SCIENTIFIC CULTURE
www.sci-cult.com

SCIENTIFIC CULTURE, Vol. 12, No. 4, (2026), pp. 1-13
Open Access. Online & Print

DOI: 10.5281/zenodo.12426101

EPIGENETIC TELEONOMY: A CONTROL-THEORETIC FRAMEWORK FOR CONTEXT-TO-CHROMATIN TRANSDUCTION

Elias Rubenstein[1]*

[1] Independent Researcher, Fort Lauderdale, Florida, USA 2026

Received: 01/12/2025
Accepted: 02/01/2026

Corresponding author: Elias Rubenstein
(science@elias-rubenstein.com)

ABSTRACT

Epigenetic research explains how chromatin states store biological history, yet a recurring challenge remains: translating context-level variables (appraisal, learning, perceived threat/safety, structured interventions) into preregisterable hypotheses about when, where, and how measurable chromatin distributions should change. This paper introduces Epigenetic Teleonomy, a control-theoretic modeling framework that treats the epigenome as a regulated dynamical system relaxing toward an operationally defined reference regime. The practical value is a preregistration-ready language that converts context-to-chromatin questions into (i) explicit reference construction rules primarily from within-subject baselines (optionally weakly regularized by population priors), (ii) a single primary distribution-level endpoint quantifying deviation from that reference, and (iii) a mediated actuation hypothesis in which protocol-defined physical input trajectories influence chromatin dynamics through measurable neurophysiological mediator trajectories and downstream nuclear effectors, without requiring a direct chromatin-binding ligand as the explanatory primitive. The framework is designed to stand or fall on two preregistered discriminative predictions: (1) expectancy-matched discrimination between algorithmically assigned congruent versus mismatch inputs undera dictionary-locked congruence definition with adequacy gates, and (2) preregistered early ordering in variance-sensitive fast-proxy modules that precedes detectable phenotype change. The result is a reviewer-readable bridge from inputs to chromatin dynamics that is falsifiable via preregistered discrimination tests and out-of-sample identification of a constrained mediator-to-module actuation map.

KEYWORDS: Epigenetics; control theory; allostasis; information theory; Kullback–Leibler divergence; Jensen–Shannon divergence; variational free energy; placebo effects; structured inputs; signal encoding; spectral representations; chromatin remodeling; system identification.

1. INTRODUCTION: THE TRANSDUCTION GAP

Epigenetics has matured from a static "histone code" metaphor into an experimentally grounded account of chromatin state, memory, and regulatory plasticity [1, 2]. Yet a persistent explanatory gap remains: the **Transduction Gap**—how context-level variables (context, appraisal, learning history, perceived threat/safety) become measurable molecular reorganizations (methylation, histone marks, accessibility) with module-level specificity. Reports of relatively rapid molecular shifts are heterogeneous across tissues, assays, and time windows; the present paper treats timescale as an empirical question to be adjudicated by preregistered sampling schedules and discrimination tests, not as an assumption.

Placebo-related effects illustrate that context and learning can recruit brain and neuromodulatory systems that influence physiology; the literature emphasizes roles for learning, expectations, and social cognition and implicates multiple neurochemical mediators, including opioid and dopamine systems [6, 7]. Similarly, psychotherapy has been shown to correlate with methylation changes in stress-related genes [12, 13]. Standard models explain how marks persist once set (e.g., maintenance processes), but often leave the directionality and targeting of coherent reorganization implicit. More broadly, recent work has argued that biological regulation can be described in computational and goal-directed terms across scales, motivating an explicit formulation of the context-to-chromatin link as a control and decoding problem [11]. A central practical obstacle is the **reference problem**: many phenotypes do not admit a true "pre-disturbance" epigenetic operating profile, and within-subject baselines can drift, vary with season/circadian timing, or be confounded by cell composition and batch effects. This paper therefore treats reference construction as a preregistered decision problem with explicit stability checks, rather than an informal assumption.

This paper proposes that chromatin regulation can be formalized as a reference-regime controlproblem: stability is maintained by feedback-driven minimization of deviation from an operational reference operating regime ("teleonomy" meaning evolved reference-regime regulation and error correction—a functional, mechanistic notion distinct from teleology, and not retrocausality) [3].

1.1. Contributions and claim hierarchy (to prevent scope drift)

To make the manuscript falsifiable and resistant to post-hoc narrative repair, the claims are explicitly hierarchical:

Primary claim (discriminative): Under empirically verified expectancy matching, congruent inputs outperform mismatch inputs on a single preregistered distribution-level endpoint ΔD with a preregistered minimal effect size δ_D (Section 7.2).

Key secondary claim (directionality): In the congruent arm, preregistered module subsets $S_k C(i)$ show the predicted directionality relative to mismatch, using a preregistered multiple-testing control (Methods).

Tertiary claim (mechanistic identification, constrained): A mediator-to-epigenome actuation map is identifiable only under a strict low-dimensional subspace constraint with out-of-sample performance criteria; failure implies broad modulation rather than selective steering (Methods).

Optional modeling layer: The SDE/control formalism is an effective description intended to guide estimators and preregistration; it is not required for the primary discriminative falsification.

2. MATERIALS AND METHODS: PREREGISTERED EXPERIMENTAL PROGRAM

2.1. Minimal empirical program (what is required vs optional)

Required for primary falsification: three-arm randomized design (C/M/X), dictionary adequacy gate, expectancy matching gate, preregistered reference construction with stability gate, a single primary endpoint ΔD, a preregistered minimal effect size δ_D, and a preregistered C vs M discrimination test.

Required for secondary claim: preregistered module subsets $S_k C(i)$ with multiplicity control and directionality tests.

Optional: SDE/state-space estimation of κ and lag kernels; mechanistic identification of $\mathbf{B}_s$ (only under strict subspace constraint).

2.2. Study design overview

Design: randomized, controlled, three-arm discriminative trial with two-active-protocol expectancy matching.

2.2.1. Arms

1. **Congruent protocol (C):** class assigned by the dictionary-locked congruence algorithm.
2. **Mismatch protocol (M):** identical

format/contact time/scaffolding; class sampled from preregistered mismatch set.

3. **Context-only control (X):** supportive interaction without protocol-defining structured content.

Key requirement: expectancy and engagement must be empirically matched between C and M (hard gate) and dictionary adequacy must pass (hard gate).

