p(\psi \mid s)/q(\psi)$ to be q-almost-surely constant, hence equal to 1; i.e., $q_\mu = p(\psi \mid s)$. ∎

Remark 16.3*.* Everything in Proposition 16.2 is textbook variational inference and is uncontested. Equation (16.2) says $-F$ is the evidence lower bound (ELBO) of machine learning; the identity and the bound appear in Dayan et al. (1995) and Neal and Hinton (1998) and ultimately descend from Feynman's variational bound in statistical mechanics. No critique of the FEP disputes this stratum; every critique concerns what is built on it. Minimizing F over μ at fixed s makes q_μ approximate the posterior (perception-as-inference); minimizing F over actions that change s is the additional move that defines active inference.

16.2 The Gaussian case: gradient descent on F is precision-weighted prediction-error dynamics

The FEP's contact with cortical physiology runs through the claim that gradient descent on F in hierarchical Gaussian models *is* predictive coding. This is elementary but worth doing in full, with no steps skipped, since it is the part of the edifice that is both well defined and empirically fruitful; nothing in this subsection goes beyond standard variational inference.

The two-level model. Let a latent cause $x \in \mathbb{R}^{n_x}$ generate data $s \in \mathbb{R}^{n_s}$ through

$$p(s,x) = p(s \mid x)\, p(x), \quad p(s \mid x) = \mathcal{N}(s;\, g(x),\, \Sigma_s)\,, \quad p(x) = \mathcal{N}(x;\, v,\, \Sigma_x)\,, \tag{16.5}$$

where g is a (possibly nonlinear) prediction map, v a prior expectation, and Σ_s, Σ_x are covariance matrices with precisions defined as their inverses, $\Pi_s = \Sigma_s^{-1}$ and $\Pi_x = \Sigma_x^{-1}$. Writing both Gaussian densities out in full, the negative log joint (the "energy") is, term by term,

$$\begin{aligned} U(s,x) \;\equiv\; -\ln p(s,x) &= \tfrac{1}{2}\,(s-g(x))^\top \Pi_s\,(s-g(x)) + \tfrac{1}{2}\,(x-v)^\top \Pi_x\,(x-v) \\ &\qquad + \tfrac{1}{2}\ln\det(2\pi\Sigma_s) + \tfrac{1}{2}\ln\det(2\pi\Sigma_x)\,, \end{aligned} \tag{16.6}$$

with the log-determinant normalization constants displayed because they stop being constants the moment the covariances are optimized (see Section 16.3).

Recognition density and the Laplace free energy. As in the original continuous-state scheme, take a fixed-form Gaussian recognition density $q_\mu(x) = \mathcal{N}(x;\mu,C)$ with the covariance C held fixed, so that the internal state is the mean μ. Its entropy is $H[q_\mu] = \frac{n_x}{2}\ln(2\pi e) + \frac{1}{2}\ln\det C$, independent of μ. Definition 16.1 gives $F = \mathbb{E}_{q_\mu}[U(s,x)] - H[q_\mu]$. Expanding U to second order about μ and using $\mathbb{E}_q[x-\mu] = 0$ and $\mathbb{E}_q[(x-\mu)(x-\mu)^\top] = C$,

$$F(s,\mu) \;=\; U(s,\mu) \;+\; \underbrace{\tfrac{1}{2}\,\mathrm{tr}\Big[C\,\nabla_x^2 U(s,\mu)\Big] - \tfrac{n_x}{2}\ln(2\pi e) - \tfrac{1}{2}\ln\det C}_{\text{treated as constant in }\mu}, \tag{16.7}$$

up to remainder terms involving third and higher derivatives of g, which vanish exactly when g is linear. The Laplace approximation consists of neglecting the μ-dependence of the trace term; in the limit $C \to 0$ the density q_μ becomes a delta function and (16.7) reduces to $F = U(s,\mu)$ exactly, i.e., pure maximum-a-posteriori estimation. Defining the two prediction errors

$$\varepsilon_s = s - g(\mu), \qquad \varepsilon_x = \mu - v, \tag{16.8}$$

and substituting (16.6) at $x = \mu$ into (16.7),

$$F(s,\mu) \;=\; \tfrac{1}{2}\,\varepsilon_s^\top \Pi_s\,\varepsilon_s + \tfrac{1}{2}\,\varepsilon_x^\top \Pi_x\,\varepsilon_x + \tfrac{1}{2}\ln\det(2\pi\Sigma_s) + \tfrac{1}{2}\ln\det(2\pi\Sigma_x) + F_0, \tag{16.9}$$

where F_0 collects the C-dependent terms of (16.7). Every term is now explicit.

The gradient, line by line. Let $g'(\mu) = \partial g/\partial x\,\big|_{x=\mu}$ denote the $n_s \times n_x$ Jacobian. From (16.8), $\partial\varepsilon_s/\partial\mu = -\,g'(\mu)$ and $\partial\varepsilon_x/\partial\mu = I$. Differentiating (16.9) term by term, using the symmetry of the precisions and the chain rule,

$$\frac{\partial}{\partial\mu}\Big[\tfrac{1}{2}\,\varepsilon_s^\top \Pi_s\,\varepsilon_s\Big] = \left(\frac{\partial\varepsilon_s}{\partial\mu}\right)^\top \Pi_s\,\varepsilon_s = -\,g'(\mu)^\top\,\Pi_s\,\varepsilon_s, \tag{16.10}$$

$$\frac{\partial}{\partial\mu}\Big[\tfrac{1}{2}\,\varepsilon_x^\top \Pi_x\,\varepsilon_x\Big] = \left(\frac{\partial\varepsilon_x}{\partial\mu}\right)^\top \Pi_x\,\varepsilon_x = \Pi_x\,\varepsilon_x, \tag{16.11}$$

while the log-determinant terms and F_0 contribute nothing. Hence

$$\nabla_\mu F(s,\mu) = -\,g'(\mu)^\top\,\Pi_s\,\varepsilon_s + \Pi_x\,\varepsilon_x. \tag{16.12}$$

Dynamics and the neural motif. Gradient descent $\dot{\mu} = -\kappa \nabla_\mu F$ on (16.12) gives

$$\dot{\mu} \;=\; \kappa \left[g'(\mu)^\top \xi_s \;-\; \xi_x \right], \qquad \xi_s = \Pi_s \, \varepsilon_s, \qquad \xi_x = \Pi_x \, \varepsilon_x. \tag{16.13}$$

Each symbol has a standing neural interpretation in this literature: $g(\mu)$ is the *top-down prediction* sent to the sensory level; ε_s is the residual computed there by simple subtraction; ξ_s is the *precision-weighted, bottom-up error* — the residual multiplied by a gain — carried by ascending connections; the transpose Jacobian $g'(\mu)^\top$ routes that error back through the same weights that generated the prediction; and ξ_x is the descending error against the prior, which regularizes the estimate toward v. The fixed point of (16.13) balances the two influences,

$$g'(\mu)^\top \, \Pi_s \, \varepsilon_s = \Pi_x \, \varepsilon_x, \tag{16.14}$$

which is exactly the stationarity condition of maximum-a-posteriori estimation under (16.5).

The L-level hierarchy. Stack the construction: let $x_0 \equiv s$ and let each level generate the one below,

$$\begin{aligned} p(s, x_1, \ldots, x_L) &= p(x_L) \prod_{l=1}^{L} p(x_{l-1} \mid x_l), \\ p(x_{l-1} \mid x_l) &= \mathcal{N}\Big(x_{l-1};\, g_l(x_l),\, \Pi_l^{-1}\Big), \end{aligned} \tag{16.15}$$

with top-level prior $p(x_L) = \mathcal{N}(x_L; v_L, \Pi_{L+1}^{-1})$. With means μ_l and errors $\varepsilon_l = \mu_{l-1} - g_l(\mu_l)$ for $l = 1, \ldots, L$ (where $\mu_0 = s$) and $\varepsilon_{L+1} = \mu_L - v_L$, the Laplace free energy is the sum of quadratic terms $F = \sum_{l=1}^{L+1} \frac{1}{2} \varepsilon_l^\top \Pi_l \varepsilon_l$ plus constants. Each μ_l appears in exactly two terms — inside $g_l(\mu_l)$ in ε_l, and linearly in ε_{l+1} — so the same two-line differentiation as (16.10)–(16.11) yields the general layer equation

$$\dot{\mu}_l \;=\; \kappa_l \left[g_l'(\mu_l)^\top \xi_l \;-\; \xi_{l+1} \right], \qquad \xi_l = \Pi_l \, \varepsilon_l, \qquad l = 1, \ldots, L. \tag{16.16}$$

Every level runs the same motif: ascending precision-weighted error from below, descending prediction to below, and correction by the error against the level above. This reproduces the Rao and Ballard (1999) architecture of state units and error units at every level. Two remarks bound the scope of this result. First, (16.16) is a statement about an *algorithm* for a chosen model class; its empirical successes support predictive coding, not the physical claims of Section 16.6. Second, the fixed points of (16.16) are maximum-a-posteriori estimates under (16.15); nothing here requires, or delivers, exact posterior beliefs.

16.3 Precision as attention: covariances as gain control

In (16.13) and (16.16) the precisions enter multiplicatively: Π_l sets how strongly the error at level l drives the estimates — a gain on the error units. This is why the covariance parameters, ordinarily an afterthought, carry the psychology in this framework. Optimizing F with respect to a

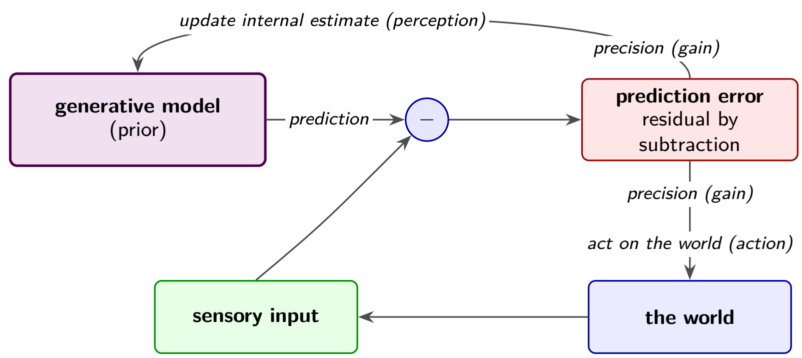


Figure 4: The process-model core of the free energy principle, as derived in Section 16.2. The generative model issues a top-down prediction; the comparator subtracts it from the sensory input; the residual, weighted by a precision that acts as a gain, becomes the prediction error, which is spent twice: ascending, it updates the internal estimate (perception, Eq. 16.13); descending, it drives action on the world, which changes the next sensory input and closes the loop. This loop is the uncontested process-model core of the framework — standard variational inference, empirically fruitful against neural data. The contested part — Markov blankets, inference ascribed to generic matter — is a separate stratum, treated in Section 16.6 and the critiques of Section 18.

precision is again standard variational inference: writing $\frac{1}{2}\ln\det(2\pi\Sigma_l) = -\frac{1}{2}\ln\det\Pi_l + \text{const}$ in (16.9),

$$\frac{\partial F}{\partial \Pi_l} = \tfrac{1}{2}\left(\varepsilon_l\,\varepsilon_l^\top - \Pi_l^{-1}\right), \tag{16.17}$$

so stationarity sets the inverse precision to the (suitably averaged) outer product of the errors: precision estimation is error-variance estimation, i.e., hyperparameter optimization on Σ_l. The FEP reading adds an interpretive layer: *attention* is the selective increase of the precision assigned to some sensory channels over others, implemented physiologically as synaptic gain, with neuromodulators as candidate carriers (Friston, 2010, Friston et al., 2017, Parr et al., 2022). The mathematics licenses the gain-control mechanism; identifying it with any particular neuromodulator is a hypothesis of the process theory, testable case by case.

16.4 What the Laplace approximation assumes, and when it fails

The derivation above rests on three assumptions worth making explicit. First, the posterior is adequately summarized by a single Gaussian centered on a mode, so that a mean and a covariance are sufficient statistics of belief. Second, the second-order expansion behind (16.7) is adequate: the μ-dependence of the trace term is negligible, which is exact if and only if g is linear. Third, the recognition covariance C is either fixed or slaved to the local curvature, so that only the mean carries dynamics. Each assumption has a characteristic failure mode. Multimodal posteriors — perceptual bistability is the standard biological example — defeat the first: a single mean tracks neither interpretation, and gradient descent can park between modes. Strong nonlinearity in g defeats the second: mode and mean separate, and the neglected curvature terms feed back on the dynamics. Heavy-tailed observation noise strains the quadratic penalties of (16.9), which over-weight outliers. And the delta-function limit $C \to 0$ discards posterior uncertainty altogether, reducing the scheme to point estimation. When these conditions fail, the dynamics (16.16) remain perfectly well defined as maximum-a-posteriori estimation, but the reading of μ as the sufficient statistic of an approximate posterior weakens accordingly; the continuous-time schemes in the literature inherit these caveats (Friston et al., 2010). None of this is special to the FEP — it is the standard fine print of Laplace methods in variational inference.

16.5 Stationary stochastic dynamics: the Helmholtz decomposition

The physics phase begins with a Langevin equation and a claim about its stationary density. This stratum is classical and correct; we derive it to fix conventions, because sign and correction-term conventions vary across the FEP corpus.

Consider the Itô stochastic differential equation

$$dx = f(x)\,dt + \omega, \qquad \left\langle \omega(t)\,\omega(t')^\top \right\rangle = 2\,\Gamma(x)\,\delta(t-t'), \tag{16.18}$$

with Γ symmetric positive semidefinite. The probability density $p(x,t)$ obeys the Fokker–Planck equation, which we write in conservation form,

$$\partial_t p = -\partial_i J_i, \qquad J_i = f_i\,p - \partial_j(\Gamma_{ij}\,p)\,, \tag{16.19}$$

summation implied. Suppose a stationary density $p^*(x) > 0$ exists (NESS). Stationarity means $\partial_i J_i^* = 0$: the stationary current is divergence-free, but need not vanish. A divergence-free vector field on $\mathbb{R}^n$ (with suitable decay) can be written as the divergence of an antisymmetric tensor field; absorbing p^* into that tensor, there exists $Q(x)$ with $Q^\top = -Q$ such that

$$J_i^* = \partial_j(Q_{ij}\,p^*)\,, \tag{16.20}$$

which is automatically divergence-free: $\partial_i\partial_j(Q_{ij}p^*) = 0$ by antisymmetry of Q_{ij} against the symmetric $\partial_i\partial_j$. Substituting (16.20) into (16.19) and dividing by p^*,

$$f_i \;=\; (\Gamma_{ij} + Q_{ij})\,\partial_j \ln p^* \;+\; \Lambda_i, \qquad \Lambda_i = \partial_j(\Gamma_{ij} + Q_{ij})\,, \tag{16.21}$$

where we used $\partial_j(A_{ij}p^*) = A_{ij}p^*\,\partial_j \ln p^* + p^*\,\partial_j A_{ij}$. Equation (16.21) is the stationary Helmholtz (gradient–solenoidal) decomposition: a dissipative gradient flow $\Gamma\,\nabla \ln p^*$ that descends the surprisal $\mathfrak{I}(x) = -\ln p^*(x)$, plus a solenoidal circulation $Q\,\nabla \ln p^*$ along its level sets, plus a housekeeping term Λ that vanishes when Γ and Q are state-independent. Detailed balance holds if and only if $Q \equiv 0$ (equilibrium); $Q \neq 0$ is the signature of a genuinely nonequilibrium steady state. This construction is classical: it appears in Graham (1977) and, in the form of a stochastic decomposition with antisymmetric matrix, in Ao (2004). In the FEP corpus the same content is written $f = (Q-\Gamma)\,\nabla\mathfrak{I}$, the sign difference on Q being a relabeling permitted by antisymmetry; the correction Λ is sometimes displayed and sometimes dropped under a constancy assumption. Nothing in (16.21) is specific to the FEP, to life, or to inference: it holds for any diffusion with a positive stationary density.

16.6 The particular-physics chain: blankets, synchronization, and expected free energy

Everything distinctive about the FEP as physics lives in the chain built on (16.21) by Friston (2013) and Friston (2019). We state it as its authors intend it, flagging each contested link; the critiques are examined in Section 18.

The particular partition. Partition the state as $x = (\eta, s, a, \mu)$: external states η, sensory states s, active states a, internal states μ, with blanket $b = (s, a)$. The pair (μ, b) is called a *particle*.

Definition 16.4 (NESS Markov blanket)**.** The partition has a Markov blanket at steady state when external and internal states are conditionally independent given the blanket:

$$p^*(\eta, \mu \mid b) = p^*(\eta \mid b)\, p^*(\mu \mid b), \qquad \text{equivalently} \qquad \partial_\eta \partial_\mu \ln p^*(x) = 0. \tag{16.22}$$

In Friston (2019) this density-level condition is intertwined with a distinct, flow-level sparsity condition — which variables appear in which components of f (e.g., $\partial_\eta f_\mu = 0$, $\partial_\mu f_\eta = 0$, with conditions on f_s and f_a encoding that sensory states are driven by external states and active states by internal ones). The relationship between the two conditions is precisely the first point of contention below.

The synchronization map. Define the conditional expectations $\boldsymbol{\eta}(b) = \mathbb{E}[\eta \mid b]$ and $\boldsymbol{\mu}(b) = \mathbb{E}[\mu \mid b]$. The monograph posits a map σ with

$$\boldsymbol{\eta}(b) = \sigma(\boldsymbol{\mu}(b))\,, \tag{16.23}$$

so that the expected internal state parameterizes a belief $q_{\boldsymbol{\mu}}(\eta)$ over external states, claimed to coincide with the conditional density $p^*(\eta \mid b)$ on the synchronization manifold. This is the content of the monograph's Free Energy Lemma: internal states *parameterize a posterior* over external states.

The claimed gradient flow. Combining (16.21) with (16.22) and (16.23), the monograph asserts that the flow of expected internal (and active) states can be rewritten as a natural gradient flow on a variational free energy,

$$\dot{\boldsymbol{\mu}} = (\Gamma_\mu + Q_\mu)\, \nabla_{\boldsymbol{\mu}} F(b, \boldsymbol{\mu})\,, \qquad F(b, \boldsymbol{\mu}) \simeq \mathfrak{I}(b) \text{ at the fixed point,} \tag{16.24}$$

so that a particle's internal dynamics "look as if" they perform variational Bayesian inference about what lies beyond the blanket — self-evidencing matter. It is this rewriting, not the bound of Proposition 16.2 nor the decomposition (16.21), that the technical critiques target: as shown below, it requires additional assumptions (notably the absence of solenoidal coupling between internal and external sectors) that were not stated in the original derivations and are not generic.

Expected free energy. Active inference over policies π (sequences of actions) is driven not by F but by the *expected free energy*

$$G(\pi) = \sum_{\tau} \mathbb{E}_{q(o_\tau,\psi_\tau|\pi)}[\ln q(\psi_\tau \mid \pi) - \ln p(o_\tau, \psi_\tau)] \,, \tag{16.25}$$

where o_τ are future observations and the expectation runs over the predictive density. Factorizing $p(o,\psi) = p(\psi \mid o)\,p(o)$ and approximating $p(\psi \mid o)$ by the model's own posterior $q(\psi \mid o, \pi)$ gives the standard decomposition

$$G(\pi) = \sum_{\tau} \Big(\underbrace{-\,\mathbb{E}_{q(o_\tau|\pi)}\, D_{\mathrm{KL}}[q(\psi_\tau \mid o_\tau, \pi) \,\|\, q(\psi_\tau \mid \pi)]}_{-\text{ epistemic value (information gain)}} \; \underbrace{-\,\mathbb{E}_{q(o_\tau|\pi)}[\ln p(o_\tau)]}_{-\text{ pragmatic value}} \Big), \tag{16.26}$$

with $p(o_\tau)$ read as a *prior preference* density and the policy posterior set to $q(\pi) \propto \exp(-\gamma\, G(\pi))$ (Friston et al., 2017). An equivalent rearrangement gives the risk-plus-ambiguity form $G_\tau = D_{\mathrm{KL}}[q(o_\tau \mid \pi) \,\|\, p(o_\tau)] + \mathbb{E}_{q(\psi_\tau|\pi)} H[p(o_\tau \mid \psi_\tau)]$. Two facts should be plainly stated for a physics readership. First, the epistemic term makes exploration fall out of the objective rather than being bolted on — this is the genuinely attractive feature of active inference as a control principle. Second, *G does not follow from F*: Millidge et al. (2021) showed that the expected free energy is not the expectation of the variational free energy of future trajectories (the natural candidate, which they call the free energy of the expected future, differs from G by sign-relevant terms), so the step from "systems minimize F" to "agents choose policies minimizing G" is an independent postulate, not a theorem.

17 What is established and what is claimed

The FEP literature interleaves four kinds of statement — theorems of variational statistics, classical results of stochastic dynamics, model-class constructions, and physical claims about generic self-organizing matter — and much of the controversy dissolves once they are held apart. Table 3 summarizes the assessment argued for in this Part.

