Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Free-Energy Geometry and Dynamical Regime Switching under Feasibility Constraints A Transferable Dynamical Motif for Nonequilibrium Systems

Toshio SATO

toshioandfep@gmail.com

Independent researcher https://orcid.org/0009-0003-4628-0167

Research Article

Keywords: DBFEP, nonequilibrium systems, regime switching, U–S plane, Z(λ) = U − λS, first-, touch, present fit / future openness

Posted Date: April 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9114900/v2

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Free-Energy Geometry and Dynamical Regime Switching under Feasibility Constraints

A Transferable Dynamical Motif for Nonequilibrium Systems

Toshio SATO

Abstract

We develop a statistical-mechanical framework for nonequilibrium systems that exhibits **regime switching** under feasibility constraints. The framework is organized around a **free-energy**–like evaluation geometry on a thermodynamic-style U–S plane, where U summarizes mismatch-/cost-like contributions and S summarizes uncertainty-/openness-like contributions. We define $Z(\lambda) = U - \lambda S$ not as a scalar objective required to decrease monotonically, but as a parametrized evaluative ruler whose rotation with λ changes the relative weighting of the two contributions.

A central claim is that the relevant type asymmetry is structural: one evaluative term is present-conditioned and point-like, whereas the other is future-oriented and trajectory-based. Selection is therefore formulated by a feasibility-aware "first-touch" rule: among admissible branches, the preferred regime is identified by the first feasible contact between an iso-Z ruler and the candidate bundle. This geometry yields delayed switching, hysteresis, and path dependence without requiring a global optimization narrative.

To show that the construction is compatible with a **dynamical system** perspective, we also present a concrete mechanical realization that implements the same rotation–contact–regime-change logic as a minimal periodic dynamics. The result is a transferable dynamical motif for constrained nonequilibrium systems in which free-energy geometry, feasibility boundaries, and regime switching can be analyzed within a common descriptive framework.

1 Introduction

Understanding how systems switch between qualitatively different regimes under resource and feasibility constraints is a central problem in nonequilibrium dynamics. In many settings, behavior is not well described as the monotonic minimization of a single smooth objective. Instead, observed trajectories reflect the interaction of heterogeneous evaluative contributions with admissible regions of state space, leading to regime switching, history dependence, and constraint-induced breakdown.

The starting point of this paper is a simple type asymmetry between two variational evaluation terms that commonly arise in constrained nonequilibrium systems. One term is present-conditioned and point-like, defined locally with respect to the current state, whereas the other is future-oriented and trajectory-based, indexed by admissible control sequences and finite horizons. Treating these terms as if they were symmetric can obscure their distinct structural roles and encourages, by default, narratives based on joint minimization that are not generally warranted in the presence of feasibility constraints.

The Dynamic Balancing Free Energy Principle (DBFEP), as initially formulated in [1] and further documented in [2], treats this asymmetry as primary. Rather than proposing a new optimization principle, DBFEP introduces a descriptive geometric layer in which the two terms are co-located on a shared state space without identification. This representation makes regime switching, constraint-induced breakdown, and history dependence explicit, without assuming the existence of a smooth global objective function. In related variational formulations, these two contributions are often referred to as variational free energy (VFE) and expected free

energy (EFE); in the present work, however, they are treated as generic present-conditioned and future-oriented variational terms, independent of any domain-specific interpretation.

Concretely, we introduce a thermodynamic-style U–S evaluation plane under a minimal bridging assumption (embedding into a canonical family). Here U summarizes mismatch-/cost-like contributions and S summarizes uncertainty-/openness-like contributions. On this plane we define

$$Z(\lambda) = U - \lambda S, \tag{1}$$

but Z is not treated as a global objective required to decrease monotonically. Rather, iso-Z contours function as an evaluative ruler: candidate branches are ranked by contact with these rulers, and changes in evaluative attitude λ rotate the ruler family.

A key point is that not all states and branches are feasible. Constraints arising from resources, fatigue, safety margins, and institutional limits carve out an admissible region. Given a feasibility-truncated bundle of candidate branches, we define the selected local regime operationally by a feasibility-aware "first-touch" rule: the preferred regime is the first feasible contact between an iso-Z ruler and the candidate bundle. As λ varies continuously, the identity of the first-touch branch can change discontinuously, yielding explicit regime switching without assuming smooth global minimization. Introducing an inertia/attitude parameter r (a coarse-grained summary of slower internal time scales) produces delayed switching and hysteresis, and constraint-induced breakdown is predicted geometrically as loss of feasible contact.