2.3. Participants and eligibility

Inclusion/exclusion criteria depend on the target phenotype; preregister a narrow phenotype to minimize heterogeneity. Exclude conditions where tissue remodeling is structurally blocked or where acute medical interventions dominate outcomes. Preregister rules for circadian sampling windows, acute infection, major medication changes, and recent major stressors.

2.4. Reference definition (primary: within-subject baseline with stability gate)

The primary reference is within-subject: $Q_i^{\text{base}}(\mathbf{y})$ is constructed from preregistered baseline samples for participant i using the reference decision tree and baseline stability gate described in the main text. Preregister the minimum baseline samples n_B and the stability test. Population priors may be used only as a weak regularizer when baseline sampling is limited, and this choice is preregistered. If Regime S is admissible, define Q^{post} only from a preregistered post-stability window.

2.5. Feature definition y, endpoint set, and distributional parameterization θ

Preregister the observable feature map and define $\mathbf{y}$ at the endpoint/module level to reduce multiple-testing burden. Then preregister the distributional parameterization: $P_y(\cdot, t) \approx P_y(\cdot \mid \theta(t))$.

Hard dimensionality constraint (operational necessity): Divergence estimation is unstable in high dimensions at typical sample sizes; therefore the endpoint dimension is preregistered and bounded:

$$m \le m_{\max}, \qquad m_{\max} = 10.$$

If a higher-dimensional representation is desired, it must be handled as an exploratory layer; confirmatory divergence endpoints remain bounded by $m_{\max}$.

2.6. Primary divergence estimator D_{KL} and mandatory sensitivity endpoints (preregistered)

The primary endpoint $D(t) = D_{KL}(P_y(\cdot, t) \,\|\, Q)$ is estimated on the preregistered low-dimensional endpoint representation.

Default estimator (confirmatory, stability-first). As the confirmatory default, summarize endpoints with *diagonal* Gaussian covariance (maximally stable at modest N):

$$P_y(\cdot \mid \theta(t)) \approx \mathrm{N}(\mu(t), \mathrm{diag}(\sigma^2(t))), Q(\cdot) \approx \mathrm{N}(\mu^*, \mathrm{diag}(\sigma^{*2})).$$

Compute the corresponding closed-form $\hat{D}_{KL}$ under the diagonal model (implemented componentwise). Full-covariance Gaussian KL is reported as a preregistered sensitivity estimator only.

Preregistered covariance regularization (mandatory, non-adaptive). For any sensitivity analysis using full covariance, preregister one fixed rule:

Shrinkage covariance: $\Sigma \leftarrow (1 - a)\Sigma_{\text{sample}} + a\,\mathrm{diag}(\Sigma_{\text{sample}})$ with a preregistered shrinkage choice (e.g., Ledoit–Wolf) and a preregistered eigenvalue floor $\lambda_{\min}$ applied to ensure PSD and stable determinants.

Mandatory symmetric sensitivity endpoints (preregistered). To guard against KL asymmetry and model misfit, report (regardless of outcome): Jensen–Shannon divergence $\mathrm{JSD}(P_y, Q)$ as a symmetric information-theoretic sensitivity endpoint at least one distribution-free distance on the same fixed representation (e.g., MMD or Wasserstein-type distance), with fixed model class and hyperparameter ranges.

Non-negotiable rule. All estimators, preprocessing, tuning rules, and the confirmatory/sensitivity roles are fixed in preregistration. Post-hoc endpoint/estimator switching is not permitted.

2.7. Assay strategy and time-scale alignment (preregistered)

Because response times differ by assay, the sampling schedule is tied to preregistered relaxation windows. A minimal preregistered multi-timepoint scaffold:

1. t_0: baseline window (repeated samples; $n_B \ge 5$).
2. t_1: early window (e.g., 1–3 hours) for mediators and fast proxies.
3. t_2: consolidation window (e.g., 48–168 hours) for distribution-level chromatin endpoints.
4. t_3: optional stability window (e.g., 7–14 days) used only for the preregistered Regime S adjudication.

Preregister tissue choice and justify surrogacy limits (e.g., blood as proxy vs. target tissue).

2.8. Mediator panel s(t) and Identification Strategy (Strict Subspace Constraint)

Mediator panel. Preregister a mediator panel $\mathbf{s}(t)$ consisting of measurable physiological outputs plausibly linked to stress/relaxation pathways

(e.g., HRV components, salivary cortisol/alpha-amylase, inflammatory cytokines), plus mandatory nuisance tracking (sleep, circadian timing, acute illness flags).

The preregistered strict subspace constraint (identifiability requirement). To resolve the identifiability crisis inherent in mapping physiological mediators to high-dimensional chromatin states with clinical sample sizes ($N \approx 50$–200), impose a Strict Subspace Constraint: the target parameter vector for identification is restricted to $\theta_{\text{sub}}(t) \in \mathrm{R}^M$ with $M \ll N$ (e.g., $M \leq 5$). The components of θ_{sub} must be fixed composite scores defined by independent prior literature (e.g., a specific epigenetic clock residual, a fixed glucocorticoid sensitivity score, a fixed inflammatory methylation index). These definitions are dictionary-locked prior to data collection; no data-driven feature extraction is permitted for identification of $\mathrm{B_s}$.

Discrete-time identification model. For sampling times t_0 and t_2, define $\Delta\theta_{\text{sub},i} = \theta_{\text{sub},i}(t_2) - \theta_{\text{sub},i}(t_0)$. Using a preregistered kernel family $h(\tau)$ to aggregate mediator history (capturing lag), define:

2.9. Expectancy matching and manipulation checks (hard gate)

Measure expectancy, credibility, and engagement at multiple timepoints (baseline, post-randomization, post-session) using preregistered instruments (e.g., a credibility/expectancy questionnaire and a fixed engagement scale). Define:

Equivalence margin: ϵ (preregistered; e.g., ϵ = 0.2 SD on expectancy).

Bayesian threshold: τ (preregistered; e.g., $BF_{01} > \tau$ with $\tau = 10$).

Hard gate: C and M are considered expectancy-matched only if both conditions hold:

1. TOST equivalence for mean expectancy difference lies within $\pm\epsilon$ across preregistered scales.
2. A preregistered Bayesian equivalence criterion supports practical equality (e.g., $BF_{01} > \tau$).