What is solid: the bound $F \geq \mathfrak{I}(s)$ (Proposition 16.2) is elementary and old; gradient descent on F in Gaussian hierarchies (Section 16.2) is a well-defined, useful algorithm; the Helmholtz decomposition (16.21) is classical nonequilibrium statistical mechanics (Ao, 2004, Graham, 1977); and, within its stated linear-Gaussian assumptions, the Bayesian mechanics of Da Costa et al. (2021) is rigorous — for multivariate Ornstein–Uhlenbeck processes on which a blanket is imposed, the synchronization map exists, is affine, and is computable, and the expected internal flow is exactly a natural gradient flow on an explicit free energy.

What is claimed but not established: that generic (or even typical) stationary systems possess Markov blankets in the sense of Definition 16.4; that flow sparsity delivers density factorization; that a blanket entails the synchronization map (16.23); that the expected internal flow is the gradient flow (16.24) without fine-tuned restrictions on the solenoidal coupling Q; and that any of this licenses reading physical dynamics as inference. The 2023 path-integral reformulation (Friston et al., 2023a,b) quietly relinquishes the NESS factorization as the foundation and restates the FEP as: the

Table 3: Status of the principal mathematical claims of the free energy principle, separated by stratum. The first three rows are uncontested mathematics whose authority predates or is independent of the FEP program; the fourth is rigorous but restricted to the linear-Gaussian setting; the remaining rows are the load-bearing physical claims of the 2013–2019 program, each of which has been shown to require additional non-generic assumptions, to fail at face value, or to hold only by construction in the reformulated path-integral setting.

Claim	Status	Authority
$F \geq \mathfrak{I}(s)$, equality iff q is the posterior	Theorem (elementary)	Textbook variational inference; Dayan et al. (1995), Neal and Hinton (1998)
Gradient descent on F in Gaussian hierarchies yields precision-weighted prediction-error dynamics	Well-defined algorithm	Friston (2005), Rao and Ballard (1999)
NESS drift decomposes as $f = (\Gamma + Q)\nabla \ln p^* + \Lambda$	Classical theorem	Ao (2004), Graham (1977)
Synchronization map and free-energy gradient flow for linear (Ornstein–Uhlenbeck) processes with an imposed blanket	Theorem under stated assumptions	Da Costa et al. (2021)
Flow-sparsity blanket $\Leftrightarrow$ density-factorization blanket	False in general; the two conditions are logically independent	Biehl et al. (2021); conceded in Friston et al. (2021)
Expected internal flow is a gradient flow on F (eq. 16.24) for generic blanketed systems	Invalid without unstated assumptions (no solenoidal coupling across the partition); non-generic even for linear systems	Aguilera et al. (2022), Biehl et al. (2021)
Free Energy Lemma: internal states parameterize the posterior over external states	False at face value; counterexamples exist	Biehl et al. (2021)
Expected free energy G follows from minimizing F	Does not follow; independent postulate	Millidge et al. (2021)
FEP as such is empirically falsifiable	Denied by proponents; principle, not law	Colombo and Wright (2021), Ramstead et al. (2023)

paths of least action of a system admitting a particular partition *can be expressed* as extremizing a free energy functional of beliefs — a statement closer to a definition, or to a choice of variational representation, than to a theorem about matter in general.

18 The critiques and the replies

The technical literature scrutinizing the FEP's physics phase is compact and of unusually high quality; each of its main results can be stated precisely.

18.1 Biehl, Pollock, and Kanai (2021): the internal mathematics as stated

Biehl et al. (2021) ("A Technical Critique of Some Parts of the Free Energy Principle," *Entropy* 23(3), 293) worked through Friston (2019) and Friston (2013) line by line and established three results.

First, *the two blanket definitions are inequivalent.* The definition via sparsity of the flow's Jacobian (which variables appear in which components of f) and the definition via factorization of the stationary density (16.22) are logically independent: neither implies the other, and the monograph slides between them.

Second, *the rewriting of the equations of motion is invalid as stated.* The monograph's claim that the flow of the expected internal states, conditioned on blanket states, depends only on blanket states — and can be written in the form (16.24) — does not follow from the assumptions given. It requires, in particular, the absence of solenoidal coupling between the internal and external sectors of Q in (16.21), an assumption used silently.

Third, *the Free Energy Lemma fails at face value.* The lemma asserts the existence of a parameterized density $q(\eta \mid \boldsymbol{\mu})$ equal to the conditional stationary density $p^*(\eta \mid b)$, with free energy reducing to surprisal on the relevant manifold; the authors exhibit counterexamples. They further prove a pointed lemma of independent interest: the vanishing of the *gradient* of the Kullback–Leibler divergence with respect to internal states does not imply that the divergence itself vanishes — so stationarity of internal states does not imply that the "belief" equals the posterior. They also observe that later formulations parameterize the variational density by different variables than earlier ones, changing the interpretation of what internal states encode. The paper is explicit that it targets specific derivations, not the possibility that some version of the program could be repaired.

18.2 The reply of Friston, Da Costa, and Parr (2021): concession and the sparse-coupling ansatz

Friston et al. (2021) ("Some Interesting Observations on the Free Energy Principle," *Entropy* 23(8), 1076) is the authorized response, and its substance is a concession plus a narrowing. The reply accepts that the observations of Biehl et al. (2021) correctly identify implicit assumptions and incomplete, heuristic proofs in the earlier formulations, and then supplies an explicit *sparse-coupling ansatz*: a set of sufficient conditions on the flow under which the flow-level blanket does induce the density factorization and under which the offending solenoidal couplings across the partition are precluded. The theorems were thus not rescued in generality; the antecedent conditions of the principle were restricted until the argument goes through. Subsequent work moved the foundation again, from expected flows at NESS to the path-integral formulation of Friston et al. (2023a).

18.3 Aguilera, Millidge, Tschantz, and Buckley (2022): non-genericity even for linear systems

Aguilera et al. (2022) ("How particular is the physics of the free energy principle?" *Physics of Life Reviews* 40, 24–50) posed the decisive quantitative question: on what fraction of parameter space do the FEP's conditions hold, in the very model class the FEP literature uses for illustration? For linear (multivariate Ornstein–Uhlenbeck) stationary processes everything is exactly solvable, and the answers are as follows.

First, the conjunction of the NESS blanket condition (16.22) with the required restrictions on solenoidal flow — no Q-coupling between internal and external blocks across the blanket — holds only on a very narrow parameter subspace, and satisfying it demands a symmetry between the perception and action channels that amounts to an *absence of perception–action asymmetry*. Systems with directed, asymmetric agent–environment coupling — the intended subject matter of the principle — generically violate the conditions.

Second, the step connecting dynamics to inference relies on an implicit equivalence between *the dynamics of the average* states and *the average of the dynamics*. The two coincide only under conditions, effectively linearity together with decoupling from the trajectory's history, that fail in general; and even in the linear systems where the construction can be completed, the inferential reading adds no predictive content beyond the dynamics themselves.

Third, even when a blanket exists, the internal states track the conditional expectation of the external states — the substance of the synchronization map (16.23) — only in special cases. The authors conclude that the FEP is not straightforwardly applicable even to the simple linear systems studied, and the paper appeared with an open commentary exchange in the same journal.

18.4 Raja, Valluri, Baggs, Chemero, and Anderson (2021): the blanket as an artifact of model choice

Raja et al. (2021) ("The Markov blanket trick: On the scope of the free energy principle and active inference," *Physics of Life Reviews* 39, 49–72) mounted the complementary, dynamical-systems objection: in applications, the blanket is not discovered in the system but induced by the modeler's choice of state-space parcellation — the "trick" of the title. Working through FEP-style treatments of canonical coupled systems, with the Watt centrifugal governor as a central example, they argue that strongly coupled nonlinear systems admit no non-arbitrary blanket; that blanket-derived boundaries need not coincide with biological boundaries; and that the principle's claimed universal scope is therefore a property of modeling practice, not of the systems modeled.

18.5 Bruineberg, Dolega, Dewhurst, and Baltieri (2022): Pearl blankets versus Friston blankets

Bruineberg et al. (2022) ("The Emperor's New Markov Blankets," *Behavioral and Brain Sciences* 45, e183, with an extensive open commentary and response) drew the conceptual distinction that organizes the philosophical debate. A *Pearl blanket* is Judea Pearl's original notion: the minimal set of nodes rendering a variable conditionally independent of the rest *within a model used by a scientist* — an epistemic instrument. A *Friston blanket* is the FEP literature's reification of that construct as a physical boundary of an agent in the world. The move from the former to the latter — from inference *with* a model to inference *within* a model — is a category transition that the mathematics does not underwrite; conclusions about the boundaries of agents or minds drawn from blanket formalism alone are therefore unsupported without independent argument.

18.6 Colombo and Wright (2021): principle status and falsifiability

Colombo and Wright (2021) ("First principles in the life sciences: the free-energy principle, organicism, and mechanism," *Synthese* 198(S14), 3463–3488) reconstructed the FEP's argument chain — from thermodynamics through homeostasis to surprisal and free energy — and showed that it depends on stipulative identifications across disciplines that lack independent justification. Their sharpest point concerns epistemic status: the FEP is defended by its proponents as a *principle*, and principles, on the proponents' own account, are not falsifiable — only their applications are. It therefore cannot function as an empirical law of biology, and its axiomatic, physics-first style sits poorly with the mechanistic explanatory norms of the life sciences.

18.7 The instrumentalist defense and its cost

The now-standard response from within and around the program concedes much of the above by deflating the principle's status. Andrews (2021) ("The math is not the territory," *Biology and Philosophy* 36(3), 30) argues the FEP is a scientific-modeling tool whose worth is instrumental, not a falsifiable empirical claim about systems. van Es (2021) (*Adaptive Behavior* 29(3), 315–329) presses the same map–territory discipline: FEP models are the scientist's instruments, and conflating model with target is an error the literature repeatedly invites. From inside the program, Ramstead et al. (2023) makes the deflation official: Bayesian mechanics is "a physics of and by beliefs," a mechanics on statistical manifolds, and the FEP is explicitly compared to a principle of least action — not itself falsifiable, but a scheme for generating falsifiable process models.

The tension should be stated plainly. The instrumentalist reading secures the FEP by surrendering the strong claim — that all persisting systems minimize free energy, as a matter of physical fact — which made the principle famous, and which motivated the physics phase in the first place. One cannot simultaneously advertise a theory of "every thing" and retreat, under technical pressure, to the position that the formalism is a lens the modeler may choose to wear. The program's own trajectory — from NESS theorems (2013–2019), through concession and restriction (2021), to a path-integral formulation in which the principle holds by construction for systems admitting a particular partition (2023) — is best read as the gradual replacement of a physical conjecture by a modeling framework.

Consensus. As of the mid-2020s the assessment across the technical exchanges is reasonably stable. As modeling mathematics — variational free energy, predictive coding, discrete-state active inference with expected free energy — the framework is legitimate, practical, and empirically productive: simulation work, behavioral model fitting, and predictive-coding-consistent neurophysiology all belong to this stratum, and all of it stands independently of the physics-phase claims. As a theorem about generic matter, the FEP is not established: the 2013–2019 derivation chain required unstated and non-generic assumptions, as its authors have in substance acknowledged, and rigorous results exist only in restricted settings. The honest one-sentence summary for a physicist is this: *the bound is trivial, the decomposition is classical, and the bridge from blankets to inference is either false in general, true by construction, or true under fine-tuned conditions — depending on which paper one reads.* How the present program reads this construction is the subject of Part VII.

19 Closing remarks: what a life scientist should take from this Part

Strip away the formalism and two different kinds of claim remain, and the single most useful skill for a reader of this literature is telling them apart. A *model that fits brains* is judged the way any model is judged: does it predict responses it was not tuned to, does it fail in informative ways, does it beat its rivals on data. A *law about matter* is a different kind of claim altogether — it asserts that something must be so for anything that persists, brains included, bacteria included, arguably rocks included. The free energy literature contains both kinds of claim, frequently in the same paper and sometimes in the same sentence, and most confusion about the principle — both the enthusiasm and the dismissal — comes from reading one kind as the other.

This is why the Markov blanket debate matters outside philosophy seminars. The blanket is the formal boundary that separates a system from its environment, and everything in the strong version of the principle flows through it. If blankets are *discovered* — if nature itself fixes where the organism ends and the world begins — then the principle is a finding about living matter. If blankets are *imposed* — if the boundary appears only because the modeler chose to carve the variables that way — then the principle is a notation, and the statement "the cell minimizes free energy" has the same standing as "the planet obeys polar coordinates." The technical work reviewed above leans toward the second reading for systems in general, while leaving the first available in special, carefully constructed cases. That is not a verdict against the framework; it is a verdict about what kind of thing the framework is.

What survives the critiques, fully intact, is the working machinery. The predictive-processing picture of the brain — prediction down, error up, gain as attention — is a productive and testable research program with an experimental record: characteristic brain responses to surprising input, quantitative fits to choice behavior, working simulations and robots. None of that machinery needs the claim about all persisting matter, and none of it falls if that claim falls. A biologist can use every practical tool in this Part while remaining entirely agnostic about the metaphysics — most practitioners now do exactly that.

And once more, because it costs someone a misread abstract every week: the free energy here is an accounting quantity for surprise, not the Gibbs or Helmholtz free energy of the chemist — the two share a variational ancestry and a name, nothing more.

The one-sentence version to carry away: used as a tool, the free energy principle has earned its place in biology; proclaimed as a law, it has not yet earned its proofs.

Part V
The Common Pattern: What the Critical Literature Established

Three principles, three restricted regimes, one shared obstruction

A plain-language word before the synthesis. The previous three Parts were three separate case files; this one reads them side by side, and the experience is a little like a physician seeing three patients in one afternoon with the same unusual presentation. Each of the three principles worked brilliantly somewhere. Each failed somewhere else. And — the finding of this Part — each failed in the *same three ways*: nobody knew its territory in advance, its central choices were made by the person applying it rather than by nature, and its most serious attempted derivation broke at a specific published step. When three unrelated theories fail identically, the failure stops being three pieces of bad luck and becomes a symptom — evidence of one missing structure underneath all three. This Part writes up that symptom precisely, because the cure proposed later will be judged by whether it treats exactly this, and not something easier.

20 Each principle is exact only in a bounded regime, and the bounds are of the same kinds

The scorecard assembled in the review Parts is uniform in a way that no single literature can see from inside.

MinEP is a theorem — but only under the full assumption list of Part II: strictly linear phenomenological laws, constant coefficients, Onsager symmetry, positive definiteness, clamped versus free forces, and state variables even under time reversal. Every classical counterexample violates one named assumption: Landauer's driven circuit violates parity, Jaynes's two-temperature network violates constancy of the coefficients across the system, nonlinear chemistry violates linearity. Beyond the linear regime the Glansdorff–Prigogine program itself established that only the force half of the entropy-production derivative retains a sign, and that this half is not the differential of any state functional: there is, provably, no far-from-equilibrium potential of the entropy-production family.

MEPP is one theorem, one durable heuristic, and one failed derivation, and the three must not be conflated. The theorem (Kohler–Ziman) is confined to linearized transport, where it is equivalent to Onsager's least dissipation and adds nothing. The heuristic (Paltridge's climate closure and its planetary extensions) works strikingly and remains underived; its own practitioners could not say in advance which entropy production to maximize, and the choice that works ignores the radiatively dominant part of the budget. The derivation (the path-information route) failed at two published steps — an unjustified monotonicity assumption and a Gaussian evaluation valid only near equilibrium — and its author's later position reclassifies the principle as an inference scheme whose output is only as good as the constraints fed in.

FEP divides the same way. Its working mathematics — the variational bound, gradient descent on a recognition density, predictive-coding process models — is standard and solid. Its physical claim — that generic persisting systems possess the blanket structure and coupling restrictions under which their internal dynamics reads as inference — failed under technical scrutiny: the two blanket definitions are inequivalent, the rewriting of the internal flow needs a silent no-solenoidal-coupling assumption, and the required conditions are non-generic even for linear systems, amounting to the absence of the very perception–action asymmetry that living systems exhibit. The program's own response has been a retreat to an explicitly instrumentalist formulation — a physics *of and by beliefs* — which secures the framework by surrendering the claim that made it a candidate law of nature.

21 The shared obstruction

Beneath the three case histories lies one structural fact, made precise in the stochastic-thermodynamics literature and met three times in this review. What selects a nonequilibrium state is not exhausted by any entropy-production functional. Near equilibrium, the dynamical fluctuation functional degenerates to (a quarter of) the excess entropy production, and the extremal principles hold to first order — this is exactly the Maes–Netočný result, and it is the honest modern statement of Prigogine's theorem. Beyond first order, a time-symmetric contribution — dynamical activity, the "frenetic" part — enters independently, and circulating currents that do no work near equilibrium begin to do work. No functional built from entropy production alone contains that information; therefore no pure extremum of entropy production, minimum or maximum, can be exact beyond linear order. The same obstruction reappears in the free energy principle's difficulties: the solenoidal (circulating) couplings that the blanket construction must assume away are precisely the sector whose work invalidates the gradient-flow reading.

This is the fact a candidate derivation must respect. A framework that "derived" MEPP as universally true would be wrong — the counterexamples are theorems too. What a correct derivation must deliver is the functional *whose restricted readings the three principles are*, together with the location of the restrictions.

22 Concerns the communities themselves state

For the record, the standing concerns as the practitioners themselves put them, gathered from the sources reviewed in Parts II–IV:

- *Scope opacity.* None of the three principles announces its domain. MinEP's assumptions were codified only after decades of misuse; MEPP's own advocates published its restrictions as a separate corrective decades after the principle's adoption; the FEP's applicability conditions were extracted by its critics, not stated by its founders.
- *Constraint dependence.* MEPP's predictions change with the choice of constraints and of which entropy production is extremized, and nothing in the principle fixes either choice. The

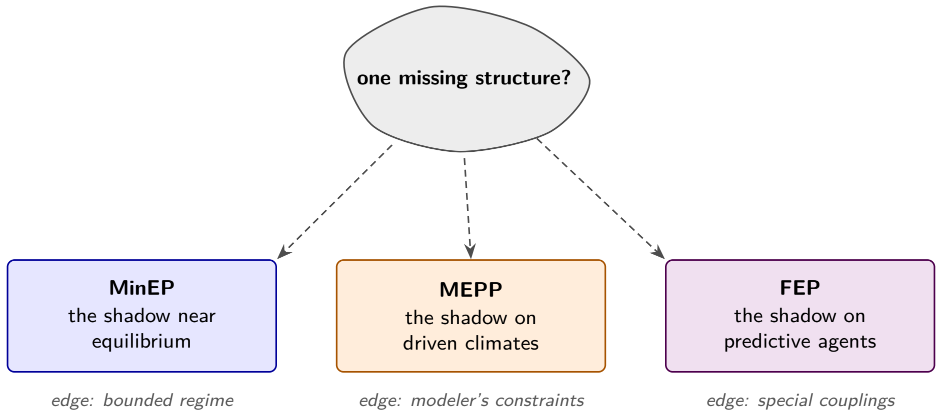


Figure 5: The shared obstruction of Section 21, drawn as a projection. One unknown structure casts three shadows: the three principles, each exact on its own screen and each with a characteristic edge — MinEP bounded to the linear regime, MEPP dependent on the modeler's constraints, FEP dependent on special, non-generic couplings. When three literatures that never read each other produce principles that fail in the same three ways, the shared failure pattern is evidence of one structure beneath all three. The candidate caster is examined in the following Parts, and the test set there is exactly this figure: whatever claims to cast the shadows must match the shadows *and* their edges — and the edges are the hard part.

same concern appears in the FEP as the arbitrariness of the blanket partition: the boundary is chosen by the modeler and the conclusions follow the choice.

- *Unfalsifiability in the strong reading.* A principle defended as unfalsifiable-by-status — only its applications can fail — cannot function as an empirical law, and this defense is now the explicit position of the FEP's proponents and the de facto position of MEPP's. The strong readings survive in the popular literature after being withdrawn in the technical one.
- *The missing bridge to fundamental physics.* Each principle has been connected *downward* to phenomena but never *upward* to a law. The path-information route was the serious attempt for MEPP and it failed; the NESS route was the serious attempt for the FEP and its antecedent conditions turned out non-generic; MinEP needs no bridge only because it claims so little.

23 What would count as an answer

The assessment fixes the specification for Part VII. A derivation worth the name must: (i) produce the three principles from one object that is itself derived, not postulated; (ii) produce their *restrictions* as theorems — the linear-regime confinement, the parity condition, the coupling conditions — rather than deriving any principle as universally true; (iii) dissolve, not adjudicate, the minimum–maximum quarrel, by exhibiting the two as dual readings of one stationary point; (iv) replace the modeler-chosen structures (which entropy production; which blanket) with structures fixed by the framework; and (v) say plainly what remains conditional or open in its own chain. Those five criteria are applied, in order, at the end of that Part.