To reduce the risk that these relationships are read as purely interpretive, we also present a concrete implementation. The same switching logic (rotation $\rightarrow$ contact $\rightarrow$ regime change) can be realized as a minimal periodic dynamics in a caster-board-like mechanism. This is not a claim that cognition is a caster board; it is an existence proof that the proposed dynamical motif is physically realizable. Finally, we introduce a projection-based methodological view: distinct observational projections of the same constrained switching dynamics can support different narrative descriptions, precisely because projection discards information. We therefore separate projection-dependent readings from projection-invariant structural claims and identify testable signatures at the level of switching dynamics.

Terminology and scope clarification. In the literature on the free-energy framework, the present-conditioned and future-oriented variational functionals are commonly referred to as variational free energy (VFE) and expected free energy (EFE), respectively. The present study originated from this decomposition; however, throughout this paper we treat these quantities as generic competing variational terms in constrained nonequilibrium systems, independent of any domain-specific interpretation. The conventional abbreviations are used only to clarify this origin and are avoided elsewhere in the manuscript.

Contributions. This paper provides a dynamical clarification of coupled variational evaluation under feasibility constraints. The main contributions are as follows:

1. **Shared geometric representation without identification.** We embed present-conditioned, point-like evaluation and future-oriented, trajectory-based evaluation into a shared U–S state space while preserving their structural asymmetry.

2. **Evaluation as a ruler rather than a global objective.** We use $Z(\lambda) = U - \lambda S$ as a parametrized evaluative ruler whose level sets rank admissible candidates without assuming monotonic minimization.

3. **Feasibility-aware selection and breakdown.** A first-touch rule selects the first feasible contact between iso-Z rulers and feasibility-truncated candidate bundles, allowing regime switching and predicting constraint-induced breakdown as loss of feasible contact.

4. **History dependence and realizability.** An inertia/attitude parameter induces delayed switching and hysteresis, and the same switching logic admits a minimal mechanical realization as an existence proof.

Paper roadmap. Sections 2–4 establish the evaluation space and introduce the shared U–S representation under minimal bridging assumptions. Sections 5–8 develop feasibility constraints, the first-touch selection principle, and inertia-induced hysteresis. Section 9 presents a concrete dynamical implementation as a mechanical existence proof. Sections 10–11 discuss projection-dependent interpretations and structurally testable predictions. Sections 12–14 relate the framework to existing variational formulations, discuss limitations, and summarize contributions.

Fig.1 Conceptual overview

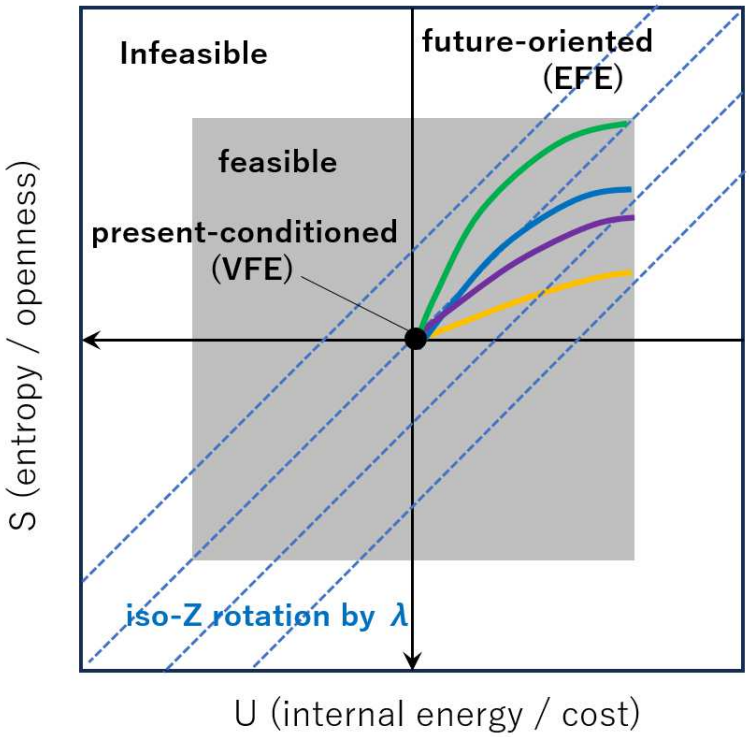


Figure 1: Conceptual overview of the DBFEP framework. Present-conditioned evaluation is embedded as a point-like (local) contribution in a shared U–S evaluation space. In contrast, future-oriented evaluation is embedded as a bundle of policy- and horizon-indexed trajectories in the same space. This explicit type asymmetry motivates a representation in which the two contributions are co-located without identification. Iso-Z contours of $Z(\lambda) = U - \lambda S$ function as evaluative rulers whose rotation with λ supports feasibility-aware selection via first feasible contact (first-touch).