Blinding integrity (mandatory). To prevent leakage, preregister: a neutral framing script presenting both C and M as two valid protocol variants, automated delivery where feasible (audio/video/app) and fidelity logs, a post-session arm-guess check (participant and delivery staff) as a blinding integrity metric. If the expectancy gate fails, the C vs M discriminative claim is not evaluated; results are reported as context-dominated.

Ethics note (preregistered). If mismatch inputs require partial disclosure to maintain expectancy matching, preregister the disclosure/debriefing procedure and require prior ethics approval for any empirical implementation

2.10. Objective congruence and mismatch construction (dictionary-locked)

Compute baseline deviation $\mathbf{e}_i(t_0) = \theta^* - \theta_i(t_0)$. Use the dictionary-locked $D = \{(k, \mathbf{v}_k, S_k)\}$ to assign congruent class $k_C(i) = \arg\max_k A_{i,k}$ with $A_{i,k}$ defined in the main text. Define mismatch set $K_M(i) = \{k : A_{i,k} \leq a_0\}$ and sample k from that set, with a_0 preregistered and tie-breaking rules fixed. Apply dictionary adequacy gates before eligibility for the primary C–M discrimination endpoint.

2.11. Intervention specification u(t) (protocol encoding)

Preregister each input as a structured sequence: timing, content units, delivery channel, instruction sequence, interaction structure, and fidelity metrics. Define mismatch inputs as class-incongruent while preserving superficial structure and measured expectancy.

2.12. Sampling schedule and endpoints (time-scale alignment)

Timepoints (example): t_0 (baseline; repeated baseline samples), t_1 (1–3 hours post-intervention for mediators/fast proxies), t_2 (48–168 hours post-intervention for distribution-level chromatin endpoints), t_3 (7–14 days optional for post-stability regime adjudication).

Primary endpoint: $\Delta D_{\text{spec}} = D(t_0) - D(t_2)$, with confirmatory analysis on Regime R and preregistered sensitivity interpretation for Regime S (if admissible).

Minimal effect size: δ_D preregistered for C–M discrimination (e.g., in standardized units).

Mandatory sensitivity endpoints: JSD and one distribution-free distance on the same fixed representation.

Secondary endpoints: early entropy ordering $\Delta H_S(t_0 \to t_1)$; phenotype change score ΔP; mediator changes $\Delta \mathbf{s}(t_0 \to t_1)$.

2.13. Multiplicity control and endpoint hierarchy (preregistered)

To prevent multiple-testing inflation, preregister: a hierarchical testing scheme (primary ΔD first; only then key secondary module-directionality tests), an alpha allocation plan (e.g., $a = 0.05$ for primary; secondary controlled via FDR or alpha-splitting with fixed thresholds), fixed rules for reporting all sensitivity analyses regardless of outcome.

2.14. Statistical analysis plan (minimal, strong, falsifiable)

2.14.1. Primary discriminative test (clean coding; context as reference)

Define indicators $I_C = 1$ for Congruent protocol, else 0, and $I_M = 1$ for Mismatch protocol, else 0; Context-only is the reference level ($I_C = I_M = 0$). Fit a preregistered model:

$$\Delta D_{\text{spec}} \sim \beta_0 + \beta_C I_C + \beta_M I_M + \beta_S S_{\text{diag}} + \beta_{CS}(I_C \times S_{\text{diag}}) + \beta_{MS}(I_M \times S_{\text{diag}}) + \mathbf{c}^\top \mathbf{Z} + \epsilon, \quad (3)$$

where $\mathbf{Z}$ includes preregistered covariates (cell composition, batch, baseline D, medication flags, circadian window, acute illness flags, etc.).

Key discrimination test: $(\beta_C - \beta_M) \geq \delta_D$ under passing gates.

Key moderation test: $(\beta_{CS} - \beta_{MS}) > 0$.

Falsification: $(\beta_C - \beta_M) < \delta_D$ or $(\beta_{CS} - \beta_{MS}) \leq 0$ under passing gates.

2.14.2. Secondary directionality tests (module subsets)

Test preregistered directionality in $S_k C\,(i)$ with preregistered multiplicity correction and preregistered effect direction. No post-hoc module discovery is permitted for confirmatory claims.

2.14.3. Early ordering test

Test whether entropy reduction precedes phenotype change: preregister an ordering criterion such that $\Delta H_S(t_0 \to t_1) < 0$ predicts $\Delta P\,(t_0 \to t_2)$ in the congruent arm more strongly than in mismatch/context arms.

2.14.4. Estimating κ and lag parameters (optional modeling layer)

Estimate κ as an effective contraction rate (Appendix A), either from decay of $D(t)$ or via state-space fitting to $d\theta(t)$ with measured mediators $\mathbf{s}(t)$ and preregistered kernel family $\mathbf{H}(\tau)$ (e.g., exponential family with parameter τ_r).

2.15. Batch and cell-composition preregistered diagnostics (fail rules)

To reduce "this is batch" objections to a checkable decision rule, preregister:

Batch-predictability check: test whether batch/plate labels predict arm assignment or ΔD_{spec} beyond a preregistered threshold; if yes, interpret results as confounded (report as failure of interpretability).

Composition-shift check: preregister a cell-mix shift threshold; if exceeded, analyze within a preregistered stratum or exclude by rule (report transparently).

2.16. Power and sample size (preregistered logic)

Because effect sizes are uncertain, preregister a staged approach: (1) pilot (feasibility + estimator variance); (2) main study powered on the interaction $(\beta_{CS} - \beta_{MS})$ and on C vs M discrimination with δ_D, with preregistered minimum detectable effects.

2.17. Data availability and reproducibility

Preregister full protocol scripts, DBP dictionary artifacts, feature definitions, QC pipeline, baseline stability tests, blinding/randomization procedure, kernel family for lag modeling, multiplicity control plan, and analysis code. Archive the locked congruence dictionary, its hash, and all input definitions with timestamped identifiers prior to recruitment.

3. FORMALIZING THE EPIGENETIC STATE SPACE

To avoid vague metaphors, we adopt a formalism compatible with Waddington's original vision [14] but stated in operational, preregisterable terms.