Table 4: The three principles side by side, as established in Parts II–IV. Each row is one principle; the columns give the regime in which it is exact, the strongest rigorous result in its favor, the identified failure of its most serious general derivation, and the community's present reading. The table's point is the uniformity of the middle two columns: every rigorous result is confined to the linear neighborhood of equilibrium or to fine-tuned coupling conditions, and every general derivation has a named, published point of failure. This is the pattern the derivation of Part VII must reproduce, not explain away.

Principle	Exact where	General derivation	Present reading
MinEP (Prigogine 1945)	Linear laws, constant symmetric coefficients, even variables, clamped/free force split	None needed near equilibrium (it is a theorem there); provably no extension: the evolution criterion is not integrable	First-order approximation from fluctuation theory; Lyapunov property of the linear regime
MEPP (Ziegler; Paltridge)	Linearized transport (Kohler–Ziman), where it restates least dissipation	Path-information route failed at the monotonicity assumption and the Gaussian step; conceded by its author	Inductive heuristic with restrictions; constitutive selection principle, dual to MinEP
FEP (Friston)	Fine-tuned coupling: blanket factorization plus no solenoidal cross-coupling; linear/Gaussian settings	NESS route required silent assumptions, non-generic even for linear systems; program retreated to path-integral formulation	Instrumentalist modeling principle, "a physics of and by beliefs"; process theories evaluated separately

24 Closing remarks: the pattern is the evidence

For the life-science reader, the one thing to carry out of this Part is a way of thinking rather than a fact. When a field's best rules share a failure pattern, the pattern is data. Three literatures that never read each other — climate physics, irreversible thermodynamics, and computational neuroscience — each produced a principle that works in a bounded regime, each left the boundary for its users to discover the hard way, and each saw its deepest derivation fail on a step involving the same physical ingredient: motion that circulates instead of descending. In a clinic, three unrelated patients with one syndrome would send you looking for a common exposure. Here the common exposure is structural: all three principles are shadows cast by something none of the three literatures had in hand — and the shape shared by the shadows is the best description we have of the thing that casts them. The five criteria at the end of this Part are that description turned into a specification. Whatever claims to be the caster — the framework of the next two Parts, or a better one someone else builds — must match the shadows *and* their edges. The edges are the hard part, and that is exactly why they are the test.

Part VI
The Framework in Brief: GDOS, GRIP, and Qi-4

What the three companion papers supply, reviewed for a reader who has not opened them

Before the derivation, the toolbox. The three principles reviewed so far will be derived from a framework built in three companion papers, and a reader should not need those papers open on the desk to follow what happens next. This Part reviews the three ingredients — a boundary law, an engine, and an information ledger — in about ten pages, in words first and equations second. A reader who wants the constructions in full will find them in Arisaka (2026, A5-GDOS), Arisaka (2026, A6-GRIP), and Arisaka (2026, A7-Qi4); a reader who wants only enough to audit Part VII will find that here.

The one idea to hold before any detail: in this framework, *no observer stands outside*. Every description of every system is written by something finite, embedded, and partial — a detector, a cell, a scientist — and the framework's starting move is to make the law of that partial view the fundamental object, rather than an approximation to some unavailable whole. Everything below unpacks that one move.

25 GDOS: the law of the partial view

25.1 The idea in plain terms

Take anything — a molecule, a mouse, a market — and ask what an observer can actually have of it. Never the whole state: always a handful of tracked quantities, sampled through instruments, with everything else untracked. Arisaka (2026, A5-GDOS) proves that the description available under those conditions can never close on itself: two situations that look identical in the tracked variables can evolve differently because they differ in the untracked ones, so the visible dynamics needs memory, hidden variables, or averaging to make up the difference. That is what "open system" means here — not a system with a hole in its wall, but any system described by something finite. Openness is the price of being partial, and since every description is partial, every described system is open. The difference between physics and biology, on this view, is not that one is open and the other closed; it is only how much survives the projection.

GDOS — the Geometric Dynamics of Open Systems — is the law of that projected description. Its derivation chain in Arisaka (2026, A5-GDOS) runs from the exact memory-carrying equation for a reduced state (the Nakajima–Zwanzig form), through the memoryless limit valid when the environment forgets quickly (the GKLS semigroup), to the level this supplement works at: a smooth flow of probability over the tracked variables.

25.2 The mesoscopic law

At that level the state is a density $\rho(z,t)$ over effective coordinates z^a (concentrations, positions, potentials, rates — whatever the observer tracks), and the law is a continuity equation, $\partial_t \rho + \nabla_a J^a = 0$: probability is moved, never created. The content is in the current J^a, which GDOS delivers in four sectors, met already in Part VII's preview and central to everything there: a *gradient* sector that flows downhill on a budget, a *curl* sector that circulates without descending, a *noise* sector that spreads, and a *gauge* sector that moves along directions the system maintains as equivalent. For a life-science reader the four have direct readings: downhill chemistry; futile cycles and oscillations; diffusion; and the freedom of a cell to relabel its internal conventions without changing its physiology. The counterexamples of Parts II–IV all live in the second and fourth sectors, which is why the four-way split is the right anatomy for this supplement.

25.3 Where the law comes from, and its grade

What separates GDOS from a phenomenological ansatz is its pedigree: it is derived as the boundary shadow of a six-dimensional bulk theory — constructed in Arisaka (2026, A2-GG6), proved anomaly-free in Arisaka (2026, A3-BI6), on the constitutional charter of Arisaka (2026, A1-Constitution) — projected through a lawful bulk-to-boundary channel. The two extra dimensions are not places: one is *scale* (how the description changes with magnification) and one is *phase* (an angle carrying quantized, protected windings). Nothing in this supplement re-derives any of that; what matters here is only that the boundary law, and the functional built on it, arrive with no adjustable landscape — the property the three reviewed literatures lacked. These imports are graded *derived* in their home papers, each with its own falsifiers, and this supplement is no more secure than they are.

26 GRIP: the engine

26.1 The idea in plain terms

A law of motion is not yet an account of why anything settles. Arisaka (2026, A6-GRIP) supplies the engine: a three-move cascade by which an open system holds its options open (DGI), loses stability and commits to one (GSSB), and is captured somewhere it can stay (GA). When the basin reached is the deepest available on its stratum, the terminus is a *ground* attractor (GGA) — the reason matter is permanent rather than merely long-lived. A chemist knows this engine as relaxation to the ground state; a developmental biologist knows it as a landscape with committed fates; the framework's claim is that these are literally the same machine at different scales.

Two regimes matter downstream. Descent on a landscape that stays put finishes a journey (GRIP-In); the reorganizations that reshape the landscape and open new basins (GRIP-Out) start the next one. The supplement uses only the first regime — steady states and their selection — but the Bénard reading of Part VIII sits exactly at the boundary between the two, which is why it is the hard case for every entropy principle.

26.2 The single budget

The engine descends one scalar ledger, and its uniqueness is a theorem, not a convenience:

$$F_{\mathsf{GFF}} = F_{\mathsf{GEP}} + F_{\mathsf{CEP}} = (\text{energy–geometry cost}) + \Theta_{\text{eff}} \times (\text{information cost}), \tag{26.1}$$

an energetic term plus an effective temperature times a relative entropy — recognizably the shape of a chemist's free energy, which is no accident: the Gibbs free energy is this ledger's reading in the regime where the informational current vanishes. The balance identity of Arisaka (2026, A6-GRIP) — ledger change = −dissipation + port power + curl work + gauge work — is the single most used equation in Part VII: the entropy-production principles hold exactly when the last two terms vanish, and every classical counterexample is a case where they do not.

26.3 The grade, stated where it matters

The cascade's closure is conditional on the one residue left open in its home paper: the monodromy statement M11, carried there in spectral-flow form, its residues reduced to one named identification and one scope convention. (Two items once stood beside it: the budget bound M10, since resolved externally by the capacity theorem of Arisaka, 2026, P2-N3, which bounds what the cascade must carry without using the cascade at all; and the warped and relative refinements R-WARP/R-REL, since closed at class level by that same paper's rigidity theorems.) This supplement quarantines the surviving condition carefully: the gauge derivations of Part VII need only the boundary law and the ledger; only statements invoking halting or attractor selection — the Bénard commitment, for instance — carry the conditional grade, and the grade table there says so row by row.

27 Qi-4: the information ledger made physical

27.1 The idea in plain terms

The third ingredient answers the question a life scientist will have been asking since Part IV: when the free energy principle says a cell "minimizes surprise," what physical quantity, if any, is being spent? Arisaka (2026, A7-Qi4) makes the informational column of the ledger exact through four pillars, of which two are load-bearing here.

Qi-QUA prices information. It proves that at the level of elementary boundary events, energy, electric charge, baryon number and information are not spent independently: all four move through one channel, in one fixed proportion, so that the four ledgers are one ledger read in four currencies. The exchange rate between the energy column and the information column is a single derived constant, $\mathsf{Q_B} = (2/5)E_{\text{Ry}}/K$ with $K = 35$ — about 0.155 eV per bit, or $5.8\,k_BT$ at body temperature — assembled from the hydrogen atom's energy scale and a ledger integer fixed upstream, with nothing adjustable. Its anchor in the laboratory is the molecule every reader of this supplement has an opinion about: the phosphoanhydride bond of ATP carries two such quanta, and the derived $2\mathsf{Q_B} = 30.0$ kJ/mol stands against the measured 30.5 ± 0.5 — agreement to 1.6%. For this supplement the point is architectural: when the inferential gauge of Part VII speaks of the cost of updating a belief, the cost is denominated in a currency with a derived exchange rate, not a metaphor.

Qi-EP is the pillar this supplement deploys: the theorem that the projected functional admits exactly the three readings — trajectory, current, inferential — whose full derivations occupy Part VII. The remaining pillars (Qi-CON, the scale-signature diagnostics with their two-knee spectrum, and Qi-HAL, the memory architecture) appear only glancingly here, in the neural reading of Part VIII.

27.2 The packet, the ports, and the rank condition

Two structures from the Qi-4 construction replace, in Part VII, the two modeler-chosen structures the critical literatures attacked. The *system packet* (SIP) replaces the Markov blanket: a system is written down as a declared five-field object — state manifold, *ports*, generator, ledger and currents, observables *with kill criteria* — so the system–environment split is constitutive data carrying its own falsifiers, not a partition hunted for in a stationary density. And the *rank condition* of the internality construction replaces the contested synchronization map: an internal model of the surroundings exists exactly when an explicit action-conditioned observability condition holds, so whether a given system supports the inference reading becomes a checkable property of its sensorimotor loop — a bacterium that can swim satisfies it; the same bacterium tethered does not.

28 The assembly, and what this Part has not done

One paragraph of assembly, since Part VII opens by using it. GDOS supplies the stage and the law of motion; GRIP supplies the engine and the unique budget the motion descends; Qi-4 makes the budget's information column physical and proves that the resulting constrained functional — transport cost plus budget, subject to continuity — admits three readings. The next Part takes that functional and performs the three readings at full length.

What this Part has not done is defend the framework's foundations. The six-dimensional construction, the anomaly proof, the constitutional charter, the derivation of the boundary channel — each is a companion paper with its own arguments and its own named falsifiers, and a skeptical reader is entitled to remain skeptical of the whole while still auditing the derivation ahead, which is exactly the posture the supplement invites: the derivation is a consequence-machine, and consequence-machines are tested by their consequences.

29 Closing remarks: three ingredients, one posture

For the life-science reader, the three ingredients of this Part are worth restating as postures rather than equations.

GDOS is the posture that *the view from inside is the fundamental one.* Biology never had the luxury of the outside view — every assay is a projection, every model organism a coarse-graining — and it has sometimes apologized for this, as if physics owned the unprojected truth. In this framework the apology runs the other way: physics, too, is a partial view, distinguished only by

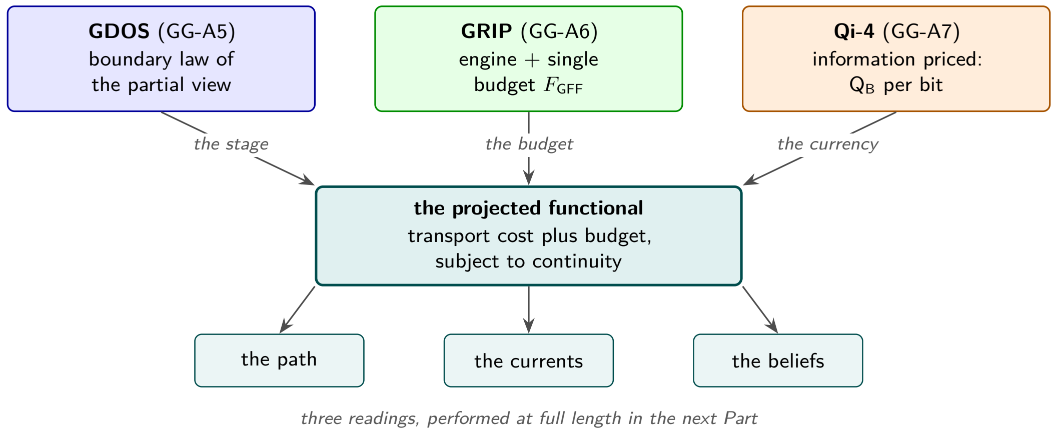


Figure 6: The assembly this Part hands to Part VII. GDOS supplies the stage: the boundary law of the partial view. GRIP supplies the engine and the single budget F_{GFF} that the motion descends. Qi-4 makes the budget's information column physical, priced at the derived rate $\mathsf{Q_B}$ per bit. Together they fix the projected functional — transport cost plus budget, subject to continuity — whose three readings (the path, the currents, the beliefs) occupy the next Part at full length. The imports are graded derived in their home papers, each with its own falsifiers; the cascade's closure is conditional on the engine's one remaining residue — the monodromy statement M11, in spectral-flow form; the budget bound M10 is resolved externally and the scope refinements are closed at class level by Arisaka, 2026, P2-N3 — and that grade travels with every statement downstream that uses it.

how little survives its projection. The tools of this supplement are therefore native to biology's situation, not borrowed finery.

GRIP is the posture that *settling is a process with a grammar*: explore, commit, rest — with a budget that is single, so that exploration, commitment and maintenance are charged to one account. Developmental biology has drawn this picture since Waddington; the framework's contribution is to make the landscape's uniqueness a theorem and its currency exact.

And Qi-4 is the posture that *information is a physical column in the books.* Nothing that remembers, senses, or predicts is doing something beyond physics; it is doing accounting in a currency whose exchange rate against the joule is fixed — and checked, at the one-and-a-half-percent level, by the bond every living cell spends all day breaking.

None of this requires a reader to believe the six-dimensional story behind it. It requires only the willingness to test one claim: that with these three ingredients, and no others, the three principles of Parts II–IV — and, more tellingly, their failures — fall out as consequences. That test is the next Part.

Part VII
The Derivation: Three Principles as Boundary Gauges of One Projected Functional

GDOS supplies the functional, GRIP the dynamics, Qi-EP the three readings

Before the mathematics, the picture. Imagine a single landscape — a terrain of cost, in which every state a system could occupy sits at some height and every way of moving between states carries a price. This Part is about one such landscape, written down once and then asked three different questions. Ask it *which path* — what is the cheapest route from here to there? — and it answers with the law of least time that optics and mechanics have used for three centuries. Ask it *which flow* — what steady pattern of currents should a driven system maintain? — and it answers with the flux-and-force law of irreversible transport, and settles on the way the old quarrel over whether such systems minimize or maximize their dissipation: the two claims turn out to be one statement read from two sides. Ask it *which guess* — how should a system revise its internal estimate of the world it senses? — and it answers with the belief-updating rule that theoretical neuroscience calls the free energy principle. Three communities, three principles, one landscape asked three questions.

Why does it matter that the landscape is derived upstream, in the companion papers, rather than proposed here? Because a cost function chosen to fit its conclusions proves nothing — the fitting did the work. This one arrives with no dial left to tune: its two terms, their relative weight, and the geometry they live on are all fixed before any of the three questions is asked. That rigidity is what makes the failures derivable along with the successes: where a principle breaks down, the landscape says so, in the same breath and with the same authority as where it holds.

The route through the Part, in words: first the imported objects and their standing; then a short refresher on the mathematical tools, for readers who want them; then the master cost function itself; then the three readings taken in turn — path, flow, guess — each proved where it holds and proved where it fails; then a table grading every step; and last, closing remarks in plain language.

This Part contains the constructive content of the supplement. The strategy is fixed by the assessment: derive the three principles from one object that is itself derived; derive their restrictions as theorems; dissolve the minimum–maximum quarrel; replace the modeler-chosen structures with structures fixed by the framework; and grade every step. The functional is imported, not invented here — the derivations below are the deployment, in review form and at full length, of the three-gauge theorem established in Arisaka (2026, A7-Qi4), standing on the boundary law of Arisaka (2026, A5-GDOS) and the runtime of Arisaka (2026, A6-GRIP).

30 The imports, with their grades

Three objects are taken from the companion papers, and nothing else is assumed.

The boundary law. Arisaka (2026, A5-GDOS) derives, from the projection of the six-dimensional bulk (Arisaka, 2026, A2-GG6; Arisaka, 2026, A3-BI6) through the open-system channel demanded by the constitutional charter (Arisaka, 2026, A1-Constitution), the law governing any boundary observer's description of any system: a Nakajima–Zwanzig kernel, reducing in the Markovian limit to a GKLS semigroup, and in the mesoscopic limit to a continuity law on an effective state manifold $\mathcal{M}$ with coordinates z^a,

$$\partial_t \rho + \nabla_a J^a = 0, \tag{30.1}$$

with the current decomposed into four sectors — gradient, curl, noise, and gauge:

$$J^a \;=\; \underbrace{-\,\rho\, D^{ab}\,\partial_b \Phi}_{\text{gradient}} \;+\; \underbrace{\rho\, u_{\mathrm{c}}^a}_{\text{curl}} \;+\; \underbrace{-\,D^{ab}\partial_b \rho}_{\text{noise}} \;+\; \underbrace{\rho\, u_{\mathrm{g}}^a}_{\text{gauge}}, \qquad \Phi \equiv \frac{\delta F_{\mathsf{GFF}}}{\delta \rho}, \tag{30.2}$$

where D^{ab} is the positive-semidefinite mobility, u_{c} is divergence-free circulation, and u_{g} generates motion along dynamically maintained gauge directions. The four-sector form is not decoration; the review's counterexamples will land in the second and fourth sectors.

The budget. Arisaka (2026, A6-GRIP) establishes that the sectors descend one scalar ledger, unique up to additive constants,

$$F_{\mathsf{GFF}}[\rho] \;=\; F_{\mathsf{GEP}}[\rho] + F_{\mathsf{CEP}}[\rho] \;=\; \int_{\mathcal{M}} \rho\, U_{\mathrm{eff}}\, dz \;+\; \Theta_{\mathrm{eff}} \int_{\mathcal{M}} \rho \ln \frac{\rho}{\rho_{\mathrm{ref}}}\, dz, \tag{30.3}$$

an energetic–geometric term plus an effective information temperature times a relative entropy, with the balance identity

$$\frac{dF_{\mathsf{GFF}}}{dt} \;=\; -\,\mathcal{D}[\rho] \;+\; \mathcal{P}_{\mathrm{port}} \;+\; W_{\mathrm{c}} \;+\; W_{\mathrm{g}}, \qquad \mathcal{D}[\rho] = \int \rho\, \partial_a \Phi\, D^{ab} \partial_b \Phi\, dz \;\geq\; 0, \tag{30.4}$$

where $\mathcal{P}_{\mathrm{port}}$ collects explicit driving and boundary flux and $W_{\mathrm{c}}, W_{\mathrm{g}}$ are the work done by the curl and gauge sectors against the ledger. The monotonicity theorem of that paper states that in the lawful sector — ledger gauge-invariant, curl current tangent to the ledger's level sets — both W terms vanish, and in the undriven limit F_{GFF} is nonincreasing.

The runtime, and the grade. The dynamics on which all of this runs — exploration, commitment, capture; $\mathsf{DGI} \to \mathsf{GSSB} \to \mathsf{GA}$ — is the GRIP cascade of Arisaka (2026, A6-GRIP), whose closure is conditional on that paper's one remaining residue — the monodromy statement M11, in spectral-flow form; the budget bound M10 is resolved externally and the warped and relative refinements are closed at class level by Arisaka (2026, P2-N3). *Every construction below that stands on the cascade inherits that conditional status.* The functional itself (30.3)–(30.4) and the gauge derivations of Sections 33–35 require only the boundary law and the ledger; the statements that invoke halting or attractor selection require the cascade and carry its grade.

31 The reader's on-ramp: the mathematical tools, briefly

Five tools carry every derivation in this Part. Each is standard, each can be stated in a paragraph, and each is tied below to the section that uses it. A reader fluent in variational calculus and information geometry should skip ahead to Section 32.