2 Background and Scope

The free-energy framework (free-energy framework) and policy-based variational inference have been presented as unifying formalisms for perception, action, and learning [3, 4]. At the same time, their scope and the status of thermodynamic analogies remain debated. The present work

does *not* aim to defend free-energy framework as a universal theory, nor to provide a detailed neural implementation.

Our focus is narrower and structural: we address a class of phenomena that arise when present-oriented evaluation (present-conditioned variational term) and future-oriented evaluation (future-oriented variational term) interact under constraints—namely (i) regime switching, (ii) feasibility-induced breakdown, (iii) inertia and hysteresis, and (iv) bounded cyclic operation that is not well captured by monotonic optimization narratives. These phenomena are familiar in nonlinear dynamical systems, and we use that tradition as a source of descriptive vocabulary and sanity checks rather than as a commitment to any single psychological narrative.

Concretely, this paper provides three deliverables: (1) a shared state-space representation (the U–S plane) in which present-conditioned variational term and future-oriented variational term can be compared without identifying them; (2) a minimal translation assumption ("bridging axiom") that motivates placing log-probability-based quantities into energy/entropy-like coordinates; and (3) an operational selection rule (λ-rotation + first-touch under feasibility) that makes switching and breakdown geometrically explicit.

Importantly, we do not introduce new computational definitions of present-conditioned variational term or future-oriented variational term, nor do we offer first-person phenomenology or specific circuit models. The contribution is a geometric and dynamical *description layer*—intended to reduce category errors (e.g., interpreting present-conditioned variational term and future-oriented variational term as jointly minimized objectives) by making type asymmetries and constraints explicit.

3 Thermodynamic Grounding

This section defines the abstract state space used throughout the paper. The aim is *not* to claim a physical thermodynamic identity,[1] but to introduce a coordinate system in which evaluative quantities and constraints can be represented with a simple geometry.

Thermodynamic language can be illuminating when one works with log-probabilities and entropy-like functionals in Bayesian mechanics, but it can also invite category errors. We therefore keep the present paper's claims strictly at the level of *geometry and selection rules* in an abstract state space.[5, 6]

3.1 The U–S state space (operational coordinates)

We work on a two-dimensional plane with two scalar coordinates:

- U **(mismatch / cost proxy):** an operational scalar that aggregates "how badly things fit" (e.g., prediction mismatch, control/effort burden, or any scalarized penalty used at the descriptive level at hand).
- S **(uncertainty / openness proxy):** an operational scalar that aggregates predictive breadth (e.g., entropy of a predictive distribution, dispersion, or any comparable uncertainty measure).

The only structural requirement is that trade-offs between mismatch-like and uncertainty-like quantities can be represented on two axes, so that evaluation can be expressed by linear "rulers" on the plane.

[1] The symbols U and S are introduced as an *operational geometry*. While thermodynamic metaphors are common in the free-energy framework literature, here U and S denote information-theoretic proxies (constructed from log-probabilities and predictive breadths) rather than biological heat exchange or physical entropy.

3.2 The evaluative quantity Z and iso-Z rulers

We define an evaluative scalar

$$Z \equiv U - \lambda S, \tag{2}$$

where λ is an *attitude* parameter that controls the relative emphasis placed on uncertainty/openness versus mismatch/cost. For a fixed λ, the iso-Z sets are straight lines on the U–S plane ("iso-Z rulers"). This construction is purely geometric: we do *not* assume that organisms monotonically minimize Z over time. Instead, λ is treated as a controlled or slowly varying quantity that rotates the ruler and can induce switching under constraints.

3.3 Feasibility is distinct from evaluation

Not all points in the U–S plane are realizable. We represent realizability by a feasible set (admissible region) and emphasize that selection is a *first-contact* operation between an iso-Z ruler and the feasible set. Consequently, feasibility changes can drive switching even when λ is held fixed, and conversely, λ changes can fail to induce switching when feasibility blocks contact.