3.1 Notation map (to prevent notation drift)

1. Latent microstate: $\mathbf{x} \in X$; microstate distribution $P(\mathbf{x}, t)$.
2. Observable assay space: $\mathbf{y}_{\text{obs}} \in Y$; induced distribution $P_y(\mathbf{y}_{\text{obs}}, t)$.
3. Low-dimensional representation: $P_y(\cdot\,; t) \approx P_y(\cdot \mid \theta(t))$ with preregistered $\theta(t)$.
4. Reference distribution: $Q(\mathbf{y})$; shorthand $\sigma_Z \equiv Q(\mathbf{y})$.

3.2. Microstate distribution $P(x, t)$

Let $\mathbf{x} \in X$ denote a high-dimensional microstate vector (e.g., methylation states at preregistered regulatory CpGs, chromatin accessibility at peaks/modules, compositional chromatin-state features). We represent the organism's instantaneous epigenetic condition as a probability distribution

$$P(\mathbf{x}, t) \quad (4)$$

rather than a single configuration, capturing stochasticity and cell-to-cell variability.

3.3. Observation model and induced distribution $P_y(y, t)$

Most experiments do not observe $\mathbf{x}$ directly. Define a preregistered, assay-accessible projection g and an observation model that includes measurement error:

$$\mathbf{y}_{\text{obs}} = g(\mathbf{x}) + \varepsilon, \varepsilon \sim \text{(preregistered error model)}. \quad (5)$$

Equivalently, assays define a conditional

measurement channel $P(\mathbf{y}_{obs} \mid \mathbf{x})$. This induces an observable-level distribution:

$$P_y(\mathbf{y}_{obs}, t) = \int_X P(\mathbf{x}, t)\, P(\mathbf{y}_{obs} \mid \mathbf{x})\, d\mathbf{x}. \quad (6)$$

All divergence and entropy endpoints below are defined on observable space (i.e., on P_y) to prevent notation drift between microstates and measured features.

Parametric control representation (to make gradients well-defined). For modeling and control, we work with a preregistered low-dimensional parameterization of the observable distribution:

$$P_y(\cdot\,; t) \approx P_y(\cdot \mid \theta(t)), \quad (7)$$

where $\theta(t)$ may comprise module-level means/covariances, fixed factor scores, fixed composite indices, histogram masses, or mixture responsibilities (specified in Methods). Dynamics and gradients are defined with respect to θ, not directly with respect to raw sample-level $\mathbf{y}_{obs}$.

3.4. Reference attractor $\sigma_Z \equiv Q(y)$ (operational reference regime)

Define a reference distribution

$$\sigma_Z \equiv Q(\mathbf{y}) \quad (8)$$

constructed *a priori* (preregistered). Because inter-individual epigenetic variability is large, the primary reference is within-subject.

Preregistered reference decision tree (operational). Because "pre-disturbance" profiles are often unavailable, $Q_i(\mathbf{y})$ for participant i is constructed by a preregistered decision rule:

1. **Type A (preferred): true pre-disturbance baseline.** If a stable baseline exists before phenotype onset (rare), use it.
2. **Type B (run-in baseline): best-available stable baseline.** Otherwise, collect a run-in baseline with repeated sampling prior to randomization.
3. **Type C (post-stabilization baseline):** If the phenotype is chronic, define Q_i from a preregistered "stability window" after clinical stabilization (with explicit limitations).
4. **Optional weak regularization: population prior.** Only if within-subject sampling is limited, use a weak age/tissue/context-matched population prior as a regularizer (never as the primary reference).

Minimum baseline sampling and stability gate (hard). To reduce regression-to-the-mean and drift artifacts, baseline construction requires:

A preregistered minimum number of baseline samples per participant, $n_B \geq 5$, with preregistered spacing and circadian window controls. A preregistered baseline stability test (CV/Drift check on θ and on D within the baseline window). If stability fails, the participant is excluded from reference-regime analyses or analyzed under a preregistered "unstable baseline" sensitivity stratum.

3.4.1. Regime-agnostic interpretation: Restoration vs. Shift (locked decision rule)

To avoid silently assuming "return to baseline", the framework preregisters two empirically adjudicated regimes:

1. **Regime R (Restoration):** Dynamics reduce deviation relative to a fixed baseline reference Q^{base}.
2. **Regime S (Shift):** Dynamics stabilize to a new regime; the relevant post-intervention reference is Q^{post} defined from a preregistered post-stability window.

3.4.2. Operationally:

Q^{base} is constructed from the preregistered run-in baseline (Type A/B/C).

Q^{post} is constructed only from preregistered post timepoints (e.g., t_2, t_3) if a post-stability criterion is met (same CV/Drift gate, preregistered).

Locked regime decision: If baseline stability passes and post-stability passes, classify as Regime R if $D(P(t_2) \mid Q^{base})$ decreases and remains near Q^{base} across the post window; classify as Regime S if $P(t_2)$ and $P(t_3)$ are mutually stable yet remain materially displaced from Q^{base} (thresholds preregistered). This regime label affects interpretation only; the primary discriminative claim is evaluated on the locked primary endpoint (Methods).

3.5. Deviation from reference as informational divergence

Define divergence from the reference at time t as

High D indicates probability mass displaced from the reference basin. Empirically, "field defects" and abnormal variability patterns in methylation have been investigated in cancer-adjacent normal tissue, supporting the idea that altered distributions—not only mean shifts—carry biological signal [4]. The use of information-theoretic functionals follows standard foundations [9,10].

Why $D_{KL}(P \mid Q)$ is primary (and what is checked). $D_{KL}(P \mid Q)$ is asymmetric; here it is used as an operational "distance from a preregistered reference regime" under the modeling convention that Q encodes the target operating regime. To protect against asymmetry artifacts and estimator

fragility, a symmetric divergence and a distribution-free distance are preregistered as mandatory sensitivity endpoints (Methods).

4. DYNAMICS: REGULATION GAIN *K* AND THE CONTROL LOOP

4.1. Potential landscape and regulated descent

Define an effective potential (error functional) $\Phi(\theta; \sigma_Z)$ whose minima correspond to stable operating regimes consistent with σ_Z. The simplest and most interpretable operational choice is:

$$\Phi(\theta; \sigma_Z) \equiv D_{KL}\ P_y(\cdot \mid \theta) \| Q\ , \qquad (10)$$

evaluated through a preregistered estimator from measured data (Methods).