Functionals and functional derivatives. A function eats a number and returns a number; a *functional* eats a whole function and returns a number — the total energy stored in a density profile, the total cost of a history. The *functional derivative* $\delta F/\delta\rho$ answers: if the input function is nudged by a small amount $\varepsilon\phi(z)$, how does the output number respond, point by point? Formally,

$$\frac{d}{d\varepsilon}F[\rho+\varepsilon\phi]\Big|_{\varepsilon=0} = \int_{\mathcal{M}} \frac{\delta F}{\delta\rho}(z)\,\phi(z)\,dz, \tag{31.1}$$

for every test function ϕ: the functional derivative is the density against which the nudge is integrated. Both terms of the ledger (30.3) can be differentiated in one line each. For the energetic term,

$$F_{\mathsf{GEP}}[\rho+\varepsilon\phi] = \int(\rho+\varepsilon\phi)\,U_{\text{eff}}\,dz = F_{\mathsf{GEP}}[\rho] + \varepsilon\int\phi\,U_{\text{eff}}\,dz \quad\Longrightarrow\quad \frac{\delta F_{\mathsf{GEP}}}{\delta\rho} = U_{\text{eff}}, \tag{31.2}$$

because the functional is linear in ρ. For the informational term the product rule acts once on each factor of $\rho\ln(\rho/\rho_{\text{ref}})$:

$$\frac{d}{d\varepsilon}\,\Theta_{\text{eff}}\int(\rho+\varepsilon\phi)\ln\frac{\rho+\varepsilon\phi}{\rho_{\text{ref}}}\,dz\Big|_{\varepsilon=0} = \Theta_{\text{eff}}\int\Big(\phi\ln\frac{\rho}{\rho_{\text{ref}}}+\rho\cdot\frac{\phi}{\rho}\Big)dz \quad\Longrightarrow\quad \frac{\delta F_{\mathsf{CEP}}}{\delta\rho} = \Theta_{\text{eff}}\Big(\ln\frac{\rho}{\rho_{\text{ref}}}+1\Big). \tag{31.3}$$

Adding the two gives the potential $\Phi = \delta F_{\mathsf{GFF}}/\delta\rho = U_{\text{eff}} + \Theta_{\text{eff}}(\ln(\rho/\rho_{\text{ref}})+1)$ that drives the gradient sector of (30.2) and appears throughout Sections 32 and 34. Note the shape of the answer: the energetic term contributes a fixed landscape, the informational term a crowding penalty that grows logarithmically where the density piles up above its reference.

Lagrange multipliers, up to a multiplier field. To extremize a function subject to a constraint, one does not eliminate a variable; one adds the constraint to the objective with an unknown coefficient and extremizes freely. The finite-dimensional prototype: minimize $f(x,y) = x^2+y^2$ subject to $x+y=1$. Form

$$\Lambda(x,y,\lambda) = x^2+y^2-\lambda\,(x+y-1), \qquad \partial_x\Lambda = 2x-\lambda = 0, \quad \partial_y\Lambda = 2y-\lambda = 0, \tag{31.4}$$

so $x=y=\lambda/2$, and the constraint fixes $x=y=\frac{1}{2}$, $\lambda=1$. The multiplier is not a bookkeeping trick: its value measures how hard the constraint pushes — how much the minimum would improve per unit of constraint relaxation. When the constraint holds at *every point of space and time* — as the continuity equation does in (32.1) — there is one multiplier per point, and the coefficient becomes a *multiplier field* $\lambda(z,t)$; the constraint enters the functional as $\int dt\int\lambda\,(\partial_t\rho+\nabla_a J^a)\,dz$, and an integration by parts moves the derivative from the varied quantity onto the multiplier.

This tool runs the stationarity computation of Section 32 and the constrained maximization inside Theorem 34.2.

The Legendre transform. Given a convex function $f(x)$, its Legendre transform trades the variable x for the slope $p = f'(x)$:

$$f^*(p) = \sup_x \left[\, p\, x - f(x)\,\right], \qquad \text{e.g.} \qquad f(x) = \tfrac{1}{2}ax^2 \implies f^*(p) = \frac{p^2}{2a}, \tag{31.5}$$

the supremum sitting at $p = ax$. Nothing is lost: transforming back recovers f. What changes is which member of a conjugate pair is treated as the control variable. Thermodynamics is built on this move: internal energy $U(S, V)$ is natural when volume is controlled, enthalpy $H(S, p) = U + pV$ when pressure is — the same physical information, reorganized for a different experimental handle. In Theorem 34.2 the flux-side and force-side entropy-production principles will turn out to be exactly such a conjugate pair: one holds the forces, the other holds the constitutive law and clamps part of the forces, and the quarrel between them evaporates the way a quarrel between energy and enthalpy would.

The Fisher information metric. A family of probability densities $q_\mu(z)$, labeled by parameters μ^i, is a surface whose points are distributions. The natural notion of distance on that surface asks: how *distinguishable* are two nearby members, given samples? The answer is the Fisher metric, $g_{ij} = \mathbb{E}_{q_\mu}[\,\partial_i \ln q_\mu\, \partial_j \ln q_\mu\,]$. The one-dimensional Gaussian makes it concrete: for $q_\mu(z) = (2\pi\sigma^2)^{-1/2} \exp[-(z-\mu)^2/2\sigma^2]$ with the mean μ as the parameter,

$$\partial_\mu \ln q_\mu = \frac{z-\mu}{\sigma^2}, \qquad g_{\mu\mu} = \mathbb{E}\left[\frac{(z-\mu)^2}{\sigma^4}\right] = \frac{\sigma^2}{\sigma^4} = \frac{1}{\sigma^2}. \tag{31.6}$$

A shift in the mean of a narrow distribution is easy to detect, so the metric is large at small σ; the same shift in a broad distribution hides in the noise, so the metric is small. In Section 35 this metric is not installed by convention, as the inference literature installs it, but derived: it is the Hessian of the ledger's own informational term.

Positive-definite Hessians and minima. At a stationary point $x_\star$ of a smooth function, the gradient vanishes and the Taylor expansion starts at second order,

$$f(x_\star + h) = f(x_\star) + \tfrac{1}{2}\, h^\mathsf{T} H\, h + O(h^3), \qquad H_{ij} = \partial_i \partial_j f(x_\star). \tag{31.7}$$

If H is *positive definite* — $h^\mathsf{T} H h > 0$ for every $h \neq 0$ — then every direction leads uphill and $x_\star$ is a strict minimum; negative definite, every direction leads downhill and it is a strict maximum; mixed signs, a saddle. Two facts recur below: a quadratic function equals its own second-order expansion, so for the quadratic objectives of Section 34 the local statement is global; and a principal submatrix of a positive-definite matrix is positive definite, which is the one-line reason the force-side minimum in Theorem 34.2 survives the clamping of part of the forces. The same logic, with sums replaced by integrals, classifies stationary points of functionals: there the Hessian is the kernel of the second variation.

32 The projected functional

The object everything else reads is the constrained transport functional on boundary histories (ρ, J) over $[0, \tau]$,

$$\boxed{\mathcal{A}_{\text{eff}}[\rho, J] \;=\; \int_0^\tau dt \left[\frac{1}{4} \int_{\mathcal{M}} (J - \rho u)^a \, (\rho D)^{-1}_{ab} \, (J - \rho u)^b \, dz \;+\; \mathcal{V}[\rho] \right], \qquad \partial_t \rho + \nabla_a J^a = 0,} \tag{32.1}$$

where u^a is the full lawful drift of (30.2) and $\mathcal{V}[\rho]$ is the scalar potential induced by the ledger. This is not a new postulate: it is the boundary image of the six-dimensional least-action principle, obtained by projecting the bulk action through the OGGEP channel — the derivation is carried out in Arisaka (2026, A7-Qi4), and the form (32.1) is also recognizable to a statistical physicist as the Onsager–Machlup–Freidlin–Wentzell cost of the boundary diffusion, which is precisely the point: the program derives from above the same functional that fluctuation theory constructs from below, and the two constructions meeting at (32.1) is what licenses everything that follows.

Stationarity with respect to J at fixed ρ is worth doing in full, because the same computation, with different decorations, runs every derivation in this Part. Adjoin the constraint with a multiplier field $\lambda(z, t)$ (second on-ramp tool),

$$\mathcal{A}_\lambda[\rho, J] \;=\; \mathcal{A}_{\text{eff}}[\rho, J] \;+\; \int_0^\tau dt \int_{\mathcal{M}} \lambda \left(\partial_t \rho + \nabla_a J^a \right) dz, \tag{32.2}$$

and nudge the current, $J \to J + \delta J$, at fixed ρ. Write the transport integrand as

$$Q(J) \;\equiv\; \frac{1}{4} (J - \rho u)^a \, (\rho D)^{-1}_{ab} \, (J - \rho u)^b; \tag{32.3}$$

it responds to the nudge, index by index,

$$Q(J + \delta J) - Q(J) = \frac{1}{4} \Big[\delta J^a \, (\rho D)^{-1}_{ab} \, (J - \rho u)^b + (J - \rho u)^a \, (\rho D)^{-1}_{ab} \, \delta J^b \Big] + O(\delta J^2), \tag{32.4}$$

and the two first-order pieces are equal: relabel $a \leftrightarrow b$ in the first and use the symmetry of $(\rho D)^{-1}$, so the linear response of the transport term is

$$\frac{1}{2} \int_0^\tau dt \int_{\mathcal{M}} (J - \rho u)^a \, (\rho D)^{-1}_{ab} \, \delta J^b \, dz. \tag{32.5}$$

The multiplier term responds through the divergence; one integration by parts (boundary flux vanishing on the lawful domain) moves the derivative onto the multiplier,

$$\int_{\mathcal{M}} \lambda \, \nabla_a \, \delta J^a \, dz \;=\; - \int_{\mathcal{M}} \partial_a \lambda \, \delta J^a \, dz. \tag{32.6}$$

Since δJ is arbitrary, the coefficient of δJ^a must vanish pointwise, and contracting the resulting condition with $2\rho D^{ca}$ (which inverts $(\rho D)^{-1}$) solves it for the current:

$$\frac{1}{2}(\rho D)^{-1}_{ab}(J-\rho u)^b - \partial_a \lambda = 0 \quad \Longrightarrow \quad J^a = \rho u^a + 2\rho D^{ab}\partial_b \lambda, \tag{32.7}$$

so histories of zero cost are exactly the lawful flows ($\lambda = \text{const}$), and the cost of any other history is the squared, mobility-weighted distance of its current from the lawful one.

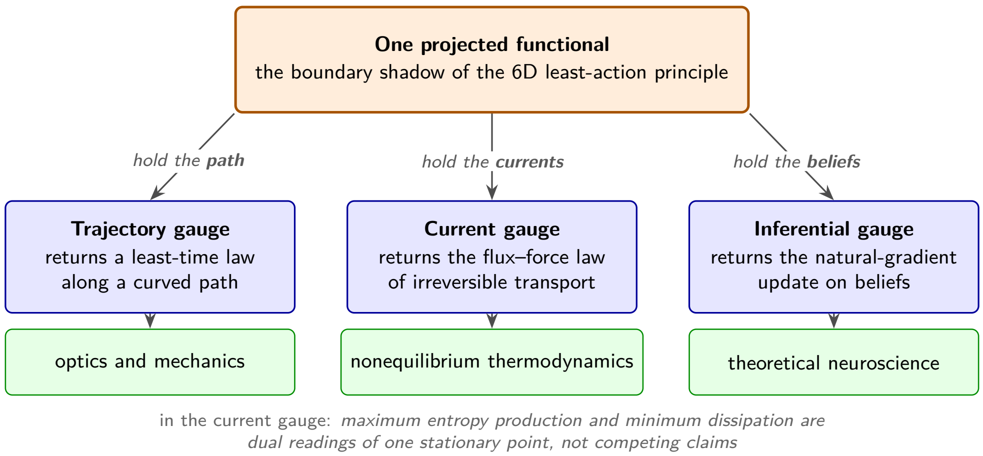


Figure 7: How three principles that look like rivals turn out to be one. The single projected functional at the top is the boundary image of the six-dimensional least-action principle; it is not a new postulate. Each arrow is the same stationarity condition asked with a different quantity held primary, and the label on the arrow names which: hold the path and the condition returns a least-time law, hold the currents and it returns the flux–force relation of irreversible transport, hold the beliefs and it returns the natural-gradient update used in the brain sciences. The bottom row names the community that has been using each form. The dashed link records what the unification settles on the way through: maximum entropy production and minimum dissipation are dual readings of one stationary point under dual control conventions, not competing claims about nature. One system may occupy all three gauges at once — a cell following a gradient traces a path, maintains fluxes, and corrects an estimate — and the three Parts of review in this supplement each live at the foot of one arrow. Carried from the Qi-EP companion, where the theorem this figure draws is established.

Three different restrictions of (32.1) — three choices of which variable is held primary — now return the three principles. The program calls these the *gauges* of Qi-EP, and the terminology is earned: each is a coordinate choice on one object, and nothing physical distinguishes them.

33 The trajectory gauge: least time

Hold the *path* primary: restrict (32.1) to sharply concentrated densities $\rho(z,t) = \delta(z - z(t))$ — the deterministic (small-noise) limit — so the functional collapses to an action for the single trajectory $z(t)$. Evaluating the transport term on the concentrated density and passing to the fixed-energy (Maupertuis) form yields the generalized optical-time functional

$$\mathcal{T}[z] \;=\; \int_\gamma \Big[\sqrt{2(E-U_{\text{eff}})\, g_{ab}\, dz^a dz^b} \;-\; a_a\, dz^a\Big], \tag{33.1}$$

with g_{ab} the state-manifold metric and a_a the projected gauge potential. Its Euler–Lagrange equation is geodesic motion in the Jacobi metric $2(E - U_{\mathrm{eff}})\, g_{ab}$ with a Lorentz-like curvature force from da and a scalar force from $\partial U_{\mathrm{eff}}$ — which is Fermat's principle when U_{eff} encodes a refractive index, and Maupertuis–Hamilton mechanics in general. This is the principle of least time, PLT, and the derivation shows its exact domain: it is the reading of (32.1) valid when the density stays concentrated, i.e., at high visibility and weak noise. Where diffusion competes with drift, the trajectory gauge is simply the wrong coordinate choice — not a failed law.

34 The current gauge: the flux–force law, and both entropy-production principles at once

Hold the *currents* primary: work at a fixed density profile (a steady state maintained by clamped boundary conditions) and vary J alone. Writing the thermodynamic force $X_a = -\,\partial_a \Phi$ and the Onsager operator $L^{ab} = \rho D^{ab}/\Theta_{\mathrm{eff}}$, the transport term of (32.1) becomes, up to a constant, the negative of the *Onsager–Rayleighian*

$$\mathcal{R}_\rho[J;X] \;=\; \int_{\mathcal{M}} \Big(J^a X_a - \tfrac{1}{2}\, J^a (L^{-1})_{ab} J^b \Big)\, dz. \tag{34.1}$$

Theorem 34.1 (Current-gauge stationarity)**.** *$\mathcal{R}_\rho$ is stationary in J exactly at the flux–force law $J^a = L^{ab} X_b$; the second variation $\delta^2 \mathcal{R}_\rho = -\,(L^{-1})_{ab} \prec 0$, so the stationary point is a maximum over fluxes; and at stationarity $\mathcal{R}_{\rho,\star} = \frac{1}{2}\dot{\Sigma}_\star$, half the entropy production of the selected state.*

Proof. Nudge the flux, $J \to J + \delta J$, and expand (34.1). The linear term in the entropy-production part is immediate; in the quadratic part the two cross pieces combine by the symmetry of L^{-1}, exactly as in (32.4):

$$\tfrac{1}{2}\big[\, \delta J^a (L^{-1})_{ab} J^b + J^a (L^{-1})_{ab}\, \delta J^b \,\big] = \delta J^a (L^{-1})_{ab} J^b. \tag{34.2}$$

Hence

$$\mathcal{R}_\rho[J + \delta J] - \mathcal{R}_\rho[J] = \int_{\mathcal{M}} \big(X_a - (L^{-1})_{ab} J^b\big)\, \delta J^a \, dz - \tfrac{1}{2} \int_{\mathcal{M}} \delta J^a (L^{-1})_{ab}\, \delta J^b \, dz, \tag{34.3}$$

with *no* higher terms: the functional is quadratic, so this expansion is exact (fifth on-ramp tool). The first variation vanishes for all δJ if and only if

$$\frac{\delta \mathcal{R}_\rho}{\delta J^a} = X_a - (L^{-1})_{ab} J^b = 0 \quad \Longleftrightarrow \quad J^a = L^{ab} X_b, \tag{34.4}$$

the flux–force law. The second variation is the exact remainder,

$$\delta^2 \mathcal{R}_\rho = -\int_{\mathcal{M}} \delta J^a (L^{-1})_{ab}\, \delta J^b \, dz \;<\; 0 \qquad \text{for } \delta J \neq 0, \tag{34.5}$$

since $L \succ 0$ implies $L^{-1} \succ 0$: the Hessian kernel is $-(L^{-1})$, negative definite, so the stationary point is a strict global maximum. Finally, substitute the law back: at $J_\star = LX$ the identity

$(L^{-1})_{ab}J^b_\star = X_a$ turns the quadratic term into half the linear one,

$$\mathcal{R}_{\rho,\star} = \int_{\mathcal{M}} \Big(J^a_\star X_a - \tfrac{1}{2}\, J^a_\star X_a \Big) dz = \tfrac{1}{2} \int_{\mathcal{M}} J^a_\star X_a \, dz = \tfrac{1}{2}\, \dot{\Sigma}_\star, \tag{34.6}$$

half the entropy production of the selected state. ■

What just happened. We let the currents wander freely and asked where the Rayleighian levels off. It levels off exactly at the ordinary linear transport law, and because the surface curves downward in every flux direction, the law sits at the top of a hill, not the bottom of a valley. The height of that hill is half the entropy production — which is why factors of one half haunt every dissipation principle descended from Rayleigh and Onsager.

Theorem 34.1 is Onsager's principle of least dissipation, recovered from the projected functional rather than postulated — and both entropy-production principles of the review now fall out of it as *the two readings of this one stationary point.*

Theorem 34.2 (The duality: MEPP and MinEP as conjugate conventions)**.** *Let $\dot{\Sigma}(J) = J^a (L^{-1})_{ab} J^b$ be the flux-side entropy production and $P(X) = X_a L^{ab} X_b$ the force-side one, with $L = L^{\mathsf{T}} \succ 0$. (i)* Flux side, forces clamped: *among all fluxes satisfying the power-balance constraint $\dot{\Sigma}(J) = J^a X_a$, the physical flux $J = LX$ maximizes $\dot{\Sigma}$, and the maximizer satisfies Ziegler's orthogonality $X_a = \lambda\, \partial\dot{\Sigma}/\partial J^a$ with $\lambda = \frac{1}{2}$ fixed by the constraint. This is the MEPP / Ziegler statement, and it* selects the constitutive relation*. (ii)* Force side, constitutive law fixed: *split the forces into clamped $X_1, \ldots, X_m$ and free $X_{m+1}, \ldots, X_n$; then on the affine subspace of clamped forces, $\partial P/\partial X_j = 2L_{jk}X_k = 2J_j$ for free j, so the stationary point of P is exactly the state with vanishing free-force fluxes — the steady state — and the Hessian $2L_{jj'}$ restricted to the free block is positive definite, so it is a minimum. This is Prigogine's theorem, and it* selects the state*. (iii) The two are Legendre-dual descriptions of the single stationary point of $\mathcal{R}_\rho$: (i) varies fluxes at fixed forces, (ii) varies free forces at fixed constitutive law, and $\dot{\Sigma}(J_\star) = P(X_\star)$ at the common point.*

Proof. (i) *Flux side.* The problem is to maximize $\dot{\Sigma}(J) = J^a (L^{-1})_{ab} J^b$ subject to the power-balance constraint $g(J) \equiv J^a X_a - \dot{\Sigma}(J) = 0$. Introduce a multiplier μ and extremize $\Lambda(J,\mu) = \dot{\Sigma}(J) + \mu\, g(J)$ freely. First the derivative of $\dot{\Sigma}$, index by index, with the symmetry of L^{-1} combining the two pieces as in (34.2):

$$\frac{\partial \dot{\Sigma}}{\partial J^c} = (L^{-1})_{cb} J^b + J^a (L^{-1})_{ac} = 2\,(L^{-1})_{cb} J^b. \tag{34.7}$$

The stationarity condition of Λ is then

$$2(L^{-1})_{cb}J^b + \mu\big(X_c - 2(L^{-1})_{cb}J^b\big) = 0 \quad \Longrightarrow \quad 2(L^{-1})_{cb}J^b = \frac{\mu}{\mu - 1}\, X_c \qquad (\mu \neq 1), \tag{34.8}$$

so every candidate extremum is proportional to the physical flux,

$$J^b = k\, L^{bc} X_c, \qquad k \equiv \frac{\mu}{2(\mu - 1)}. \tag{34.9}$$

The constraint fixes k: on the candidate,

$$J^a X_a = k\, X_a L^{ab} X_b, \qquad \dot{\Sigma}(J) = k^2\, X_a L^{ab} X_b, \qquad J^a X_a = \dot{\Sigma}(J) \implies k = k^2 \implies k = 1 \tag{34.10}$$

(the root $k = 0$ is the trivial equilibrium point $J = 0$, excluded for nonzero X by maximality below). Hence $J = LX$, the linear law. For Ziegler's orthogonality, note that $\dot{\Sigma}$ is homogeneous of degree two, so Euler's homogeneous-function theorem gives

$$J^a \frac{\partial \dot{\Sigma}}{\partial J^a} = 2\,\dot{\Sigma}(J); \tag{34.11}$$

contracting the orthogonality ansatz $X_a = \lambda\, \partial\dot{\Sigma}/\partial J^a$ with J^a and using the constraint,

$$\dot{\Sigma} = J^a X_a = \lambda\, J^a \frac{\partial \dot{\Sigma}}{\partial J^a} = 2\lambda\,\dot{\Sigma} \quad \implies \quad \lambda = \tfrac{1}{2}, \tag{34.12}$$

the advertised value, fixed by the constraint and by nothing else. That the extremum is a maximum is seen by expanding along the constraint surface: parameterize $J = LX + \delta$ and impose $g(J) = 0$ exactly,

$$(LX + \delta)^a X_a = (LX + \delta)^a (L^{-1})_{ab} (LX + \delta)^b \quad \implies \quad \delta^a X_a = -\,\delta^a (L^{-1})_{ab}\,\delta^b, \tag{34.13}$$

after canceling XLX from both sides and collecting the terms linear and quadratic in δ. Substituting this admissibility condition back into $\dot{\Sigma}$,

$$\dot{\Sigma}(LX + \delta) = X_a L^{ab} X_b + 2\,\delta^a X_a + \delta^a (L^{-1})_{ab}\,\delta^b = \dot{\Sigma}(LX) - \delta^a (L^{-1})_{ab}\,\delta^b \;\le\; \dot{\Sigma}(LX), \tag{34.14}$$

with equality only at $\delta = 0$, by positive definiteness of L^{-1}: every admissible departure from the linear law strictly lowers the entropy production, so $J = LX$ is the strict constrained maximum.