Fig.2 Thermodynamic grounding

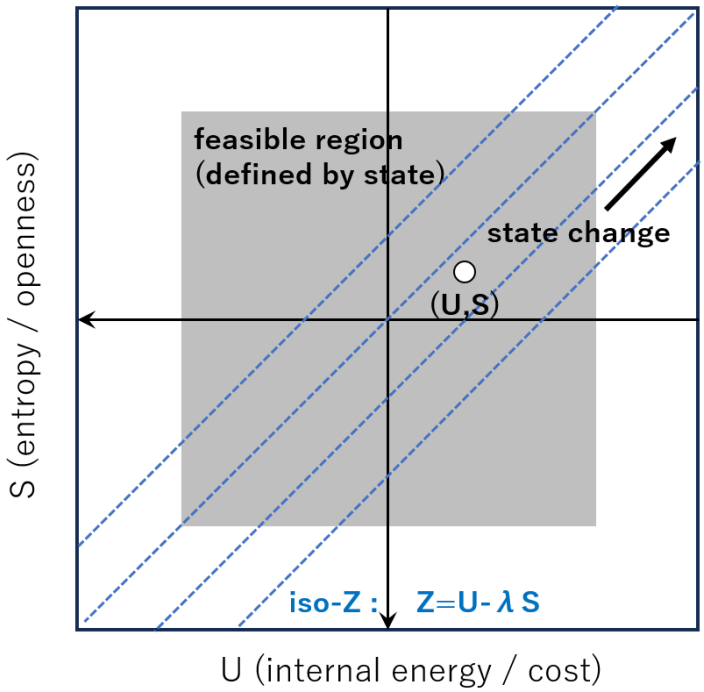


Figure 2: Thermodynamic grounding used as a formal state-space definition. The U–S plane provides an abstract coordinate system on which evaluation and constraints are represented. Iso-Z contours ($Z = U - \lambda S$) define equal-evaluation lines, while the feasible region is restricted by state constraints.

4 Bridging Axioms

This section states minimal assumptions ("bridging axioms") that connect variational quantities in policy-based variational inference to the operational coordinates (U, S) used in this paper. The axioms are *sufficient* for our constructions, not necessary: DBFEP is intended to remain

meaningful even if a reader prefers a different modeling formalism, so long as a mismatch-like scalar and an uncertainty-like scalar can be defined.[7, 8]

4.1 Bridging Axiom A1 (canonical-family embedding; sufficient, not necessary)

Axiom A1. At an appropriate descriptive level, relevant beliefs and predictions can be represented by a canonical family (e.g., exponential-family form) so that log-probabilities and entropy-like terms provide natural scalar summaries.

Role of A1 in this paper. A1 is used only to motivate why a two-axis (U, S) description is not arbitrary: in canonical families, one can often identify a "data-fit / energy-like" component and an "uncertainty / entropy-like" component. We do *not* claim that biological systems literally compute thermodynamic potentials, nor do we require a unique mapping from a specific generative model to (U, S).

4.2 Operational correspondences (what is—and is not—claimed)

Under A1 we treat U as a mismatch/cost proxy and S as a predictive-breadth proxy. The only properties we rely on are: (i) trade-offs between these proxies can be drawn on two axes, (ii) evaluation contours can be expressed as linear rulers of the form $Z = U - \lambda S$, and (iii) feasibility constraints restrict admissible regions in the plane.

4.3 Status of $Z = U - \lambda S$

The quantity Z is introduced as an evaluative coordinate that is convenient for describing switching and selection. Importantly, DBFEP does *not* assume that Z is globally minimized over time. Instead, λ can be controlled or swept to represent changing evaluative stance, and selection is realized by *first-touch contact* between an iso-Z ruler and the feasible set. This framing isolates the paper's central contribution: a constrained geometric selection rule that can generate discontinuous switching, inertia-driven lag, and history dependence without requiring an explicit global optimization narrative.

5 Embedding present-conditioned variational term and future-oriented variational term in the U–S Plane

We now place present-conditioned variational term and future-oriented variational term on the shared U–S state space. The goal is *not* to identify them, but to make their structural differences explicit while enabling comparison on the same plane.

5.1 present-conditioned variational term as a present-conditioned point (or local neighborhood)

present-conditioned variational term evaluates present model–observation consistency and is anchored to the current time (i.e., to current observations and beliefs). For this reason it is represented as a point in the U–S plane (or, if one wishes to acknowledge estimation uncertainty, a small neighborhood around that point). In the sense relevant here, this representation is policy-independent: it does not require committing to a particular future policy to be defined at the present step.