4.2. Mediated and lagged actuation (operational transduction path)

A delivered input $\mathbf{u}(t)$ is not assumed to act directly on chromatin coordinates as a universal primitive. Instead, the operational hypothesis is that inputs are transduced through measurable mediator channels (autonomic, endocrine, immune), which then gate downstream transcription-factor and remodeling activity. We therefore model a mediated actuation path:

$$\mathbf{s}(t) = \mathrm{M}\ \mathbf{u}(t), \mathbf{c}(t)\ , \qquad (11)$$

where $\mathbf{s}(t)$ is a preregistered mediator vector and $\mathbf{c}(t)$ are measured context covariates (expectancy, credibility, engagement).

Because epigenetic coordinates do not generally respond on neural timescales, we incorporate a distributed lag (hysteresis) via a response kernel. The resulting stochastic dynamics are:

1. $\kappa \geq 0$ is **regulation gain** (effective error-correction strength; optionally normalized for reporting).
2. $\mathbf{s}(t)$ are **measured mediator signals** (e.g., HRV, cortisol/catecholamines, inflammatory cytokines), preregistered.
3. $\mathbf{B}_s$ maps mediator fluctuations to epigenetic coordinates; it is **unknown** and treated as an **empirically identifiable** mapping only under preregistered identifiability constraints (Methods).
4. $\mathbf{H}(\tau)$ is a response kernel capturing relaxation time(s) τ_r (assay-dependent, preregistered).
5. $\Sigma^{1/2} d\mathbf{W}(t)$ is stochastic perturbation (metabolic/environmental noise).

Status of the actuation map $\mathbf{B}_s$ (bounded claim). In biological systems, the actuation map from measured mediator channels to epigenetic change is generally not derivable from first principles. Accordingly, $\mathbf{B}_s$ is treated here as an *unknown, empirically identifiable, and potentially sparse* mapping whose existence and specificity are *hypotheses to be tested*. The framework preregisters (i) structural constraints on $\mathbf{B}_s$ (low-dimensionality/sparsity and stability), (ii) an explicit identification procedure with out-of-sample validation, and (iii) failure criteria under which any observed effects are interpreted as broad, non-specific modulation rather than selective steering (Methods).

4.3. Physical execution (no information–energy identification)

This framework does not treat information as an energy source. The energy budget remains endogenous (ATP, chromatin remodelers, transcriptional and signaling networks). The claim is that structured, informationally salient inputs can **bias** which endogenous programs engage and how strongly through mediator channels, without assuming that chromatin-level specificity must originate from a direct chromatin-binding ligand. ATP-dependent chromatin remodeling complexes provide a concrete substrate for state transitions and reconfiguration [5].

4.4. On the meaning of $\nabla\Phi$ (effective description)

The gradient-flow notation is an effective description, not a claim that cells explicitly compute derivatives. Φ is an operational functional whose decrease summarizes relaxation toward a reference regime under feedback constraints. A natural candidate in related literature is variational free energy [10]; accordingly, Φ may be interpreted as a free-energy-like functional under preregistered approximations, without asserting a literal optimization algorithm at the molecular level.

5. INFORMATIONAL INPUTS AND DISCRIMINATION TESTING

5.1. Functional carrier equivalence (restricted and testable)

"Carrier equivalence" is not the claim that substance never matters. The restricted claim is: Different physical carriers can be functionally equivalent *only if* they converge on overlapping decoding pathways and produce empirically similar mediator trajectories $\mathbf{s}(t)$ under controlled context conditions, and only then produce similar downstream epigenetic dynamics.

A convergent transduction chain can be written as:

Input → Perception/Inference → Mediator

Output $\mathbf{s}(t)$ → Nuclear Effectors → Chromatin Remodeling.

Context and learning can modulate symptoms and physiology via brain systems and neurochemical mediators, providing a plausible physiological channel for context-to-chromatin transduction [6,7, 10].

5.2. Protocol-defined informational inputs (Physically instantiated trajectory framework)

To avoid methodological narrowing while keeping the framework falsifiable, we define the intervention class operationally as any *preregistered, physically instantiated* input trajectory $\mathbf{u}(t)$ delivered to the organism under controlled conditions. The term "informational input" is used in a strictly operational sense: it denotes reproducible structure in the delivered physical carrier that can be quantified independent of outcomes (timing, intensity, sequencing, and—where applicable—spectro-temporal representations).

Physical premise (reviewer-tight). The framework adopts a minimal physical premise: any input that is *experienced* by an organism must be physically instantiated, i.e., realized as a time-evolving state of some carrier (mechanical, electromagnetic, electrochemical, or molecular).

In that sense, **all experienceable inputs are time-varying physical trajectories** and can be represented in an equivalent basis (e.g., a spectral basis for continuous-time signals) by standard signal decompositions [8,9]. This is a representation-level statement about physical trajectories and does not posit a single privileged microscopic mechanism.

Carrier-agnostic operational specification. Depending on modality, $\mathbf{u}(t)$ may be specified as: an acoustic waveform (air-pressure dynamics) with preregistered amplitude envelope, timing and prosodic constraints, a visual sequence (electromagnetic input) with preregistered luminance/contrast dynamics and timing, a linguistic script with preregistered temporal structure and delivery constraints, a behavioral action schedule with preregistered sequencing and fidelity metrics, a pharmacological exposure characterized by identity, dose, and a preregistered pharmacokinetic time course (a controlled molecular-state trajectory at the organism level). The framework does *not* assume a single privileged carrier; it tests whether different carriers that produce matched expectancy and comparable mediator trajectories can nevertheless be discriminated by downstream chromatin endpoints under dictionary-locked congruence rules.

Coupling/entrainment is an empirical hypothesis. Biological mediator systems are oscillatory across scales, and stimulus-to-mediator coupling can manifest as entrainment, coherence, or cross-frequency interactions in preregistered mediator signals [15–17]. The present paper does not require a universal "resonance mechanism". Any coupling between input features and mediator dynamics is treated as an empirical question adjudicated by preregistered mediator analyses and out-of-sample identification criteria.

5.2.1. Complex structured inputs: scripts, sequences, and interaction protocols

Complex inputs (e.g., multi-step scripts, interpersonal interaction protocols, or extended sequences) are treated as *structured dynamical programs* rather than as semantic entities. They are preregistered by: a fixed sequence of units (steps) with timestamps and duration bounds, physically measurable delivery constraints (e.g., amplitude/prosody bounds, pacing, pauses), interaction structure constraints (turn-taking, response windows, permitted deviations), fidelity metrics (adherence scores, timing deviation, completed-step proportion).