(ii) *Force side.* Display the force-side production and differentiate with respect to a free force X_j, $j > m$, the clamped forces held fixed:

$$P(X) = X_a L^{ab} X_b, \qquad \frac{\partial P}{\partial X_j} = \delta^a_j\, L^{ab} X_b + X_a L^{ab}\,\delta^{\,b}_j = (L^{jb} + L^{bj}) X_b = (L + L^{\mathsf{T}})^{jb} X_b. \tag{34.15}$$

The symmetrization step is where the physics enters: Onsager symmetry $L = L^{\mathsf{T}}$ collapses the combination,

$$(L + L^{\mathsf{T}})^{jb} X_b \;\overset{L=L^{\mathsf{T}}}{=}\; 2\, L^{jb} X_b = 2\, J^j, \tag{34.16}$$

so the gradient of P in the free directions is (twice) the physical flux, and the stationary point of P is exactly the state with vanishing free-force fluxes — the steady state. Without symmetry the gradient of P is $(L + L^{\mathsf{T}})X$, which is not the physical flux LX, and the theorem fails — which is the parity restriction of the review, now visible as algebra. The character of the extremum follows from a second differentiation,

$$\frac{\partial^2 P}{\partial X_j\,\partial X_l} = 2\,L^{jl}, \qquad j,l = m+1,\dots,n, \tag{34.17}$$

twice the free-block principal submatrix of L; a principal submatrix of a positive-definite matrix is positive definite (fifth on-ramp tool), so the steady state is a strict minimum of P on the affine subspace of clamped forces.

(iii) *The Legendre bridge.* The two sides are conjugates of one bilinear form (third on-ramp tool). Compute the transform of $\dot\Sigma$ at the doubled pairing:

$$\sup_J \big[\,2J^a X_a - \dot\Sigma(J)\,\big] \quad \text{is attained where} \quad 2X_a = 2(L^{-1})_{ab}J^b, \quad \text{i.e. } J = LX, \tag{34.18}$$

and the value there is

$$2\,X_a L^{ab} X_b - X_a L^{ab} X_b = X_a L^{ab} X_b = P(X)\,: \tag{34.19}$$

the force-side production *is* the Legendre transform of the flux-side production, and at the shared stationary point $\dot\Sigma(J_\star) = P(X_\star)$, the common value. Reading (i) fixes the forces and varies the fluxes; reading (ii) fixes the constitutive law and varies the free forces; the two extremizations are over complementary variables of the same bilinear form. ■

What just happened. Two communities have been computing at the same stationary point while holding different things fixed. Fix the forces and vary the fluxes, and the extremum hands you the transport law — that is the maximum-entropy-production reading. Fix the transport law and vary the free forces, and the extremum hands you the steady state — that is Prigogine's minimum. The relation between them is the same as between energy and enthalpy: one Legendre transform apart, carrying identical information under different experimental handles, so no measurement can side with one against the other.

Remark 34.3 (What the duality settles). The seventy-year quarrel of the review dissolves without a winner: MEPP and MinEP are not rival claims about nature but conjugate conventions about what is held fixed at one stationary point — the flux-side reading fixes the transport law, the force-side reading fixes the state. This is the program's derivation of exactly the resolution that the critical literature (Polettini's constitutive-versus-phenomenological dichotomy; the circuit analysis of Bruers, Maes and Netočný) reached from below, and the agreement of the two routes is a check, not a coincidence.

Theorem 34.4 (Where the principles must fail)**.** *The reduction of* (32.1) *to the Rayleighian* (34.1) *used two properties of the lawful drift: the curl sector does no work against the ledger,* $W_\mathrm{c} = \int \rho\, u_\mathrm{c}^a\, \partial_a \Phi\, dz = 0$*, and the ledger is gauge-invariant,* $W_\mathrm{g} = 0$*. Where either fails, no functional of the entropy-production family is stationary at the physical state, and both extremum principles fail together.*

Proof. Split the lawful drift of (30.2) into its dissipative part and the two retained sectors,

$$\rho u^a = \rho u_\mathrm{d}^a + \rho u_\mathrm{c}^a + \rho u_\mathrm{g}^a, \tag{34.20}$$

where u_{d} collects the gradient and noise sectors, and carry the split through the first variation (32.5) of the transport cost (the constant factor Θ_{eff} converting $(\rho D)^{-1}$ to L^{-1} is absorbed throughout):

$$\delta\mathcal{A} = \tfrac{1}{2}\int (J-\rho u_{\mathrm{d}})^a (L^{-1})_{ab}\,\delta J^b\,dz \;-\; \tfrac{1}{2}\int (\rho u_{\mathrm{c}}+\rho u_{\mathrm{g}})^a (L^{-1})_{ab}\,\delta J^b\,dz. \tag{34.21}$$

The first integral is the Rayleighian variation of Theorem 34.1; the second is the cross term the quadratic reduction discarded. Setting the total to zero for all admissible δJ gives the physical stationary current

$$J_\star^a = L^{ab}X_b + \rho u_{\mathrm{c}}^a + \rho u_{\mathrm{g}}^a. \tag{34.22}$$

Now test whether any entropy-production functional is extremal there. Expand $\dot\Sigma(J) = J^a(L^{-1})_{ab}J^b$ about $J_\star$, using (34.7):

$$\dot\Sigma(J_\star+\delta)-\dot\Sigma(J_\star) = 2\,\delta^a(L^{-1})_{ab}J_\star^b+\delta^a(L^{-1})_{ab}\,\delta^b = 2\,\delta^a X_a + 2\,\delta^a(L^{-1})_{ab}(\rho u_{\mathrm{c}}+\rho u_{\mathrm{g}})^b + \delta^a(L^{-1})_{ab}\,\delta^b. \tag{34.23}$$

The first term is removed by admissibility exactly as in (34.13); the third is the benign quadratic. The middle term is the obstruction: it is *linear* in δ, of either sign, and survives at $J_\star$ unless $\rho u_{\mathrm{c}} + \rho u_{\mathrm{g}}$ is L^{-1}-orthogonal to every admissible variation. A stationary point with a surviving linear term is no extremum. To see that the orthogonality condition is the work condition, take the admissible variations generated by shifting the potential, $\delta J^a = L^{ab}\,\partial_b\chi$; then

$$\delta^a(L^{-1})_{ab}(\rho u_{\mathrm{c}}+\rho u_{\mathrm{g}})^b = \partial_a\chi\,(\rho u_{\mathrm{c}}+\rho u_{\mathrm{g}})^a, \qquad \chi = -\,\Phi \implies \int(\cdots)\,dz = -\,(W_{\mathrm{c}}+W_{\mathrm{g}})\big/\Theta_{\mathrm{eff}}, \tag{34.24}$$

so the vanishing of the cross term for this family of variations is precisely the condition $W_{\mathrm{c}} = W_{\mathrm{g}} = 0$: where either sector does work against the ledger, no functional of the entropy-production family is stationary at the physical state, and both extremum principles fail together. ■

What just happened. The theorem does not say the extremum principles sometimes fail; it computes the exact tax that makes them fail. When circulation or gauge motion does work against the ledger, a term linear in the variation survives at the physical state — and a point with a surviving linear term sits on a slope, not on a summit. Landauer's circuit, Jaynes's network, and the frenetic obstruction of the modern fluctuation literature are three names for that one slope.

Remark 34.5 (The counterexamples, located)*.* Theorem 34.4 is the framework's account of every counterexample assembled in the review. Landauer's driven circuit fails MinEP because an inductive current is odd under time reversal: in boundary coordinates it is curl-sector circulation that does work, $W_{\mathrm{c}} \neq 0$. Jaynes's two-temperature network fails because a nonuniform temperature field makes the ledger's level sets and the circulation non-orthogonal — in the language of (30.4), the coefficients cease to be constant across the system and the work terms survive. And the frenetic obstruction — the modern statement that selection beyond linear order depends on time-symmetric activity no entropy-production functional contains — is the observation that W_{c} and W_{g} are time-symmetric sector couplings invisible to $\dot\Sigma$. The program does not rescue the extremum principles from these cases; it derives that they must fail there, which is what a correct derivation owes.

Remark 34.6 (Beyond quadratic order). The Rayleighian (34.1) is the quadratic restriction of (32.1); the full functional is not quadratic, and its stationarity conditions beyond quadratic order are not expressible as an extremum of any entropy-production functional — in agreement with, and for the same structural reason as, the macroscopic-fluctuation-theory result that what survives far from equilibrium is a corrected minimum-cost principle for the full functional, not an entropy extremum. The linear-regime confinement of both principles is thus a theorem of the projection, not an empirical accident.

35 The inferential gauge: the free energy principle

Hold the *beliefs* primary. Let the boundary density be parameterized on a statistical manifold: $\rho = q_\mu$, coordinates μ^i — the internal states of an agent in the vocabulary of Part IV, the registered parameters of the system packet in the program's own. Two computations turn (32.1) into the free energy principle's working core.

The metric is derived, not chosen. The claim is that the second variation of the ledger's informational term (30.3) in the belief coordinates is the Fisher metric of the fourth on-ramp tool, and the computation deserves its three lines. Write $h(\mu) = D_{\mathrm{KL}}[\, q_\mu \,\|\, q_{\mu_0}] = \int q_\mu \ln(q_\mu/q_{\mu_0})\, dz$ and expand about the base point μ_0. The first derivative, by the product rule, is

$$\frac{\partial h}{\partial \mu^i} = \int \partial_i q_\mu \, \ln \frac{q_\mu}{q_{\mu_0}}\, dz + \int q_\mu \, \partial_i \ln q_\mu \, dz, \tag{35.1}$$

and the second integral vanishes identically in μ by the score identity — differentiating the normalization of q_μ under the integral sign,

$$\int q_\mu \, \partial_i \ln q_\mu \, dz = \int \partial_i q_\mu \, dz = \partial_i \int q_\mu \, dz = \partial_i(1) = 0. \tag{35.2}$$

At $\mu = \mu_0$ the logarithm in the first integral of (35.1) is $\ln 1 = 0$, so the whole first derivative vanishes at the base point: the first-order term of the expansion is zero, as it must be for a divergence that attains its minimum, zero, at $\mu = \mu_0$. The metric therefore comes from the second derivative. Differentiating (35.1) again (the score term contributes nothing, being identically zero),

$$\frac{\partial^2 h}{\partial \mu^j \partial \mu^i} = \int \partial_j \partial_i q_\mu \, \ln \frac{q_\mu}{q_{\mu_0}}\, dz + \int \partial_i q_\mu \, \partial_j \ln q_\mu \, dz, \tag{35.3}$$

and at $\mu = \mu_0$ the first integral again dies on $\ln 1 = 0$, while the second, with $\partial_i q = q\, \partial_i \ln q$, is exactly the expectation of the product of scores. The second variation at the base point is therefore

$$\frac{\partial^2}{\partial \mu^i \partial \mu^j} D_{\mathrm{KL}}[q_\mu \,\|\, q_{\mu_0}] \Big|_{\mu=\mu_0} \;=\; \int q_{\mu_0} \, \partial_i \ln q_{\mu_0} \, \partial_j \ln q_{\mu_0} \, dz \;\equiv\; g^F_{ij}(\mu_0), \tag{35.4}$$

the Fisher information metric — so the geometry that the inference literature installs by convention arrives here as the Hessian of the CEP term. Restricting (32.1) to the manifold gives the inferential action

$$\mathcal{A}_{\inf}[\mu] \;=\; \int_0^\tau dt\,\Big[\tfrac{1}{2}\,g^F_{ij}(\mu)\,\dot\mu^i\dot\mu^j \;+\; \mathcal{F}(\mu;s)\Big]\,, \qquad \mathcal{F}(\mu;s) \;=\; F_{\mathsf{GFF}}[q_\mu]\big|_{\text{ports }s}, \tag{35.5}$$

The working equations are the overdamped limit of this action, and the limit is worth taking explicitly. The Euler–Lagrange equation of (35.5) — vary $\mu^i(t)$, integrate the $\dot\mu$-term by parts, remember that the metric depends on μ — is

$$g^F_{ij}\,\ddot\mu^j + \Big(\partial_k g^F_{ij} - \tfrac{1}{2}\,\partial_i g^F_{jk}\Big)\dot\mu^k\dot\mu^j \;=\; \partial_i\mathcal{F}(\mu;s), \tag{35.6}$$

a second-order equation: an inertial term, a metric (Christoffel-type) term quadratic in the velocity, and the gradient of the free energy. The overdamped regime is the familiar one from mechanics. In the prototype

$$m\,\ddot x + \gamma\,\dot x + V'(x) = 0, \qquad \frac{|m\ddot x|}{|\gamma\dot x|} \sim \frac{m}{\gamma\,\tau} \;\longrightarrow\; 0 \quad \text{as } \gamma\tau \gg m, \tag{35.7}$$

a particle in molasses forgets its inertia: on histories varying over a timescale τ long compared with m/γ, the acceleration term is down by one power of the timescale ratio, and the surviving balance is friction against force, $\gamma\dot x = -V'(x)$. Here the role of the friction is played by the mobility that the transport term of (32.1) carries onto the manifold — the same $(\rho D)^{-1}$ weight, restricted to belief coordinates, that sets the rate κ at which the registered states exchange with the ports. On relaxational histories, slow compared with that port-coupling time, $|\ddot\mu| \sim |\Delta\mu|/\tau^2$ while $|\dot\mu| \sim |\Delta\mu|/\tau$, so both left-hand terms of (35.6) — the inertial term and the velocity-quadratic metric term — are suppressed by one power of the timescale ratio relative to the friction–gradient balance, and dropping them leaves the first-order flow

$$\dot\mu^i \;=\; -\,\kappa\,(g^F)^{ij}\,\partial_j\mathcal{F}(\mu;s), \qquad \dot a^\alpha \;=\; -\,\kappa_a\,\frac{\partial\mathcal{F}}{\partial a^\alpha}, \tag{35.8}$$

natural-gradient descent of a free energy in the internal states, and gradient descent in the action variables: the perception and action equations of the free energy principle, with the bound $\mathcal{F} \geq -\ln p(s)$ holding by the same one-line Jensen argument as in Part IV, since $\mathcal{F}$ differs from the variational free energy by the identification of the reference density with the generative model's marginal.

What changes, relative to the construction criticized in Part IV, is everything the critics attacked.

- *The blanket is replaced by declared ports.* The program does not assume that a statistical boundary exists in the stationary density of a generic system — the claim shown to be non-generic. The system–environment split enters as *constitutive data*: the ports of the system packet SIP (Arisaka, 2026, A6-GRIP), declared with the system and carrying, in the packet's final field, the criteria under which the description is judged wrong. Where the free energy literature's boundary is a partition found (or not) in a stationary density, the program's is part of what it means to write the system down — the same move by which thermodynamics escapes the ambiguity of "the system" by declaring its walls.

- *The synchronization map is replaced by a rank condition.* The contested lemma — that internal states parameterize a posterior over external ones — is not assumed. The program's internality construction (Arisaka, 2026, A7-Qi4) makes the internal model's existence *conditional and checkable*: the external perturbation is recoverable from the sensory history if and only if an explicit action-conditioned observability matrix has full column rank, and where the rank fails, no unique internal model exists. The finding of the linear-systems critique — that generic coupled systems do not satisfy the inference reading — is thereby absorbed as a classification: such systems sit below the rank condition, and the framework says so rather than assuming otherwise.

- *The solenoidal obstruction is the same obstruction.* The couplings that had to be silently assumed away in the blanket construction are, in boundary coordinates, curl-sector couplings — the same W_c of Theorem 34.4. The inferential gauge holds exactly where the current gauge's extremum principles hold, and fails with them, for the same reason. One obstruction, three literatures.

- *What is not derived is not claimed.* The expected free energy of the active-inference process theories does not follow from (35.5), just as it does not follow from the variational free energy in the original program — a point the inference literature itself has established. In the present framework, prospective policy selection is carried by a different mechanism (the internalized regime switch of the runtime), and no claim is made that (35.8) contains it.

36 The three-gauge theorem, and the grade of every step

Assembling Sections 33–35:

Theorem 36.1 (Three gauges of one projected functional; deployment of the Qi-EP principal theorem)**.** *The constrained functional* (32.1)*, projected from the bulk action through the GDOS channel, returns under three restrictions: (i) with the path held primary, the least-time law* (33.1) *(*PLT*: optics and mechanics); (ii) with the currents held primary, the flux–force law (Theorem 34.1), carrying MEPP and MinEP as its flux-side and force-side readings (Theorem 34.2), valid exactly where the curl and gauge sectors do no ledger work (Theorem 34.4); (iii) with the beliefs held primary, the natural-gradient inference law* (35.8) *(*FEP*), on the Fisher geometry derived from the ledger's own informational term, under the same sector condition and the rank condition of the internality construction.*

The status of each element, in the program's grading:

37 The five criteria, checked

Against the specification of Part V (§23): (i) the three principles are derived from one functional that is itself derived — the projection of an upstream action, not a postulate; (ii) the restrictions arrive as theorems — the linear-regime confinement (Remark 34.6), the parity and inhomogeneity failures (Theorem 34.4), the coupling and rank conditions of the inference reading; (iii) the minimum–maximum quarrel is dissolved by Theorem 34.2, in agreement with the resolution the critical literature

Table 5: The grade of every step in the derivation chain of this Part, in the program's standard grading. "Derived" means proved here or in the cited companion from the stated imports; "conditional" names the condition; "argued" means supported but not proved, and stated as such; "not claimed" marks the adjacent results this Part deliberately does not assert. The right-hand column is the load-bearing one: a review whose subject is three literatures' overclaims must not add its own, and this table is where the boundary of the present claim is drawn.

Statement	**Grade**	**Condition / authority**
Boundary continuity law and four-sector current	Derived	Arisaka (2026, A5-GDOS); Arisaka (2026, A7-Qi4)
Single scalar ledger $F_{\mathsf{GFF}} = F_{\mathsf{GEP}} + F_{\mathsf{CEP}}$; balance identity	Derived	Arisaka (2026, A6-GRIP)
Projected functional (32.1) from the bulk action	Derived	Arisaka (2026, A7-Qi4), via OGGEP
Trajectory gauge ⇒ PLT	Derived	small-noise limit; this Part
Current gauge ⇒ flux–force law; least dissipation	Derived	Theorem 34.1
MEPP/MinEP duality at one stationary point	Derived	Theorem 34.2
Failure locations of both principles (parity, inhomogeneity, nonlinear order)	Derived	Theorem 34.4; Remarks 34.5–34.6
Inferential gauge ⇒ natural-gradient FEP; Fisher metric from CEP	Derived	(35.4)–(35.8)
Internal model existence	Conditional	rank condition of the internality construction (Arisaka, 2026, A7-Qi4)
Statements invoking halting or attractor selection	Conditional	GRIP closure, conditional on M11 (spectral-flow form) (Arisaka, 2026, A6-GRIP); M10 resolved, scope refinements closed (Arisaka, 2026, P2-N3)
Exhaustiveness of the three gauges	Argued	from the packet's structure (state, flow, registration); not proved
Expected free energy / policy selection from (35.5)	Not claimed	does not follow; carried by a different mechanism
Universal validity of MEPP	Not claimed	false, and derived to be false outside its regime

reached from below; (iv) the modeler-chosen structures are replaced — the choice of which entropy production by the ledger fixed upstream, the blanket by the declared ports of the packet, the synchronization map by a rank condition; and (v) the conditional and open elements are stated in Table 5 and nowhere waived. What a reader should retain is not that the three principles are true — each is true only where its regime theorem says — but that the three regimes, the three laws, and the three failures now follow from one object.