5.2 future-oriented variational term as a policy-indexed bundle of trajectories

future-oriented variational term depends on predicted future outcomes under specific policies. Each policy π (and, if relevant, each horizon τ) induces a distinct predictive distribution and therefore a distinct path through U–S as predictions unfold. Accordingly, we represent future-oriented variational term as a *bundle of trajectories* $(U(\tau,\pi), S(\tau,\pi))$ indexed by policy (and optionally by horizon).

5.3 Structural asymmetry (type difference)

The embedding makes a type difference visible: present-conditioned variational term is present-oriented and point-like, whereas future-oriented variational term is future-oriented and trajectory-like. This asymmetry is often obscured when both are discussed as if they were quantities of the same kind, which can encourage "joint objective minimization" readings. In DBFEP, we treat the point-like versus bundle-like character as primary and build the selection geometry on top of it.

5.4 Sign convention for value-like display (optional)

For geometric clarity in some figures, we may display $-\text{future}-\text{orientedvariationalterm}$ as a value-like quantity. This is only a sign convention for visualization; the policy dependence and trajectory character of future-oriented variational term, as well as the switching logic developed in Sections 6–8, do not depend on this choice.

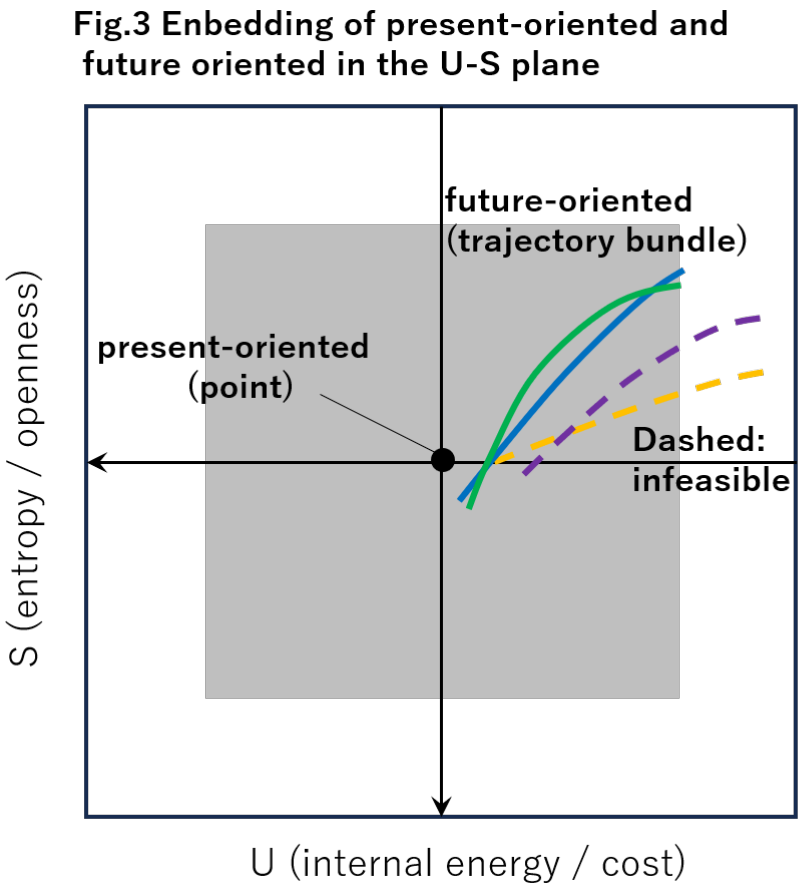


Figure 3: Embedding of present-oriented and future-oriented evaluations in the shared U–S state space. The present-conditioned variational term (present-conditioned variational term) is represented as a point, while the future-oriented variational term (future-oriented variational term) is represented as a bundle of policy–horizon branches. Selection is governed by the first-touch rule: for a fixed attitude λ, an iso-Z evaluative ruler is translated in the improvement direction, and the branch that makes the first feasible contact is selected. Solid branches denote trajectories that admit such first feasible contact and are therefore realized. Dashed branches denote trajectories that, although lying entirely within the feasible region, are never selected because the iso-Z ruler always makes first feasible contact with another branch before reaching them. In particular, vertical position alone does not determine selection: a branch may appear more favorable in absolute terms yet remain unselected if it never achieves first feasible contact under the evaluation geometry.

6 State Constraints and Feasibility

The evaluative geometry introduced above does not determine behavior unless feasibility is taken into account. We represent feasibility through state constraints and show how constraints shape switching and breakdown.