6. RESULTS

6.1. Primary output: preregistered discrimination endpoint

The primary result of the framework is a preregistration-ready decision rule based on the distribution-level endpoint ΔD_{spec} and a fixed minimal effect size δ_D, evaluated under passing dictionary adequacy and expectancy gates.

6.2. Secondary outputs: directionality and early ordering

Secondary results are (i) preregistered directionality tests in the locked module subset $S_k C$ (i) under preregistered multiplicity control, and (ii) preregistered early ordering in fast-proxy modules (entropy reduction preceding detectable phenotype change), with responders defined independently of ΔD.

7. FALSIFIABLE PREDICTIONS

7.1. Context effects as generic gain vs. protocol-specific selective module recruitment

Context effects can increase global regulatory engagement (increase effective gain; reduce generalized noise) without necessarily producing selective module reorganization. Protocol-specific effects are hypothesized to produce **selective**

recruitment visible as differential reorganization in preregistered endpoints *conditional on* identifiability constraints, dictionary adequacy gates, and successful expectancy matching.

7.2. Prediction 1: Discriminative diagnostic moderation under expectancy matching

Let $S_{\text{diag}} \in [0, 1]$ be a preregistered diagnostic certainty score (from a standardized algorithm). Define

$$\mathrm{E}[\Delta D_{\text{spec}} \mid \text{Congruent}] = g(S_{\text{diag}}), \quad g' > 0. \quad (14)$$

7.2.1. Independent congruence definition (dictionary-locked, algorithmic)

To prevent post-hoc reinterpretation, congruence is defined independently of outcomes via a preregistered, dictionary-locked target library with adequacy gates. Let the **Congruence Dictionary** be

$$\mathrm{D} = \{(k, \mathbf{v}_k, \mathrm{S}_k)\}_{k=1}^{K} \quad (15)$$

where each entry k defines (i) a fixed target direction $\mathbf{v}_k$ in parameter space and (ii) an associated preregistered module subset S_k (modules expected to respond if k is correct).

Dictionary Build Protocol (DBP): external, fixed, auditable; no leakage. Each v_k is derived *before recruitment* under a preregistered Dictionary Build Protocol (DBP) and is archived with a timestamped public identifier (e.g., OSF/Zenodo) prior to recruitment and prior to any outcome analysis. For each entry k, DBP records:

1. the concrete source list (datasets/papers/repos with DOI/Accession),
2. inclusion/exclusion rules and transformation steps,
3. the fixed feature space $\mathbf{y}$ (module/score vocabulary) and platform mapping rules (assay harmonization),
4. the estimator producing $\mathbf{v}_k$ from sources and the normalization rule $\mathbf{v}_k \propto \Delta D_{\text{spec}}/|\Delta D_{\text{spec}}|$,
5. a final dictionary hash (freeze identifier) and timestamp.

No post-hoc edits to D are permitted, and no information from trial outcomes may influence dictionary construction (directly or indirectly).

Dictionary adequacy gate (hard). To prevent degenerate mismatch assignment, a participant is eligible for C–M discrimination only if all preregistered adequacy criteria pass:

- minimum classes $K \geq K_{\min}$,
- minimum mismatch-set size $|\mathrm{K}_M(i)| \geq K_{M,\min}$,
- minimum alignment spread $\mathrm{Var}_k(\mathrm{A}_{i,k}) \geq v_{\min}$.

If the gate fails, the participant enters a preregistered "dictionary-inadequate" stratum and is excluded from the primary discrimination analysis (reported transparently).

Operational congruence rule. Let the participant's baseline error vector be $\mathbf{e}_i(t_0) = \theta^* - \theta_i(t_0)$, where θ^* is the individual reference (personal baseline attractor parameterization). Define the alignment score

The **congruent** assignment is $k_C(i) = \arg\max_k \mathrm{A}_{i,k}$ (ties broken by a preregistered rule). The **mismatch** assignment is sampled uniformly from a bottom-quantile set

$$\mathrm{K}_M(i) = \{k : \ \mathrm{A}_{i,k} \leq a_0\}, \quad (17)$$

with preregistered threshold a_0 (e.g., $a_0 = 0$ or the 20th percentile of $\mathrm{A}_{i,k}$ across k). This guarantees mismatch by construction while preserving format, contact time, and expectancy matching.

Hard falsification rule (no escape-hatch). If, under expectancy matching and dictionary adequacy, the congruent arm fails to outperform mismatch on the preregistered primary endpoint by at least δ_D and the preregistered module subset $\mathrm{S}_{k_C(i)}$ fails to show the predicted directionality relative to mismatch, then the specificity claim (and the validity of D for this phenotype) is rejected for this input class. Failure cannot be reinterpreted as "incorrect congruence" because congruence is dictionary-locked and algorithmically assigned.

Expectation-matched discriminative design (two-active-protocol framing):

1. **Protocol A (C):** class $k_C(i)$ assigned by the dictionary-locked congruence algorithm.
2. **Protocol B (M):** a class $k \in \mathrm{K}_M(i)$ sampled from the preregistered mismatch set, with identical format, contact time, and scaffolding.
3. **Context-only (X):** supportive interaction without protocol-defining structured content.

Falsification: If ΔD^{CM} spec
under matched expectancy and engagement, the discrimination claim fails (effects reduce to nonspecific context).

7.3. Prediction 2: Early ordering in preregistered fast-proxy modules

Define a preregistered set of variance-sensitive and plausibly fast-responding proxy modules S, chosen by explicit criteria:

1. **Assay-feasibility criterion:** modules measured reliably at early windows in the chosen assay/tissue.
2. **Prior constraint:** fixed list derived from

independent prior literature or prior non-overlapping cohorts; no data-driven module discovery for primary/secondary endpoints.

3. **Multiplicity control:** module count and testing hierarchy are locked prior to unblinding (Methods). For each module $i \in S$, preregister a discretization into K metastable bins (or mixture components). Let $p_{i,k}(t)$ denote the estimated probability mass of module i in bin/component k at time t after cell-composition adjustment and QC. Define module entropy:

with preregistered weights w_i (e.g., uniform or based on baseline variance).