38 Closing remarks: one object, three shadows

Return to the opening picture, which can now be completed. There is one landscape of cost. Shine a light on it from three directions and it casts three shadows. Ask which path, and the shadow is the law of least time — the rule by which light finds its way through a lens and a ball finds its arc through the air. Ask which flow, and the shadow is the flux-and-force law of irreversible transport, carrying with it, as two readings of one stationary point, both of the entropy-production principles that spent seventy years being defended against each other. Ask which guess, and the shadow is the belief-updating rule of the brain sciences, moving an internal estimate downhill on a landscape whose very geometry — the metric that says which estimates count as close — comes from the cost function itself rather than from a modeler's taste. Three shadows, three literatures, one object. None of the shadows is the object, and each is faithful only from its own angle.

The deeper lesson of this Part is that deriving the failures matters more than deriving the successes. Any framework can be tuned until it reproduces a law that is known to hold; the fitting does the work, and the agreement proves little. What cannot be tuned is a failure. When the same object that yields the three principles also computes, with no additional freedom, exactly where each one must break — the coil in the circuit, the network of warm and cold resistors, the strongly driven chemistry, the system whose senses cannot pin down its world — then the boundaries arrive with the same authority as the laws, and the counterexamples that once looked like wreckage become confirmations. A theory earns trust at the places where it says no.

And the cell? A cell following a nutrient gradient traces a path, maintains a pattern of flows, and corrects an estimate of where the food is — all at once, all the time. Nothing in this Part says the cell is pulled in three directions by three competing imperatives, and nothing could: the three descriptions are one description, and a system cannot be in conflict with itself over what is, mathematically, a choice of coordinates. Where the three readings seem to disagree about a living system, the disagreement is in the reader, not in the cell — or else the system has wandered outside the gentle regime where the readings apply, and the honest statement is not that one principle defeated another but that all three fell silent together, at a boundary this Part has drawn rather than blurred.

What should remain, after the algebra fades, is the economy of the thing: not three principles, patched and policed at their edges, but one object, asked three questions, answering each only as far as the question deserves. A landscape does not argue with its shadows.

Part VIII
The Gauges in the Laboratory and the Field

Seven real observations, read through the derived gauges, each with the measurement that could refute the reading

This Part is where the argument meets things you can point at. So far the three principles and their proposed common source have lived in functionals and theorems; here they must earn their keep against a thermodiffusion cell, a resistor network, a pan of convecting fluid, three planetary atmospheres, the enzyme that makes ATP, a swimming bacterium, and a predicting brain. For each case the format is the same and deliberately blunt: what is observed, how the old principle read it, how the derived functional reads it, and — the column that matters most — what measurement would prove the new reading wrong. A life-science reader should feel at home in at least three of the seven: the ATP machine, the bacterium, and the brain are the cases where the framework's claims come closest to a wet lab, and they are written to be readable on their own.

None of the seven is offered as a confirmation; they are placements, and the discipline of the program is to say which rows could bite.

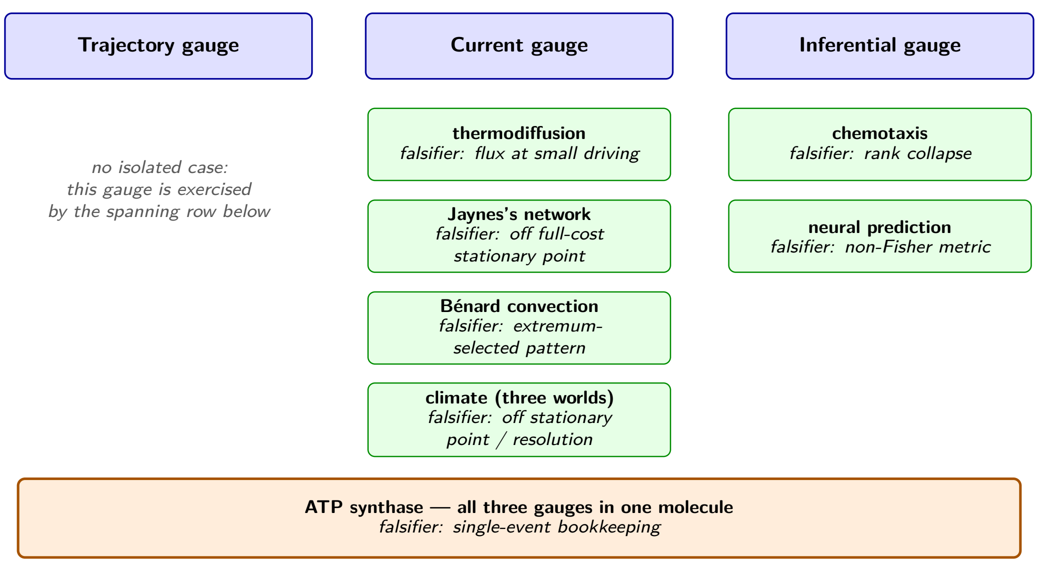


Figure 8: The case map of this Part. Each of the seven observations is placed under the gauge whose reading it exercises: thermodiffusion, Jaynes's network, Bénard convection, and planetary climate under the current gauge; chemotaxis and neural prediction under the inferential gauge; and ATP synthase spanning all three columns — the one system in this Part that runs a path, maintains fluxes, and corrects an estimate at once. Each box carries the keyword of the measurement that could refute its reading, expanded case by case in the sections below. The placements are placements, not confirmations: none of the seven is offered as evidence for the framework, and the discipline of the Part is that every reading closes with its falsifier stated in the text.

39 Thermodiffusion: the textbook MinEP state

Observed. A fluid mixture held between plates at different temperatures develops a stationary concentration gradient — the Soret effect — carrying heat but no net matter flow.

The old reading. Prigogine's theorem, in its home regime: with the thermal force clamped and the diffusive force free, the stationary state minimizes entropy production, and the minimum is realized by the vanishing of the matter flux.

The gauge reading. Force-side reading of Theorem 34.2(ii), exactly as in the classical proof — the derivation adds only the provenance of the quadratic form (the ledger's Hessian) and the statement of scope: the reading holds while the coefficients are constant across the cell and all variables are even.

What would refute it. A measured stationary Soret state with a persistent nonzero matter flux at small driving, or a minimum displaced from the zero-flux state beyond the quadratic corrections the functional itself predicts.

40 Jaynes's two-temperature network: the counterexample as a theorem

Observed (thought experiment, realizable on a bench). A resistor network with elements held at different temperatures distributes currents by Kirchhoff's laws — minimizing total heat production, not total entropy production; the two prescriptions differ whenever the temperatures do.

The old reading. A refutation of MinEP's generality, and it is.

The gauge reading. Theorem 34.4 with the inhomogeneity mechanism: a temperature field varying across the system makes the ledger's weight $1/T$ nonuniform, the level sets and the circulation cease to be orthogonal, and the work terms survive — the extremum principle is derived to fail here. The physical selection (Kirchhoff) remains the stationary point of the full transport cost (32.1), which does not degenerate to an entropy functional in this case.

What would refute it. A driven network in which the physical current distribution deviated from the stationary point of the full quadratic cost — not merely from the entropy extremum — at leading order.

41 Bénard convection: the boundary of both principles

Observed. A fluid layer heated from below conducts at small temperature difference and, above a critical threshold, spontaneously organizes into convection rolls whose transport exceeds conduction.

The old readings. For MinEP, the historical falsifier: the convective branch produces more entropy than the conductive one, and it is the convective branch that is stable beyond threshold. For the maximum tradition, the historical motivation, in the selection-principle form.

The gauge reading. The onset is a regime boundary of the current gauge, not a paradox: below threshold the curl sector does no ledger work and the force-side minimum governs; the instability is the point at which a circulating sector begins to do work ($W_\mathrm{c} \neq 0$), the quadratic reduction fails

(Theorem 34.4), and selection passes to the full functional — in runtime language, a commitment event of the cascade, graded conditional per Table 5. The framework does not predict the convective pattern from an entropy extremum, and says why no such prediction can be exact.

What would refute it. A driven-fluid steady state beyond threshold demonstrably selected by an entropy-production extremum alone, robustly across geometries — the derivation forbids it.

42 Planetary heat transport: Paltridge's climates on three worlds

Observed. With no fluid dynamics — no rotation rate, no tuned diffusivity — an energy-balance model closed by maximizing the entropy production of the meridional transport reproduces the equator-to-pole temperature contrasts of Earth, Mars, and Titan.

The old reading. The flagship success of MEPP, and its standing embarrassments: nothing in the principle says which entropy production to maximize (the choice that works ignores the radiatively dominant part of the budget), or why an unconstrained extremum should know about atmospheres at all.

The gauge reading. A flux-side reading of Theorem 34.2(i) with the radiative forces clamped: the turbulent transport sector, fast and strongly mixing relative to the radiative relaxation, is the sector for which the constitutive selection applies, and the ledger (30.3) — fixed upstream, not chosen per application — identifies the *material* entropy production as the extremized quantity, because the radiative sector enters the boundary functional as clamped ports, not as varied currents. The constraint choice that Paltridge made empirically is, in this reading, the ports declaration of the packet; what was arbitrary becomes stated — and therefore falsifiable.

What would refute it. An atmosphere (observed or simulated at adequate resolution) whose material transport, with radiative ports genuinely clamped and scale separation genuinely present, sits systematically off the flux-side stationary point — or a demonstrated dependence of the selected state on resolution that survives the stated scale-separation condition. Both concerns are live in the climate literature, and the reading stands exposed to them.

43 ATP synthesis: all three gauges in one molecule

Observed. The rotary enzyme that synthesizes ATP pays for each bond with an integer number of proton crossings set by its rotor size — different integers on different organisms, tuned to the available electrochemical gradient — while the free energy of the bond itself is universal: the measured 30.5 ± 0.5 kJ/mol standard hydrolysis free energy against the program's derived $2\mathsf{Q}_\mathsf{B} = 30.0$ kJ/mol, an agreement at the 1.6% level (Arisaka, 2026, A7-Qi4).

The old readings. Bioenergetics treats the number as measured; the free energy literature treats the cell as an inference engine; the thermodynamic literature treats the chemiosmotic machine as a near-equilibrium transducer.

The gauge reading. The one system in this Part that exercises all three gauges at once, which is why the program's home papers use it. The proton path through the rotor is trajectory-gauge (a least-time path through a curved potential); the steady turnover is current-gauge (flux–force operation of a linear transducer near its stall force, with the duality of Theorem 34.2 fixing the operating point); and the regulatory network around it is inferential-gauge (the cell's estimate of its energy state updated by natural-gradient descent). The three descriptions do not conflict because they are three restrictions of (32.1) — which is the concrete answer to the entrance question of why a cell obeying three principles is not pulled in three directions.

What would refute it. An elementary catalytic event whose four-way bookkeeping violates the single conversion factor outside the stated tolerance — the falsifier of the underlying pillar, inherited here.

44 Chemotaxis: the inferential gauge with a checkable rank condition

Observed. A swimming bacterium climbs a chemical gradient it cannot spatially resolve, by comparing present against remembered concentration and modulating its tumbling rate.

The old reading. The free energy literature's minimal agent: an internal state performing inference on an external one across a blanket.

The gauge reading. Equation (35.8) with a one-dimensional belief coordinate (the methylation state as the estimate of gradient sign), and — the difference that matters — the internality construction's rank condition doing real work: the bacterium's internal model exists *because* self-motion raises the observability rank; a tethered cell, sampling but not moving, falls below the rank condition and the framework says it can hold no unique gradient estimate. The blanket is not assumed; the ports (receptors, motor) are declared; the existence of the inference reading is a checkable condition on the sensorimotor loop.

What would refute it. A demonstration that gradient-climbing performance survives the closure of the action channel (rank collapse) — immotile cells inferring gradients they cannot move through — or an identified chemotactic architecture whose update law deviates from natural-gradient structure where the metric is measurable.

45 Neural prediction: the gauge the principle was built for

Observed. Cortical responses attenuate to predicted input and amplify prediction error; hierarchical sensory processing carries top-down predictions and bottom-up residuals; behavioral and neural data are fit, with quantitative success, by precision-weighted prediction-error dynamics.

The old reading. The FEP's home territory — and, per Part IV, the place where the working mathematics was always solid regardless of the foundational claims.

The gauge reading. The inferential gauge in its hierarchical Gaussian restriction, where (35.8) reduces to exactly the predictive-coding dynamics of the process theories — so everything that literature has empirically established transfers unchanged. What the derivation adds is the boundary of the reading: the same solenoidal condition as everywhere else (strongly circulating neural dynamics — oscillatory coupling across the declared ports — takes the system outside the gradient-flow regime), and a bridge that is now checkable in principle: the metric in (35.8) is fixed by the ledger, so precision weighting is not a free parameter family but a measurable geometry.

What would refute it. Identified neural inference dynamics whose metric structure, where measurable, is not Fisher — or inference maintained in regimes where the circulating sector demonstrably does ledger work, without the degradation the reduction predicts.

46 What the seven cases share

In every case the derived reading does three things the original principle could not: it says *which* functional (fixed upstream, never chosen per application), it says *where the reading stops* (a stated sector or rank condition, not a surprise), and it names a measurement that could kill it. And in every case the failure literature of Parts II–IV reappears not as embarrassment but as confirmation of a boundary the derivation drew in advance. That inversion — counterexamples as theorems — is the practical content of deriving a principle instead of adopting it.

47 Closing remarks: what the seven cases teach

Seven cases, one lesson, and it is worth saying it in a biologist's terms. In every case the derived reading did something the original principle could not do: it named its functional (no choosing which entropy to extremize after seeing the data), it stated its own limits (sector and rank conditions, announced in advance), and it exposed a throat — a specific measurement that would kill it. That third property is the one to prize. Working scientists are rightly wary of frameworks that explain everything; the seven readings here are valuable precisely because each one can fail, and says how. Notice also what the cases collectively show about living systems in particular: the enzyme, the bacterium and the brain were not three different kinds of exhibit but the same exhibit at three scales — an open system running one budget while a path is taken, flows are maintained, and an estimate is corrected. The claim that a cell obeys three principles at once stopped being a philosophical curiosity in this Part and became an experimental posture: pick the gauge your assay can see, check its conditions, and the other two readings are not rivals but the same measurement awaiting a different instrument. If one sentence survives from this Part into a reader's own work, let it be the question each case ended on: *what would refute this reading?* — asked before the reading is believed.

Part IX
Significance and Uniqueness of the Approach

What changes when a principle becomes a theorem with a boundary — with the four frameworks compared column by column

This Part steps back and asks what has actually been gained — in plain terms first, because the gain is easiest to see from outside the mathematics.

For a century, the sciences of open systems have run on borrowed confidence. A rule would work — in a climate model, in a plasticity calculation, in a neural simulation — and its users would extend it credit, because working is the best argument there is. But credit is not a foundation. When the rule failed somewhere new, there was no way to know whether the failure was the rule's edge or the user's error; when two rules disagreed, there was no court to hear the case; and when a skeptic asked *why* the rule worked, the honest answer was that nobody knew. Parts II–IV documented all three costs in detail. What Part VII changed is the credit arrangement: the three rules are now consequences of one functional that is fixed by something other than the rules themselves — and a consequence, unlike a postulate, comes with a warranty card that says exactly what is covered and what voids it.

48 Five things the derivation changes

1. The functional is fixed upstream. Every prior attempt to found these principles chose its functional at the point of use: which entropy production, which constraints, which partition. The projected functional of Part VII is fixed by the bulk action and the boundary channel — objects constructed in companion papers for reasons that have nothing to do with thermodynamic extremum principles — and arrives with no dial left to turn. This is the difference between a rule and a consequence, and everything else follows from it.

2. Domains of validity become theorems. The linear-regime confinement of both entropy-production principles, the parity condition, the inhomogeneity failure, the coupling and rank conditions of the inference reading: in the source literatures these were discovered by collision, over decades. Here each is a stated theorem of the reduction, and the classical counterexamples land where the theorems say. A user of any of the three principles can now ask, *before* applying it, whether the sector conditions hold — a question the principles themselves could never answer.

3. The quarrels dissolve without winners. Minimum versus maximum entropy production was never a disagreement about nature; it was a coordinate dispute at one stationary point, and the duality theorem exhibits the two as Legendre-dual conventions. Likewise the inference reading does not compete with the thermodynamic ones: the three gauges are restrictions of one functional, so a cell that follows a path, maintains fluxes, and updates an estimate receives one instruction, not three.

4. The modeler-chosen structures are replaced by declared ones. The two structural criticisms that cut deepest in the review — that MEPP's constraints and the FEP's blanket are put in by hand, so the conclusions follow the modeler's choices — are met not by better choices but by removing the choice. The ledger fixes which entropy production; the packet's ports fix the system boundary as constitutive data carrying kill criteria; the rank condition decides, case by case, whether an internal model exists at all. What was arbitrary is now stated, and what is stated can be wrong.

5. Falsifiability is restored. An unfalsifiable principle generates unfalsifiable applications. A derived functional reverses this: every reading in Part VIII closes with the measurement that would refute it, and the underlying chain carries its own named falsifiers upstream — including one already standing at the 1.6% level in every living cell.

49 The four frameworks, column by column

Table 6 sets the three reviewed frameworks beside the present one across twelve dimensions. Reading advice: the table is not a scoreboard on which the fourth column wins every row — rows 9 and 12 record, deliberately, where the fourth column's own liabilities sit. The instructive rows are 3 through 6, where the three reviewed frameworks' answers are "chosen by the modeler" or "discovered by collision," and the change offered here is structural rather than a matter of degree.

Row 12 deserves its own paragraph, because it is the honest price tag. The three reviewed frameworks are cheap to adopt — MinEP is textbook material, MEPP is a closure anyone can bolt onto a box model, the FEP ships as software. The present framework is expensive: its functional is only as trustworthy as a six-dimensional construction, an anomaly proof, and a constitutional charter that most readers of this supplement have never examined. That expense is not hidden here; it is the trade. What the reader buys with it is rows 3 through 8 — and the supplement's position is that after a century of unpriced principles, the field's problem has been too little upstream structure, not too much.

50 What is unique to this route

Three features distinguish the derivation from every prior attempt reviewed in Parts II–IV.

First, *direction*. The path-information route and the blanket route both ran from below: assume a microscopic or stochastic structure, and climb toward the principle. Both failed at the climb — the Gaussian step, the silent coupling assumptions. The present route runs from above: the functional descends from an action fixed by anomaly cancellation and constitutional constraints, and the principles are met on the way down, each in its regime. Meeting the fluctuation-theoretic structure (Onsager–Machlup, macroscopic fluctuation theory) at the same functional from the opposite direction is the consistency check the from-below routes never had.

Table 6: The three reviewed frameworks and the present one, compared across twelve dimensions. Each row is one question a working scientist might put to a framework before adopting it. Sources: the MinEP column condenses Part II, the MEPP column Part III, the FEP column Part IV, and the GDOS column Parts VI–VII, including its liabilities (rows 9 and 12). The table is the supplement's summary judgment and is meant to be quotable on its own; a reader who checks any cell against the corresponding Part should find nothing stronger in the cell than in the Part.

	MinEP	**MEPP**	**FEP**	**GDOS/GRIP/Qi-4**
1. Central object	entropy production $P(X)$	entropy production $\dot{\Sigma}(J)$	variational free energy F	one projected functional $\mathcal{A}_{\text{eff}}$
2. Status of the object	phenomenological	phenomenological / inferential	postulated ("principle")	derived from an upstream action
3. What selects the functional	convention (linear regime)	modeler (which EP, which constraints)	modeler (generative model, blanket)	fixed upstream; no per-application choice
4. System boundary	implicit (clamped walls)	implicit (box model chosen by hand)	Markov blanket, contested	declared ports of the packet, with kill criteria
5. Domain of validity	discovered by collision; codified late	unknown in advance; case by case	extracted by critics; contested	stated as theorems (sector and rank conditions)
6. The counterexamples	external embarrassments, later cataloged	external (choice-of-EP, resolution dependence)	external (blanket inequivalence, non-genericity)	internal theorems: derived to fail where they fail
7. Relation to the other two	presented as rival to MEPP	presented as rival to MinEP	unrelated literature	all three as gauges of one functional; duality theorem
8. Best rigorous result	the 1945 theorem (linear regime)	Kohler–Ziman (linear regime)	the variational bound (mathematics, uncontested)	the three gauge theorems + duality + failure location
9. Weakest point	scope narrowness	no derivation; constraint dependence	foundational claim withdrawn to instrumentalism	conditional imports (M11, spectral-flow form); gauge exhaustiveness argued, not proved
10. Information's status	absent	absent	central but unpriced	priced: Q_{B} derived, anchored at 1.6% by ATP
11. Falsifiability	yes, within scope	weak (constraints adjustable)	disputed; defended as unfalsifiable-by-status	named falsifiers per reading and upstream
12. Cost of adoption	none (textbook)	low	moderate (modeling apparatus)	high: a six-dimensional framework whose own foundations must be tested

Second, *the failures are load-bearing.* No prior framework derived the counterexamples. Here Landauer's circuit, Jaynes's network, the convective instability, and the non-genericity finding of the linear-systems critique of the FEP are not accommodated but required — the sector and rank conditions that produce the principles also produce, and locate, their breakdowns. A derivation that could not fail where the principles fail would be wrong; this one fails there by theorem.