6.1 Feasible regions (admissibility)

The aggregated constraint variable `state` delimits a subset of the U–S plane corresponding to realizable states. Evaluation and switching are meaningful only within this feasible region. Feasibility is distinct from evaluation: a candidate may be evaluatively preferred yet infeasible given current limits.

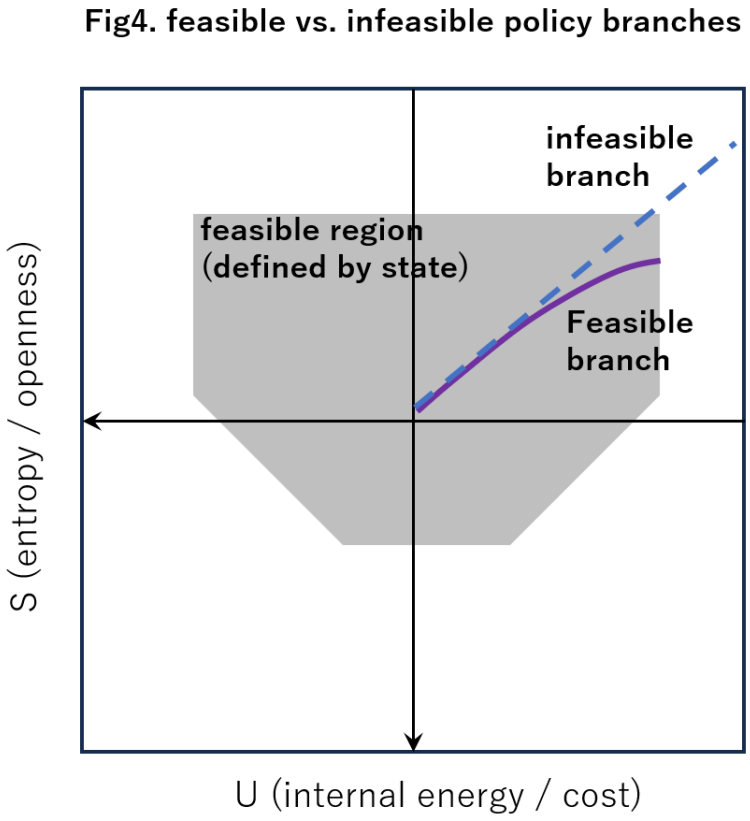


Figure 4: Feasible and infeasible policy branches under state constraints. The feasible region defines admissible states; policy branches that exit this region become infeasible (dashed). Dashed segments may also indicate trajectories that temporarily lie within the feasible region but inevitably violate constraints over time, leading to constraint-induced breakdown as a geometric loss of feasible contact.

6.2 Geometry of breakdown (constraint-induced failure)

Overlay iso-Z rulers and policy branches on the feasible region. As λ changes, iso-Z rulers rotate (Section 7). If the first evaluatively relevant contact between a ruler and a policy branch would occur only in an infeasible region, then the corresponding switch cannot be realized. In this framework, breakdowns (including sudden loss of cyclic operation) are geometric consequences of feasibility constraints rather than failures of inference or evaluation.

In this paper, feasibility refers solely to whether a state is admissible given constraints (resources, safety margins, or limits), and does not imply that the state will be reached, maintained, or selected by the dynamics. Actual behavior is determined by the intersection of feasibility with the switching dynamics defined by the first-touch rule and inertia. Accordingly, dashed segments in subsequent figures indicate trajectories that are not realized under the switching dynamics, even if portions lie within the feasible region.

7 Lambda-Rotation and Iso-Z Geometry

Having established feasibility constraints, we now describe evaluation geometrically on the shared U–S plane. Central to this description is the interpretation of Z through its iso-Z rulers and their rotation under changes in evaluation attitude λ.