Responder definition (preregistered; independent of ΔD). For Prediction 2, "responders" are defined *before* unblinding by fixed thresholds on a phenotype/clinical outcome score only (or an independent clinical composite), e.g.,

$$\text{Responder}_i := (\Delta P_i \geq p_0),$$

with preregistered p_0 and a preregistered rule for handling missingness. This prevents circular selection on the primary divergence endpoint.

Prediction: in responders,

$$\Delta H_S(t_{\text{early}}) < 0 \text{ before } \Delta\text{Phenotype}(t) \text{ becomes detectable.} \quad (20)$$

8. DISCUSSION: WHAT THE FRAMEWORK DOES AND DOES NOT CLAIM

8.1. No retrocausality, no thermodynamic violation

"Teleonomy" here means reference-regime error correction produced by evolved architecture (allostasis), not backwards causation [3]. The system consumes energy to maintain and restore order; delivered inputs are transduced through mediator channels that bias endogenous program engagement.

8.2. Why a distribution-level divergence endpoint is useful

A distribution-level endpoint formalizes "distance from the reference basin" and avoids one-locus explanations, permitting explicit preregistered predictions about stability, drift, and recovery—compatible with modular, multi-factorial regulation [4]. Because divergence estimation can be fragile, the Methods preregister symmetric and distribution-free sensitivity endpoints and numerical safeguards.

8.3. Formal rival-hypothesis set (to prevent rhetorical framing)

To prevent narrative framing of "context-only" as a rhetorical category, the framework preregisters a minimal rival set:

1. **H0 (Context-only):** under passing expectancy gates, $\Delta \equiv \mathrm{E}[\Delta D \mid C] - \mathrm{E}[\Delta D \mid M] \leq \delta_D$ and $(\beta_{CS} - \beta_{MS}) \leq 0$.
2. **H1 (Selective recruitment):** under passing gates, $\Delta > \delta_D$ and preregistered directionality in $S_k C(i)$ holds relative to mismatch with preregistered multiplicity control.
3. **H2 (Dictionary irrelevant/mismatch):** $\Delta \leq \delta_D$ while context-only improvements may still occur (C, M, X can all improve); the specificity claim fails for this phenotype and dictionary.

8.4. Scope limitations (explicit)

1. The mediator-to-epigenome mapping $\mathbf{B}_s$ is unknown and must be identified under preregistered identifiability constraints; failure to identify it argues against selective recruitment.
2. Specificity is claimed at the preregistered endpoint level (primary ΔD; key secondary $S_k C(i)$ directionality), not as a guaranteed locus-level targeting claim.
3. Lagged responses are expected; early readouts use fast proxies, while stable methylation endpoints belong to later windows.
4. Congruence is dictionary-locked and algorithmic; it is not redefined after outcomes are known.

8.5. Likely failure modes (preregister to avoid post-hoc narratives)

1. Tissue specificity (blood vs. target tissue) and limited surrogacy.
2. Cell composition confounds and incomplete adjustment.
3. Lag/time-scale mismatch (assay not aligned with τ_r).
4. Weakly defined $Q(\mathbf{y})$ (reference not operationalizable or unstable at baseline).
5. Expectancy mismatches between arms (invalidating the discriminative test).
6. Over-flexible module definitions or endpoint switching (multiplicity inflation disguised as "discovery").

8.6. Preregistered falsification criteria (skeptic-facing checklist)

The following criteria are preregistered as decision rules adjudicable from preregistration artifacts, the locked dictionary, and the reported analyses:

1. **Dictionary adequacy gate passes (required):** $K \geq K_{\min}$, $|K_M(i)| \geq K_{M,\min}$, and alignment spread above $v_{\min}$. Otherwise the participant is not eligible for primary C–

M discrimination (reported transparently).

2. **Expectancy gate passes (required):** C and M satisfy preregistered equivalence on expectancy/credibility/engagement (TOST within $\pm\epsilon$ and Bayesian equivalence $BF_{01} > \tau$). If not, the C vs M specificity claim is not evaluated and results are reported as context-dominated.
3. **Primary discrimination succeeds or fails cleanly:** under passing gates, the preregistered primary endpoint satisfies $\Delta D^C - \Delta D^M \geq \delta_D$ (with preregistered model and covariates). If $\Delta D^C - \Delta D^M < \delta_D$, the specificity claim fails for this input class and phenotype under the locked dictionary.
4. **Directionality in the locked module subset:** preregistered directionality in $S_{kC}(i)$ holds relative to mismatch with preregistered multiplicity control; failure counts against selective recruitment even if the primary endpoint shifts.
5. **Sensitivity endpoints agree in sign:** Jensen–Shannon divergence and the preregistered distribution-free distance yield concordant qualitative conclusions (at minimum, no sign-flip indicating estimator fragility).
6. **Batch and cell-composition fail rules do not explain the effect:**

- Batch-predictability check: if preregistered batch/plate identifiers predict arm assignment or ΔD above a preregistered threshold, results are treated as confounded (reported as failure of interpretability).
- Composition-shift check: if preregistered cell-mix shift exceeds a preregistered threshold, the participant enters a preregistered stratum or is excluded by rule (reported transparently).

7. **Identification claim is conditional and falsifiable:** the mediator-to-module mapping passes preregistered out-of-sample criteria only under the strict subspace constraint; if CV performance does not exceed the permuted-mediator null, the mechanistic identification claim is rejected and interpretation is limited to broad modulation.

9. CONCLUSION

Epigenetic Teleonomy reframes chromatin dynamics as a control-theoretic relaxation problem: the epigenome is modeled as a distribution evolving under endogenous energy expenditure while being regulated toward a preregistered within-subject reference regime. Regulation gain κ summarizes effective error correction. Protocol-defined input trajectories are treated as physically encoded, preregisterable programs that are transduced through measurable mediator channels and may (as a falsifiable hypothesis) produce selective recruitment under preregistered constraints. The framework stands or falls on preregistered discrimination tests with a dictionary-locked congruence definition, explicit lag modeling, mandatory expectancy matching and dictionary adequacy gates, symmetric/sensitivity endpoints, estimator safeguards, and out-of-sample identification of a constrained mediator-to-module actuation map.