Third, *one obstruction, named.* The deepest result of the critical literatures — that selection beyond linear order depends on time-symmetric activity invisible to any entropy functional — appears here as a single geometric statement: the curl and gauge sectors of the boundary current do work. That one condition governs the current gauge and the inferential gauge alike, which is why three literatures kept meeting the same wall from three sides.

51 Significance beyond the three principles

Three wider consequences, stated with their grades.

For nonequilibrium thermodynamics. The duality theorem gives the field a clean grammar for a recurring confusion: extremum statements about *constitutive relations* (flux side) and extremum statements about *state selection* (force side) are different kinds of statement, conjugate at one point, and the field's historical quarrels largely dissolve under the distinction. This much agrees with the best modern critical literature and is graded derived; what the framework adds beyond it — the upstream fixing of the functional — is only as strong as the framework's own foundations.

For the biology of energy. If information carries a derived price, then bioenergetics and information processing are one subject with one currency, and questions that sound philosophical become budgetary: how much of a cell's ATP turnover services its memory and prediction, and does the ledger balance? The 1.6% ATP anchor makes this more than rhetoric, and the reading is exposed to a named laboratory test at the single-event level.

For the sciences of mind. The inference reading survives here in exactly the form its best critics said it should: as a conditional description, holding where a checkable rank condition and sector condition hold, silent elsewhere. That is a demotion from "all persisting matter infers" and a promotion from "useful modeling scheme": inference becomes a physical regime a system can be in or out of — and which regime a given preparation occupies is an experimental question.

52 Closing remarks: what has been bought, and for whom

For the life-science reader, the gain of this whole exercise can be put in one sentence: *three rules you may have met as slogans now have a common source, exact domains, and price tags.* If you have used minimum entropy production to reason about a gently driven pathway, this framework tells you when that reasoning was licensed. If you have met maximum entropy production as "life accelerates the degradation of gradients," it tells you what that reading is worth — a flux-side selection statement, valid under stated conditions, not a purpose written into nature. And if you have run free-energy models of perception, it tells you what the model's "free energy" has to do with the free energy in your biochemistry textbook: they are two readings of one ledger, one taken with beliefs held primary and one with currents, and the exchange between them is not a pun but a theorem.

None of this obliges a biologist to adopt a six-dimensional physics. The comparison table is explicit that adoption is expensive and that the framework's own foundations carry open conjectures. But the supplement's minimal offer stands independent of full adoption: the grammar — one functional, three gauges, sector conditions, declared ports, rank conditions — can be used as a checklist by anyone, today, to ask of any extremum-principle argument in the life sciences the two questions that were never askable before: *which reading of what functional is this, and are its conditions met?* A field that asks those two questions routinely will waste less of its next century than these three literatures wasted of their last.

Part X
Conclusions

The whole argument retold in plain language, and what it means

This closing Part is written for the reader the rest of the supplement has been hardest on: the life scientist who came for the ideas and has been wading through other people's mathematics to get them. Here is the whole argument again, in order, with no equations — and then what we think it means.

53 The story so far, in plain language

Begin with the question that started everything. *Which way does an open system go?* Not a system in a textbook box, sealed and settling to equilibrium — an open one: a cell fed by a gradient, a planet fed by a star, a brain fed by its senses. Equilibrium physics has one magnificent answer for sealed boxes — downhill in free energy — and no general answer for anything alive enough to be interesting.

Three sciences filled the gap with three rules. The first rule says: if the pushing is gentle, the system settles into the laziest steady state available — the one that wastes the least. That is Prigogine's minimum principle, and it is genuinely true, provably, in its gentle regime. A resting cell idling at basal metabolism is not far from this picture: hold the gradients mild and steady, and the state that persists is the thriftiest one on offer.

The second rule says nearly the opposite: push hard, give the system many ways to carry the load, and it will find the way that destroys the gradient *fastest*. That is maximum entropy production. Feed it nothing but sunlight and bookkeeping and it reproduces the temperature gap between equator and poles — on Earth, on Mars, on Titan. Nobody has ever derived it, its failures are as striking as its successes, and its own advocates concluded it cannot be proved from what we currently know.

The third rule comes from the study of brains: whatever persists must behave *as if* it were predicting its inputs and correcting its errors — minimizing, in the jargon, a free energy that measures surprise. As working mathematics this powers half of computational neuroscience. As a claimed law about all persisting matter it did not survive inspection: the assumptions that make a lump of matter into an inference engine turn out to be special, not generic, and the principle's defenders now describe it as a way of modeling systems rather than a fact about them.

So the field stood, roughly, at this scoreboard: one true theorem with a small territory; one powerful heuristic with no foundation; one powerful modeling language with a withdrawn foundation; and — the detail this supplement kept returning to — a living cell that is simultaneously a gently driven chemical network, a hard-driven engine, and a predictor, with no explanation anywhere of why it is not torn apart by three different instructions.

The constructive claim of this supplement is that the three rules are one thing seen three ways. The program's companion papers build, from considerations that have nothing to do with any of these literatures, a single master cost function for open systems — a landscape with a built-in accounting system whose columns are energy, charge, matter, and information. Ask that one object "which *path* does the system take?" and it answers with the least-time law of optics and mechanics. Ask it "which *flows* does the system carry?" and it answers with the flux–force law of thermodynamics — and the minimum and maximum principles fall out as the two sides of that single answer, one holding the pushes fixed, one holding the plumbing fixed, never rivals at all. Ask it "which *guess* does the system update?" and it answers with the error-correcting rule of the brain sciences, with the vague parts of that story replaced by checkable conditions. One landscape; three questions; three famous answers.

And — this is the part we would ask a skeptical reader to weigh most heavily — the derivation also produces the *failures*. The places where each rule breaks were not explained away; they were derived. The circulating currents that ruin the minimum principle in a driven circuit, the convection cells that embarrassed it historically, the special couplings that the brain-science principle silently needs — all of them sit exactly where the mathematics of the reduction says they must, in the two sectors of the flow that circulate and relabel rather than descend. A theory that predicts only successes explains nothing; the arithmetic of this one required its own limits.

54 What this means for a life scientist

Suppose you grant the argument — or grant it provisionally, as one should. What changes at the bench?

First, *a checklist replaces a mood.* Extremum arguments circulate in biology in a low-precision form: this pathway is efficient, that network maximizes throughput, this organism minimizes surprise. The gauge grammar converts each such claim into two answerable questions: which reading of which functional is being invoked, and are its conditions — gentle driving, no working circulation, an intact sensing-and-acting loop — actually met in the preparation at hand? A claim that cannot answer is not wrong; it is not yet a claim.

Second, *information gets a price.* The most consequential single number in this supplement for a biologist is the derived exchange rate between energy and information — about six times the thermal energy scale per bit, two quanta per ATP bond, checked to one and a half percent. If that number survives, then memory, sensing and prediction are not free riders on metabolism; they are line items, and questions like "what fraction of this neuron's budget services its predictions" stop being metaphors and become measurements someone can attempt.

Third, *the brain-science language and the biochemistry language reconcile.* A generation of students has been told, in one building, that free energy is what ATP hydrolysis releases, and in the next building, that free energy is what perception minimizes — with the resemblance waved off as coincidence. On the account given here it is not a coincidence: the two are readings of one ledger, taken with different quantities held primary, and the theorem connecting them is short. Fields that share a currency can trade; that is the practical meaning of unification.

Fourth, *a posture toward principles.* The deepest lesson of the three histories is not about entropy at all. It is that working rules promoted to laws without derivation cost their fields decades — in misapplication, in empty quarrels, in retreats dressed as reformulations. Biology, importing physics' principles secondhand, has often imported the promotion along with the rule. The protection is not mathematical sophistication; it is the habit of asking every principle for its passport: where are you valid, what would refute you, and what fixed you.

55 What we are *not* concluding

Symmetry requires this section, and one page of candor buys the rest. We are not concluding that maximum entropy production is true — the derivation says it is true only in a bounded regime and false outside it, and the boundedness is the result. We are not concluding that every organism computes probabilities — the inference reading holds where a checkable condition on the sensing-and-acting loop holds, and is silent elsewhere. We are not concluding that the framework behind the derivation is established — it stands on companion papers whose own foundations carry named open conjectures, and this supplement inherits every one of them, saying so at each point of use. And we are not concluding that the three gauges are the only readings the master functional admits; that count is argued, not proved. What we conclude is narrower and, we believe, more durable: *there exists at least one explicit, checkable construction under which the three great extremal principles of open systems, and their known failures, follow from a single object* — and its existence changes what the three literatures can reasonably demand of any future foundation, including a better one than this.

56 A closing word

Step far enough back and the century this supplement has reviewed looks like a single long attempt to answer a child's question: *why does anything bother?* Why does the storm organize, the flame hold its shape, the cell repair its membrane, the mind press its guesses against the world? Equilibrium physics, for all its power, answers only why things stop. The three principles reviewed here were three attempts to say why things *go* — and each, in its own dialect, gave a version of the same strange reply: that going has a direction, that the direction is set by an accounting, and that the books are kept in quantities no one can touch — entropy, likelihood, cost.

What the derivation adds is the suggestion that the three dialects were never separate languages, and that their common tongue prices a currency the sciences have always handled but never valued: information. If that is right, then the divide our institutions draw — matter in one faculty, meaning in another; the energetic sciences here, the informational ones there — is not a divide in nature. A cell metabolizing and a cell remembering are running the same ledger. A brain predicting and a protein folding are descending the same kind of landscape. The world does not contain physical things and informational things; it contains open systems, keeping one set of books in several currencies, and every science we have built is a column of that ledger read with something different held fixed.

There is a long tradition, running from Heraclitus's fire to Schrödinger's *What is Life?*, of suspecting that persistence — not substance — is the deep fact about the world: that a thing is not what it is made of but what it manages to keep doing. The literatures reviewed here are that suspicion, worked out in modern tools, three separate times. Perhaps the fitting last word for a supplement about their unification is that the suspicion itself may be the theorem: that to exist, for an open system, simply *is* to run the ledger — to spend gradient, bank structure, and settle, for a while, into the deepest place the landscape allows. Nothing in the universe stays. Some of it, for a time, keeps its books well enough to seem to. We are among its better accountants, and it should not surprise us — being ourselves entries in the ledger — that when we finally audited the three great rules of staying, we found a single hand had written them.

Part XI
References

Cited works, each with a relevance note linking it to the present review

Unlike a conventional reference list, every entry below carries a short *Relevance* note saying what the cited work contributes to the present review. Section A lists the papers of the GG-Theory program on which the derivation Part stands; Section B lists all other works, alphabetically by first author. The bibliographic details of Section B follow the verification discipline stated in Part I: where a page range or edition could not be checked against a primary source, it is omitted rather than asserted.

A. Papers of the GG-Theory program

Arisaka, K. (2026). *The GG-Constitution: Four Axioms and Seven Postulates of the GG-Theory Program, From Which GG-6, BI-6, OGN-5, and GDOS Descend as Theorems*. Preprint, available at https://preprints.arisaka-gg.org/a1-constitution/.
doi:10.5281/zenodo.22063867.
Relevance. The parent constitutional paper of the GG-Theory program. Establishes the four axioms AX-1 LC, AX-2 GGEP, AX-3 OGGEP, AX-4 MSMF, together with the seven postulates POS-1–POS-7 (GI, UEP, AF, CP, RG, OGN, MF) and the central Arisaka equation $J_E/\mathsf{Q_B} = J_e = J_B = J_{Qi}$ that ties the four boundary currents through one energy–information conversion factor. Cited here as the upstream charter: the boundary channel reviewed in Part VI descends from the observer axiom AX-3 OGGEP, and the conversion constant $\mathsf{Q_B}$ that prices the ledger's information column enters this supplement through the Arisaka equation established here.

Arisaka, K. (2026). *GG-Theory with GG-6: Geometry–Gauge Locking at the 6D Bulk (GG-6) at the Heart of Grand Geometric Theory (GG-Theory)*. Preprint, available at https://preprints.arisaka-gg.org/a2-gg6/.
doi:10.5281/zenodo.22064298.
Relevance. The bulk-side foundational paper of the GG6 Trio. Establishes the 6D bulk metric $\mathrm{d}s_6^2 = \mathrm{e}^{-2r}\eta_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu + \mathrm{d}r^2 + \mathrm{d}\theta^2$, the canonical four-component gauge group $G_{\mathrm{tot}} = \mathrm{Spin}(1,5) \times \mathrm{SU}(5) \times \mathrm{U}(1)_B \times \mathrm{U}(1)_{Qi}$, the magic-dimension ladder $\mathrm{Spin}(1,5) \cong \mathrm{SL}(2,\mathbb{H})$, and the OGN-5 Diophantine cascade $(M,N,L;J,K) = (1,3,5;7,35)$ with the dual-lock identity $K = LJ = M^2+N^2+L^2 = 35$.[1] These are the ultraviolet inputs from which the boundary law reviewed

[1]Notation, kept uniform across the cohort: N is the OGN-5 ledger's second entry, and every ledger identity is written with it. $N_c = 3$ is the color triplicity, the three colors of $\mathrm{SU}(3)_C$ — what that entry counts, named where the color reading is the point. $N_g = 3$ is the generation (family) count; it is *not* a ledger entry and is not fixed by the ledger (GG-P2-N3). $N_f = 2N_g = 6$ is the quark-flavor count of a beta function at high scale. Three of these are three and one is six, and they are three different quantities.

in Part VI is projected, and the integer ledger, through $K = 35$, fixes the conversion constant $\mathsf{Q_B} = (2/5)\, E_{\mathrm{Ryd}}/K$ that this supplement anchors against ATP synthesis.

Arisaka, K. (2026). *Anomaly Freedom of BI-6: The Modified Bianchi Identity (BI-6) for the Six-Force, Six-Dimensional Bulk (GG-6)*. Preprint, available at `https://preprints.arisaka-gg.org/a3-bi6/`.
doi:10.5281/zenodo.22064300.
Relevance. The anomaly-inflow companion of the GG6 Trio. Establishes BI-6 as the four-component Green–Schwarz-modified Bianchi identity $\mathrm{d}H_3 = X_4$ on the 6D bulk, with the four-component coefficient vector $(a, b, c, d) = (1, -1, -2/5, -2/5)$ and the trace-ratio lock $|a/c| = |a/d| = 5/2$ pinned by the OGN-5 integer ledger. Cited here, beside the bulk construction, as the consistency half of the six-dimensional story this supplement imports but does not defend. (This entry absorbs the legacy preprint code GD-2026-5, under which the BI-6 package circulated before the cohort renaming; older internal references to "Paper V" resolve here.)

Arisaka, K. (2026). *GDOS: Geometric Dynamics of Open Systems — 6D GG-6 Bulk Projected to 4D Boundary for Quantum Measurement, General Relativity, and QFT*. Preprint, available at `https://preprints.arisaka-gg.org/a5-gdos/`.
doi:10.5281/zenodo.22064304.
Relevance. The stage of the whole derivation: the boundary law of the partial view. From the constitutional observer axiom AX-3 OGGEP, derives the CPTP bulk-to-boundary projector $\mathcal{D}_{\mathsf{OGGEP}} = \mathcal{R} \circ \mathrm{Tr}_{\mathrm{env}}$, the Nakajima–Zwanzig kernel, and — in the mesoscopic limit — the continuity law with its four-sector current decomposition, which Part VII takes as its starting equation. The curl and gauge sectors of that decomposition locate every classical counterexample reviewed in Parts II–IV, and the sector conditions of the three gauge theorems are read directly off this paper's boundary equation.

Arisaka, K. (2026). *The GRIP on GDOS — GRIP: the Geometric Engine of Symmetry Breaking, from the Big Bang to Mind*. Preprint, available at `https://preprints.arisaka-gg.org/a6-grip/`.
doi:10.5281/zenodo.22064306.
Relevance. The engine of this supplement. Source of the runtime cascade GRIP (DGI → GSSB → GA), the scalar ledger $F_{\mathsf{GFF}} = F_{\mathsf{GEP}} + F_{\mathsf{CEP}}$ with its four-sector current decomposition, the SIP packet, and the balance identity from which the three gauge theorems of Part VII are derived. The cascade-closure residue that grades this supplement's conditional rows — the monodromy statement M11, in spectral-flow form — is this paper's; the budget bound M10 is resolved externally and the scope refinements are closed at class level by Arisaka, 2026, P2-N3.

Arisaka, K. (2026). *The Qi-4 Pillars — Qi-QUA, Qi-EP, Qi-CON, Qi-HAL: The Quantum-Information Backbone of GDOS*. Preprint, available at `https://preprints.arisaka-gg.org/a7-qi4/`.
doi:10.5281/zenodo.22064308.
Relevance. The quantum-information half of the runtime this supplement imports: the four Qi-4 pillars (Qi-QUA, Qi-EP, Qi-CON, Qi-HAL) descending from OGGEP on the memory-bearing

$U(1)_{Qi}$ sector. Qi-QUA prices the ledger's information column at the derived rate $\mathsf{Q_B}$ per bit — the exchange rate this supplement anchors against ATP synthesis — and Qi-EP is the home paper of the three-gauge theorem deployed at full length in Part VII. The internal-model rank condition graded conditional in the grade table is this paper's internality construction.

Arisaka, K. (2026). *Why Three Generations: A Three-Leg Derivation of $N_g = 3$ from BI-6, MSMF, and GDOS Morse Stability.* Preprint, available at https://preprints.arisaka-gg.org/p2-n3/. doi:10.5281/zenodo.22064323.
Relevance. Cited here for a single result: its capacity theorem, which establishes the budget bound that the runtime's cascade closure required as its conjecture M10 — externally, without using the cascade — so that the conditioning inherited by this supplement reduces to the monodromy statement M11 in its spectral-flow form; the same paper's rigidity theorems close the warped and relative refinements at class level. First cited in Part VI.

B. All other works

Županović, P., Juretić, D., and Botrić, S. (2004). *Kirchhoff's loop law and the maximum entropy production principle.* Physical Review E 70, 056108.
Relevance. The network MaxEP claim shown in the relevant section of that paper to be Onsager least dissipation in network clothing.

Aguilera, M., Millidge, B., Tschantz, A., and Buckley, C. L. (2022). *How particular is the physics of the free energy principle?* Physics of Life Reviews 40, 24–50.
Relevance. Exact analysis of linear Ornstein–Uhlenbeck systems showing the FEP's conditions hold only on a narrow parameter subspace requiring no perception–action asymmetry, and that flow-of-average and average-of-flow are conflated.

Andrews, M. (2021). *The math is not the territory: navigating the free energy principle.* Biology and Philosophy 36(3), 30.
Relevance. The clearest statement of the instrumentalist defense: the FEP as a modeling tool judged by fruitfulness, not a falsifiable claim about systems.

Ao, P. (2004). *Potential in stochastic differential equations: novel construction.* Journal of Physics A 37, L25–L30.
Relevance. Constructs the antisymmetric-matrix decomposition of stochastic dynamics into potential-gradient and circulatory parts; establishes that this stratum of the FEP is standard nonequilibrium theory.

Ashby, W. R. (1952). *Design for a Brain.* Chapman & Hall, London.
Relevance. Frames adaptive behavior as homeostasis of essential variables under ultrastable dynamics; the cybernetic root of the FEP's identification of persistence with bounded surprisal.

Bertini, L., De Sole, A., Gabrielli, D., Jona-Lasinio, G., and Landim, C. (2004). *Minimum Dissipation Principle in Stationary Non-Equilibrium States.* Journal of Statistical Physics 116, 831–841.

Relevance. Generalizes Onsager–Machlup to nonequilibrium stationary states within macroscopic fluctuation theory, the framework in which MinEP survives only as a near-equilibrium reduction.

Bertini, L., De Sole, A., Gabrielli, D., Jona-Lasinio, G., and Landim, C. (2015). *Macroscopic fluctuation theory.* Reviews of Modern Physics 87.

Relevance. The comprehensive account of the quasi-potential and its Hamilton–Jacobi equation as the exact variational characterization of nonequilibrium steady states, of which MinEP is the near-equilibrium quadratic shadow.

Biehl, M., Pollock, F. A., and Kanai, R. (2021). *A technical critique of some parts of the free energy principle.* Entropy 23(3), 293.