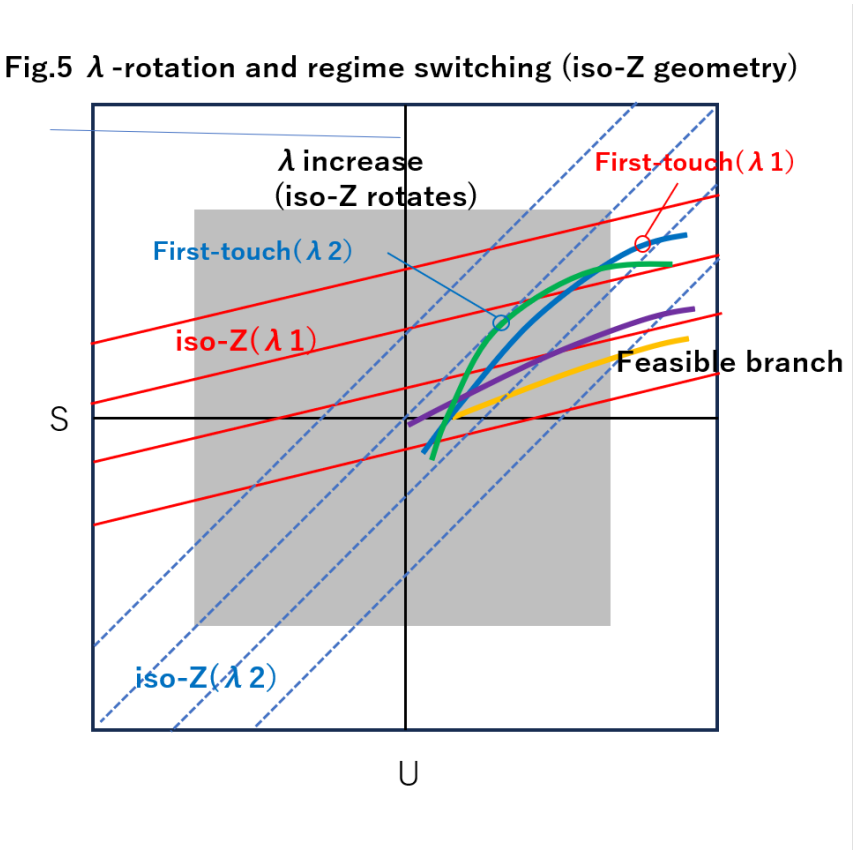


Figure 5: Regime switching via λ-rotation and the first-touch rule. Selection is operationalized by a geometric contact procedure: for a fixed evaluation attitude λ, the family of iso-Z rulers is translated in the improving direction (decreasing U and/or increasing S) until it makes the first feasible contact with the policy branch bundle. As λ varies continuously, rotating the ruler family, the identity of the contact point can jump discontinuously between branches. This explains discrete regime switching as a geometric consequence of the first-touch rule under feasibility constraints, without requiring a global optimization search over the entire plane.

7.1 Iso-Z rulers (ranking, not an objective to be minimized over time)

Iso-Z sets are straight lines in the U–S plane whose slope depends on λ. Rather than representing a quantity that must monotonically decrease, these lines serve as evaluative rulers that rank states relative to one another. Changing λ rotates the ruler family.

7.2 First-touch rule (operational selection under feasibility)

When iso-Z rulers are overlaid on the present-conditioned variational term point and the feasible portions of future-oriented variational term branches, the selected candidate is determined by a *first-touch* procedure: fix λ (thus fix the ruler orientation), translate the ruler in the improving direction, and choose the *first feasible point of contact* with the candidate set. This rule does not require a global search over the entire plane; if the unconstrained optimum would lie outside the feasible region, the first feasible contact determines the selected mode.

7.3 Switching via λ-rotation (continuous attitude, discrete regime change)

As λ varies continuously, the identity of the first-touch branch may change discontinuously. This discontinuity corresponds to regime switching: smooth changes in evaluation attitude can induce abrupt policy changes.

8 Inertia and Hysteresis

We now describe why smooth changes in evaluation attitude can produce delayed responses and history dependence. In DBFEP these effects are modeled as *inertia* in the effective attitude variable and as *hysteresis* in regime switching under feasibility constraints.

8.1 Attitude inertia (effective λ lags the commanded $\lambda(t)$)

We distinguish the commanded attitude $\lambda(t)$ (what the controller "wants") from an effective attitude $\lambda_{\text{eff}}(t)$ (what is realized in evaluation and switching). Finite adjustment rates, friction-like costs, and internal consolidation processes can make $\lambda_{\text{eff}}(t)$ follow $\lambda(t)$ only with delay. On the U–S plane this delay appears as a phase lag between the rotation of iso-Z rulers implied by $\lambda(t)$ and the actual switching events implied by $\lambda_{\text{eff}}(t)$.

8.2 History dependence and hysteresis (switching thresholds depend on sweep direction)

Because selection is implemented by a first-touch rule under feasibility, the identity of the selected branch can depend on the direction in which λ is swept. When λ increases versus decreases, the first feasible contact may occur on different branches even at the same nominal λ value. This produces hysteresis loops: switching thresholds differ for upward and downward sweeps, and the current regime retains memory of the past path of λ.