Author Contributions: Conceptualization, methodology, formal analysis, writing—original draft preparation, writing—review and editing: Elias Rubenstein.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable to this theoretical modelling paper. Any future empirical implementation of the proposed study design will require prior ethics approval by an appropriate review board.

Informed Consent Statement: Not applicable to this theoretical modelling paper. Any future empirical implementation will require informed consent from all participants.

Data Availability Statement: No new data were generated or analyzed. All derivations and estimators are specified in Appendix A, and all preregistration artifacts (dictionary with DBP, protocols, analysis code) are intended to be archived with a timestamped public identifier prior to recruitment.

Conflicts of Interest: The author declares no conflict of interest.

10. ABBREVIATIONS AND NOTATION

10.1. Assays / biology

- ATAC-seq: Assay for Transposase-Accessible Chromatin (sequencing)
- CpG: Cytosine–phosphate–guanine site
- DNA: Deoxyribonucleic acid
- HPA: Hypothalamic–pituitary–adrenal axis
- HRV: Heart rate variability

10.2. Statistics / modeling / design

- D_{KL} (KLD): Kullback–Leibler divergence
- JSD: Jensen–Shannon divergence
- SDE: Stochastic differential equation
- TOST: Two One-Sided Tests (equivalence testing)
- QC: Quality control
- $\mathbf{W}(t)$: Wiener process (Brownian motion)

10.3. Notation

- $\mathbf{x} \in X$: high-dimensional epigenetic microstate vector (latent biological configuration space)
- $P(\mathbf{x}, t)$: microstate distribution at time t
- $\mathbf{y}_{obs}$: assay-level observations (noisy)
- $P(\mathbf{y}_{obs} \mid \mathbf{x})$: measurement channel (assay noise model)
- $P_y(\mathbf{y}, t)$: induced observable distribution over Y
- $Q(\mathbf{y})$: preregistered reference distribution (primary: within-subject with stability gate; optional weak population regularizer)
- $\sigma_Z \equiv Q(\mathbf{y})$: reference attractor notation used for the operational reference regime
- $\theta(t)$: preregistered low-dimensional parameterization of $P_y(\cdot, t)$ (module moments, fixed indices, bin masses, mixture responsibilities)
- $\Phi(\theta; \sigma_Z)$: effective potential / error functional (operationally chosen as divergence to Q)
- $\kappa \geq 0$: regulation gain (effective strength of error-corrective control)
- $\mathbf{u}(t)$: delivered protocol-defined input trajectory
- $\mathbf{c}(t)$: measured context covariates (expectancy/engagement/credibility)
- $\mathbf{s}(t)$: measured mediator vector (autonomic/endocrine/immune signals)
- $\mathbf{B}_s$: unknown mediator-to-endpoint mapping (estimated under preregistered identifiability constraints)
- $\mathbf{H}(\tau)$: response kernel (lag / relaxation)
- Σ: diffusion/noise covariance in parameter dynamics
- δ_D: preregistered minimal effect size for C–M discrimination on the primary endpoint

A APPENDIX A. THE MATH LINK: D_{KL}, Φ, LYAPUNOV STABILITY, AND NOISE

A.1. Deterministic descent (core Lyapunov statement)

Let $\Phi(\theta; \sigma_Z)$ be continuously differentiable and bounded below, with a (local) minimum at θ^* corresponding to the reference basin implied by $Q(\mathbf{y})$. Consider the deterministic dynamics (set inputs and noise to zero):

estimated by fitting ln $D(t)$ vs time over preregistered windows.

REFERENCES

[1] Bird, A. (2007). Perceptions of epigenetics. *Nature*, 447(7143), 396–398. doi:10.1038/nature05913.

[2] Jenuwein, T., & Allis, C. D. (2001). Translating the histone code. *Science*, 293(5532), 1074–1080. doi:10.1126/science.1063125.

[3] McEwen, B. S., & Wingfield, J. C. (2003). The concept of allostasis in biology and biomedicine. *Hormones and Behavior*, 43(1), 2–15. doi:10.1016/S0018-506X(02)00024-5.

[4] Teschendorff, A. E., et al. (2016). DNA methylation outliers in normal breast tissue identify field defects that are enriched in cancer. *Nature Communications*, 7, 10478.doi:10.1038/ncomms10478.

[5] Clapier, C. R., et al. (2017). Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nature Reviews Molecular Cell Biology*, 18, 407–422. doi:10.1038/nrm.2017.26.

[6] Wager, T. D., & Atlas, L. Y. (2015). The neuroscience of placebo effects: connecting context, learning and health. *Nature Reviews Neuroscience*, 16, 403–418. doi:10.1038/nrn3976.

[7] Ashar, Y. K., et al. (2017). Brain Mechanisms of the Placebo Effect. *Annual Review of Clinical Psychology*, 13, 73–98. doi:10.1146/annurev-clinpsy-021815-093015.

[8] Landauer, R. (1961). Irreversibility and Heat Generation in the Computing Process. *IBM Journal of Research and Development*, 5(3), 183–191. doi:10.1147/rd.53.0183.

[9] Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27, 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x.

[10] Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11, 127–138. doi:10.1038/nrn2787.

[11] Levin, M. (2019). The Computational Boundary of a "Self". *Frontiers in Psychology*, 10, 2688. doi:10.3389/fpsyg.2019.02688.

[12] Quevedo, Y., et al. (2022). Potential epigenetic mechanisms in psychotherapy. *Frontiers in Human Neuroscience*, 16, 955005. doi:10.3389/fnhum.2022.955005.

[13] Roberts, S., et al. (2019). DNA methylation of FKBP5 and response to exposure-based psychological therapy. *American Journal of Medical Genetics Part B*, 180(2), 150–158. doi:10.1002/ajmg.b.32650.

[14] Waddington, C. H. (1957/2014). *The Strategy of the Genes*. Reprint edition. Routledge. ISBN: 9781138017313.
[15] Buzsáki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science*, 304(5679), 1926–1929. doi:10.1126/science.1099745.
[16] Canolty, R. T., & Knight, R. T. (2010). The functional role of cross-frequency coupling. *Trends in Cognitive Sciences*, 14(11), 506–515. doi:10.1016/j.tics.2010.09.001.
[17] Fries, P. (2015). Rhythms for cognition: communication through coherence. *Neuron*, 88(1), 220–235. doi:10.1016/j.neuron.2015.09.034.