Relevance. Proves the two blanket definitions inequivalent, the rewriting of expected flows invalid without unstated assumptions, and the Free Energy Lemma false at face value, with counterexamples.

Bruers, S. (2007). *A discussion on maximum entropy production and information theory.* Journal of Physics A: Mathematical and Theoretical 40, 7441.

Relevance. Reconstructs both Dewar derivations, shows the 2003 machinery yields MinEP in an explicit model, and supplies the taxonomy of inequivalent MaxEP statements.

Bruers, S., Maes, C., and Netočný, K. (2007). *On the Validity of Entropy Production Principles for Linear Electrical Circuits.* Journal of Statistical Physics, doi:10.1007/s10955-007-9412-z.

Relevance. The parity classification — MinEP for even variables, a maximum-type dual for odd ones — absorbing Landauer's circuit counterexample and situating Jaynes's example beyond the resolution of either principle.

Bruineberg, J., Dolega, K., Dewhurst, J., and Baltieri, M. (2022). *The Emperor's New Markov Blankets.* Behavioral and Brain Sciences 45, e183.

Relevance. Distinguishes Pearl blankets (epistemic instruments within a model) from Friston blankets (reified physical boundaries), showing the mathematics does not underwrite the metaphysical reading.

Colombo, M. and Wright, C. (2021). *First principles in the life sciences: the free-energy principle, organicism, and mechanism.* Synthese 198(S14), 3463–3488.

Relevance. Reconstructs the FEP's argument chain, exposes its stipulative cross-disciplinary identifications, and makes precise the point that a principle unfalsifiable by design cannot serve as an empirical law.

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

Relevance. The good-regulator theorem; the FEP's claim that organisms are models of their environments is this result restated probabilistically.

Da Costa, L., Friston, K., Heins, C., and Pavliotis, G. A. (2021). *Bayesian mechanics for stationary processes.* Proceedings of the Royal Society A 477(2256), 20210518.
Relevance. The first mathematically careful Bayesian mechanics: for linear Gaussian processes with an imposed blanket, the synchronization map is exact and the expected internal flow is a genuine free-energy gradient flow.

Dayan, P., Hinton, G. E., Neal, R. M., and Zemel, R. S. (1995). *The Helmholtz machine.* Neural Computation 7(5), 889–904.
Relevance. The entry of variational free energy into neural modeling: a recognition density trained to bound the log evidence of a generative model, mathematically identical to Friston's F.

de Groot, S. R. and Mazur, P. (1962). *Non-Equilibrium Thermodynamics.* North-Holland, Amsterdam.
Relevance. The standard treatise of linear irreversible thermodynamics; its Chapter V states the MinEP theorem with the constancy and boundary-condition hypotheses explicit.

Dewar, R. C. (2003). *Information theory explanation of the fluctuation theorem, maximum entropy production and self-organized criticality in non-equilibrium stationary states.* Journal of Physics A: Mathematical and General 36, 631–641.
Relevance. First path-information derivation attempt; source of the construction and of the contested step (11.25).

Dewar, R. C. (2005). *Maximum entropy production and the fluctuation theorem.* Journal of Physics A: Mathematical and General 38, L371–L381.
Relevance. Second derivation attempt, claiming Ziegler's orthogonality far from equilibrium via a Gaussian step later shown to confine the result to linear response.

Dewar, R. C. (2009). *Maximum entropy production as an inference algorithm that translates physical assumptions into macroscopic predictions: don't shoot the messenger.* Entropy 11, 931–944.
Relevance. The author's concession that the Gaussian approximation is valid only close to equilibrium, and the retreat to MEP-as-inference.

Doyle, P. G., and Snell, J. L. (1984). *Random Walks and Electric Networks.* Mathematical Association of America, Washington.
Relevance. Standard modern source for Thomson's and Dirichlet's network principles, the dual pair proved in the relevant section of that paper.

Essex, C. (1984). *Radiation and the irreversible thermodynamics of climate.* Journal of the Atmospheric Sciences 41, 1985–1991.
Relevance. The structural objection that the dominant radiative entropy production is not extremized by the climate MEPP; the which-entropy-production problem.

Friston, K. (2005). *A theory of cortical responses.* Philosophical Transactions of the Royal Society B 360(1456), 815–836.
Relevance. First full statement of the brain-level program: hierarchical cortical processing as empirical-Bayes inference minimizing free energy.

Friston, K. (2009). *The free-energy principle: a rough guide to the brain?* Trends in Cognitive Sciences 13(7), 293–301.

Relevance. The widely read tutorial statement linking the variational objective to attention, action, and hierarchical message passing.

Friston, K. (2010). *The free-energy principle: a unified brain theory?* Nature Reviews Neuroscience 11(2), 127–138.

Relevance. The most-cited formulation of the FEP, claiming to subsume predictive coding, efficient coding, and optimal control under one objective.

Friston, K. (2013). *Life as we know it.* Journal of the Royal Society Interface 10(86), 20130475.

Relevance. The pivot to physics: random dynamical systems with Markov blankets at nonequilibrium steady state claimed to appear to perform Bayesian inference; origin of the contested derivations.

Friston, K. (2019). *A free energy principle for a particular physics.* arXiv:1906.10184; monograph, 148 pp.

Relevance. The full physics-phase construction — particular partition, synchronization map, free-energy gradient flow — and the specific target of the technical critiques.

Friston, K., Daunizeau, J., Kilner, J., and Kiebel, S. J. (2010). *Action and behavior: a free-energy formulation.* Biological Cybernetics 102(3), 227–260.

Relevance. The founding continuous-state active-inference paper: action minimizes the same free energy that perception minimizes, by changing the data rather than the beliefs.

Friston, K., Da Costa, L., and Parr, T. (2021). *Some interesting observations on the free energy principle.* Entropy 23(8), 1076.

Relevance. The authorized reply: accepts that the critique identifies implicit assumptions and heuristic proofs, and supplies a sparse-coupling ansatz restricting the principle's antecedent conditions.

Friston, K., Kilner, J., and Harrison, L. (2006). *A free energy principle for the brain.* Journal of Physiology—Paris 100(1–3), 70–87.

Relevance. First use of the phrase as a principle; introduces action as the second route to minimizing free energy, the germ of active inference.

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., and Pezzulo, G. (2017). *Active inference: a process theory.* Neural Computation 29(1), 1–49.

Relevance. The discrete-state (POMDP) formulation with expected free energy over policies; the basis of essentially all empirical active-inference work.

Friston, K., Da Costa, L., Sajid, N., Heins, C., Ueltzhöffer, K., Pavliotis, G. A., and Parr, T. (2023). *The free energy principle made simpler but not too simple.* Physics Reports 1024, 1–29.

Relevance. The path-integral reformulation that retreats from the nonequilibrium steady-state density factorization to a paths-of-least-action statement of the principle.

Friston, K., Da Costa, L., Sakthivadivel, D. A. R., Heins, C., Pavliotis, G. A., Ramstead, M., and Parr, T. (2023). *Path integrals, particular kinds, and strange things.* Physics of Life Reviews 47, 35–62.

Relevance. Extends the path-integral formulation to a taxonomy of particles; the current form in which the principle holds by construction for systems admitting a particular partition.

Glansdorff, P. and Prigogine, I. (1954). *Sur les propriétés différentielles de la production d'entropie.* Physica 20, 773–780.

Relevance. Introduces the decomposition $dP/dt = d_X P/dt + d_J P/dt$, the first step of the attempted extension beyond the linear regime.

Glansdorff, P. and Prigogine, I. (1964). *On a general evolution criterion in macroscopic physics.* Physica 30, 351–374.

Relevance. Proves $d_X P/dt \leq 0$ without linearity or reciprocity and establishes that $d_X P$ is not an exact differential — the precise sense in which MinEP admits no far-from-equilibrium generalization.

Glansdorff, P. and Prigogine, I. (1971). *Thermodynamic Theory of Structure, Stability and Fluctuations.* Wiley-Interscience, London and New York.

Relevance. The mature statement of the program: MinEP confined to the linear regime, the local-potential methods, and the stability theory from which dissipative structures emerged.

Goody, R. (2007). *Maximum entropy production in climate theory.* Journal of the Atmospheric Sciences 64, 2735.

Relevance. Climate-side critique: the missing dynamical constraints, the steady-versus-chaotic distinction, and the planetary data assessment.

Graham, R. (1977). *Covariant formulation of non-equilibrium statistical thermodynamics.* Zeitschrift für Physik B 26, 397–405.

Relevance. A classical source for the gradient–solenoidal decomposition of stationary stochastic dynamics that the FEP's physics phase takes as its starting point.

Grinstein, G., and Linsker, R. (2007). *Comments on a derivation and application of the "maximum entropy production" principle.* Journal of Physics A: Mathematical and Theoretical 40, 9717–9720.

Relevance. Identifies the error in the 2005 derivation and the unjustified self-organized-criticality claim of 2003; the pivotal critique.

Helmholtz, H. von (1868). *Zur Theorie der stationären Ströme in reibenden Flüssigkeiten.* Verhandlungen des naturhistorisch-medicinischen Vereins zu Heidelberg 5.

Relevance. The minimum dissipation theorem of Stokes flow, establishing the pattern that the minimum property belongs to the linear regime alone.

von Helmholtz, H. (1856–1867). *Handbuch der physiologischen Optik.* Voss, Leipzig; three volumes, completed 1867.

Relevance. Introduces unconscious inference, the doctrine that percepts are conclusions drawn from sensory premises and prior experience; the conceptual ancestor claimed by predictive coding and the free energy principle alike.

Jaynes, E. T. (1980). *The Minimum Entropy Production Principle.* Annual Review of Physical Chemistry 31, 579–601.
Relevance. The definitive scope critique: distinguishes Onsager's and Prigogine's principles and shows by the hot-and-cold resistor network that the physical steady state minimizes heat production rather than entropy production.

Kirchhoff, G. (1848). *Ueber die Anwendbarkeit der Formeln für die Intensitäten der galvanischen Ströme in einem Systeme linearer Leiter auf Systeme, die zum Theil aus nicht linearen Leitern bestehen.* Annalen der Physik und Chemie 75, 189–205.
Relevance. The earliest minimum-dissipation theorem for a driven steady state — currents in a linear resistor network minimize the Joule heat production — and the benchmark against which Jaynes's counterexample is measured.

Kleidon, A., and Lorenz, R. D. (eds.) (2005). *Non-equilibrium Thermodynamics and the Production of Entropy: Life, Earth, and Beyond.* Springer, Heidelberg.
Relevance. The field-defining edited volume of the 2003–2006 wave, collecting the climate, planetary, and theoretical strands.

Klein, M. J. and Meijer, P. H. E. (1954). *Principle of Minimum Entropy Production.* Physical Review 96, 250–255.
Relevance. First statistical-mechanical verification of Prigogine's theorem, whose first-order-only agreement is the earliest quantitative marker of MinEP's status as a near-equilibrium approximation.

Landauer, R. (1975). *Inadequacy of entropy and entropy derivatives in characterizing the steady state.* Physical Review A 12, 636–638.
Relevance. The blowtorch theorem: steady states of inhomogeneously heated bistable systems depend on kinetic details invisible to the entire entropy family of functionals.

Landauer, R. (1975). *Stability and entropy production in electrical circuits.* Journal of Statistical Physics 13, 1–16.
Relevance. Exhibits the failure of MinEP arbitrarily close to equilibrium in circuits with inductors — later diagnosed as a violation of the tacit time-reversal-parity hypothesis.

Lorenz, R. D., Lunine, J. I., Withers, P. G., and McKay, C. P. (2001). *Titan, Mars and Earth: entropy production by latitudinal heat transport.* Geophysical Research Letters 28, 415.
Relevance. The comparative-planetology success of the two-box MEP model worked out in the relevant section of that paper.

Maes, C. and Netočný, K. (2007). *Minimum entropy production principle from a dynamical fluctuation law.* Journal of Mathematical Physics 48, 053306.

Relevance. The modern derivation and delimitation: MinEP as the leading order of the occupation-time rate functional around detailed balance, Eq. (8.4), with the frenetic contribution as the obstruction at higher order.

Maes, C. and Netočný, K. (2013). *Minimum entropy production principle.* Scholarpedia 8(7), 9664.
Relevance. Authoritative encyclopedic summary by the authors of the fluctuation-theoretic derivation; source for the consolidated validity conditions adopted in this Part.

Martyushev, L. M. (2010). *The maximum entropy production principle: two basic questions.* Philosophical Transactions of the Royal Society B 365, 1333–1334.
Relevance. The advocacy position restated: the principle cannot be proved, only corroborated; derivations require assumptions less obvious than the principle.

Martyushev, L. M., and Seleznev, V. D. (2006). *Maximum entropy production principle in physics, chemistry and biology.* Physics Reports 426, 1–45.
Relevance. The standard advocacy review; assembles the Ziegler formulation and the application record this Part assesses.

Martyushev, L. M., and Seleznev, V. D. (2014). *The restrictions of the maximum entropy production principle.* Physica A 410, 17–21.
Relevance. The advocates' own codification of the principle's restrictions: locality, fixed forces, separated scales.

Millidge, B., Tschantz, A., and Buckley, C. L. (2021). *Whence the expected free energy?* Neural Computation 33(2), 447–482.
Relevance. Shows the expected free energy G is not the expectation of the variational free energy of future trajectories, so the policy objective of active inference is an independent postulate rather than a consequence of the principle.

Neal, R. M. and Hinton, G. E. (1998). *A view of the EM algorithm that justifies incremental, sparse, and other variants.* In M. I. Jordan (ed.), Learning in Graphical Models, 355–368. Kluwer.
Relevance. Shows expectation–maximization is coordinate descent on one free energy functional; the template for the FEP's unification of inference and learning under a single objective.

Nicolis, C., and Nicolis, G. (1980). *On the entropy balance of the earth-atmosphere system.* Quarterly Journal of the Royal Meteorological Society 106.
Relevance. Early critique questioning whether Paltridge's extremum is recoverable from the underlying dynamics.

Onsager, L. (1931). *Reciprocal Relations in Irreversible Processes. I.* Physical Review 37, 405–426.
Relevance. Derives the reciprocity $L_{ik} = L_{ki}$ from microscopic reversibility — hypothesis (A3), on which the identity $\partial P/\partial X_j = 2J_j$ and hence the whole theorem depends.

Onsager, L. (1931). *Reciprocal Relations in Irreversible Processes. II.* Physical Review 38, 2265–2279.

Relevance. Formulates the principle of the least dissipation of energy, the flux-side variational principle of Proposition 5.1, to be distinguished from Prigogine's force-side theorem.

Onsager, L. and Machlup, S. (1953). *Fluctuations and Irreversible Processes.* Physical Review 91, 1505–1512.
Relevance. Recasts least dissipation as a statement about fluctuation paths — the seed of the large-deviation viewpoint from which MinEP is later derived as a first-order approximation.

Ozawa, H., Ohmura, A., Lorenz, R. D., and Pujol, T. (2003). *The second law of thermodynamics and the global climate system: a review of the maximum entropy production principle.* Reviews of Geophysics 41, 1018.
Relevance. The canonical geophysical review of the climate MEPP, its planetary tests, and its turbulence applications.

Paltridge, G. W. (1975). *Global dynamics and climate—a system of minimum entropy exchange.* Quarterly Journal of the Royal Meteorological Society 101, 475–484.
Relevance. The founding climate application: a zonal energy-balance model closed by an entropy extremum instead of dynamics.

Paltridge, G. W. (1978). *The steady-state format of global climate.* Quarterly Journal of the Royal Meteorological Society 104, 927–945.
Relevance. Restates the 1975 extremum as maximum entropy production of the internal transport, the canonical form of the climate MEPP.

Paltridge, G. W. (1981). *Thermodynamic dissipation and the global climate system.* Quarterly Journal of the Royal Meteorological Society 107, 531.
Relevance. Third installment of the climate MEPP series, developing the dissipation formulation.

Parr, T., Pezzulo, G., and Friston, K. (2022). *Active Inference: The Free Energy Principle in Mind, Brain, and Behavior.* MIT Press, Cambridge, MA.
Relevance. The consolidated textbook treatment of active inference as a process theory and modeling framework, the stratum of the program that stands independently of the physics-phase claims.

Polettini, M. (2013). *Fact-checking Ziegler's maximum entropy production principle beyond the linear regime and towards steady states.* Entropy 15, 2570–2584.
Relevance. Counterexamples beyond linear response and the constitutive-versus-phenomenological dichotomy adopted here as the standard MinEP/MaxEP resolution.

Prigogine, I. (1945). *Modération et transformations irréversibles des systèmes ouverts.* Bulletin de la Classe des Sciences, Académie Royale de Belgique 31, 600–606.
Relevance. The priority document: first statement and proof of the minimum entropy production theorem, including the monotonic decrease of P during relaxation.

Prigogine, I. (1947). *Étude thermodynamique des phénomènes irréversibles.* Desoer, Liège.

Relevance. The doctoral thesis in book form and the theorem's canonical early exposition, containing the thermodiffusion example of the relevant section of that paper.

Prigogine, I. (1978). *Time, Structure and Fluctuations* (Nobel lecture, December 8, 1977). Science 201, 777–785.
Relevance. Prigogine's own retrospective, presenting the linear-regime theorem together with the breakdown of its variational structure as the gateway to dissipative structures.

Prigogine, I. and Lefever, R. (1968). *Symmetry Breaking Instabilities in Dissipative Systems. II.* Journal of Chemical Physics 48, 1695–1700.
Relevance. The Brusselator: explicit demonstration that beyond a finite distance from equilibrium the stable regime is not the branch of least entropy production.

Raja, V., Valluri, D., Baggs, E., Chemero, A., and Anderson, M. L. (2021). *The Markov blanket trick: on the scope of the free energy principle and active inference.* Physics of Life Reviews 39, 49–72.
Relevance. Argues the blanket is induced by the modeler's parcellation rather than discovered in the system, with the Watt governor as central counterexample to universal scope.

Ramstead, M. J. D., Sakthivadivel, D. A. R., Heins, C., Koudahl, M., Millidge, B., Da Costa, L., Klein, B., and Friston, K. (2023). *On Bayesian mechanics: a physics of and by beliefs.* Interface Focus 13(3), 20220029.
Relevance. The programmatic statement of Bayesian mechanics, comparing the FEP to a principle of least action on belief spaces and making the deflationary reading official.

Rao, R. P. N. and Ballard, D. H. (1999). *Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects.* Nature Neuroscience 2(1), 79–87.
Relevance. The canonical hierarchical predictive-coding model that the Gaussian gradient dynamics of the FEP reproduce; an algorithm that neither requires nor implies the principle.

Rayleigh, Lord (J. W. Strutt) (1873). *Some General Theorems relating to Vibrations.* Proceedings of the London Mathematical Society 4.
Relevance. Introduces the dissipation function for linearly damped systems; Onsager's variational function is modeled on it.

Sawada, Y. (1981). *A thermodynamic variational principle in nonlinear non-equilibrium phenomena.* Progress of Theoretical Physics 66, 68.
Relevance. Proposes maximum entropy production as a selection principle among nonlinear nonequilibrium steady states, the "non-variational" branch of Bruers' taxonomy.

Schnakenberg, J. (1976). *Network theory of microscopic and macroscopic behavior of master equation systems.* Reviews of Modern Physics 48, 571–585.
Relevance. Defines the standard entropy production functional for Markov jump processes, Eq. (8.3), used in the stochastic formulation of MinEP.

Singh, M. S., and O'Neill, M. E. (2022). *The climate system and the second law of thermodynamics.* Reviews of Modern Physics 94, 015001.
Relevance. The modern consensus assessment: theoretical and modeling issues preclude broad acceptance of MEPP in climate science.

van Es, T. (2021). *Living models or life modelled? On the use of models in the free energy principle.* Adaptive Behavior 29(3), 315–329.
Relevance. Presses the map–territory discipline: FEP models are the scientist's instruments, and conflating model with target is an error the literature repeatedly invites.

Ziegler, H. (1958). *An attempt to generalize Onsager's principle, and its significance for rheological problems.* Zeitschrift für angewandte Mathematik und Physik 9, 748–763.
Relevance. First formulation of the orthogonality principle generalizing Onsager's least dissipation to nonlinear dissipation functions in continuum mechanics.

Ziegler, H. (1961). *Zwei Extremalprinzipien der irreversiblen Thermodynamik.* Ingenieur-Archiv 30, 410–416.
Relevance. Companion statement of the two extremum principles (force space and flux space) whose flux-space member is the maximum entropy production form used throughout this Part.

Ziegler, H. (1983). *An Introduction to Thermomechanics*, 2nd ed. North-Holland, Amsterdam.
Relevance. Textbook consolidation of the orthogonality principle and its plasticity applications; the canonical reference for Definition 11.2.

Ziman, J. M. (1956). *The general variational principle of transport theory.* Canadian Journal of Physics 34, 1256–1273.
Relevance. States and proves the one uncontested MaxEP theorem: the stationary solution of the linearized Boltzmann equation maximizes entropy production under the self-consistency constraint.