8.3 Projection caveat (what the panels show and what they hide)

The visual signatures used in later figures (e.g., phase-lag in a time–attitude projection and hysteresis in an attitude–regime projection) are *projections* of a higher-dimensional process. Each panel isolates one relation and necessarily suppresses others. Accordingly, we interpret these panels as complementary diagnostics for inertia and hysteresis rather than as complete representations of the underlying entanglement among feasibility, evaluation, and switching.

8.4 Implication for interpretation

Inertia and hysteresis are not "errors" of evaluation; they are expected dynamical consequences of finite adjustment rates and constraint-limited selection. This framing helps avoid over-reading switching as evidence of inconsistent objectives or pathological inference: the same selection rule can yield delayed and history-dependent behavior even when the evaluative geometry is fixed.

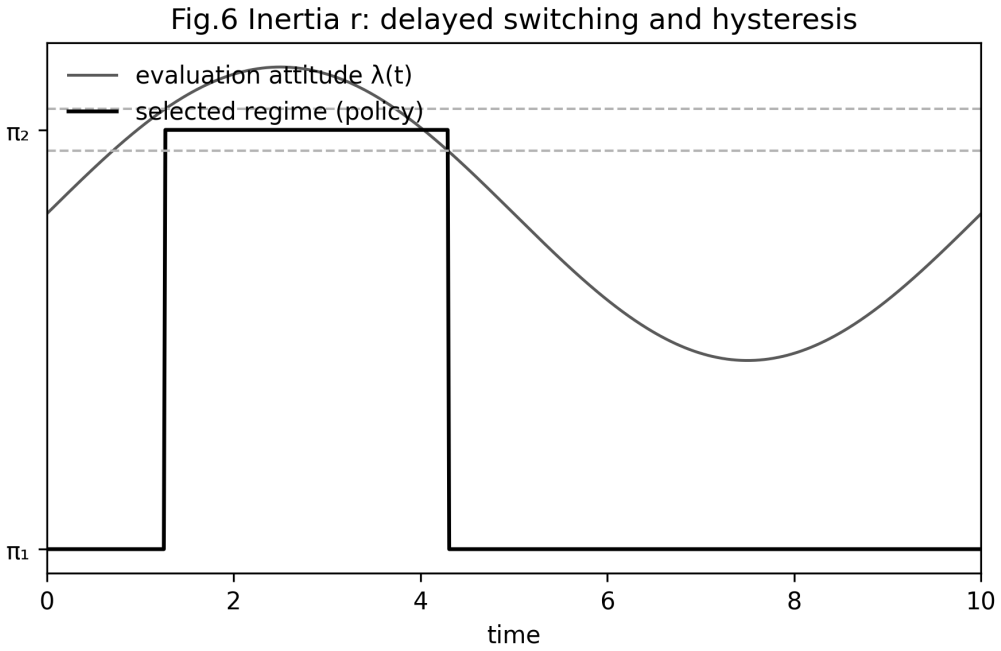


Figure 6: Inertia r produces delayed switching and hysteresis. Under time-varying evaluation attitude $\lambda(t)$, the selected regime switches at different thresholds depending on direction, yielding history dependence and non-instantaneous transitions.

9 Casterboard-like Mechanical Instantiation

This section provides a concrete *mechanical instantiation* of the DBFEP ingredients.Up to this point, we have formalized the evaluative logic of DBFEP as a geometric operation on the abstract $U - S$ plane. However, to demonstrate that DBFEP constitutes a "transferable dynamical motif" rather than a mere interpretive framework, it is necessary to show how this geometric selection rule translates into concrete, time-evolving behavior.

To bridge the gap between abstract evaluation and physical realization, this section maps the functional roles defined in the $U - S$ plane onto a three-dimensional $X - Y - Z$ coordinate system representing the dynamics of a mechanical system. Here, the x-axis represents the balance of goals (future-oriented variational term-side purpose), the z-axis represents the mode of engagement (present-conditioned variational term-side means), and the y-axis represents the resulting net progression (drift) emerging from their dynamic coupling. This shift in coordinates is not a replacement for the $U - S$ geometry, but a necessary projection intended to provide an "existence proof" that the proposed motif is physically implementable and yields observable signatures such as tacking-like progression.

The goal is *not* to claim that the abstract U–S variables are mechanically identical to a particular device coordinate system. Rather, we use a casterboard-like mechanism to demonstrate that the qualitative signatures emphasized in this paper can arise from a simple coupled dynamics with constraints.

fig.7 Caster–like mechanical instantiation side view (schematic)

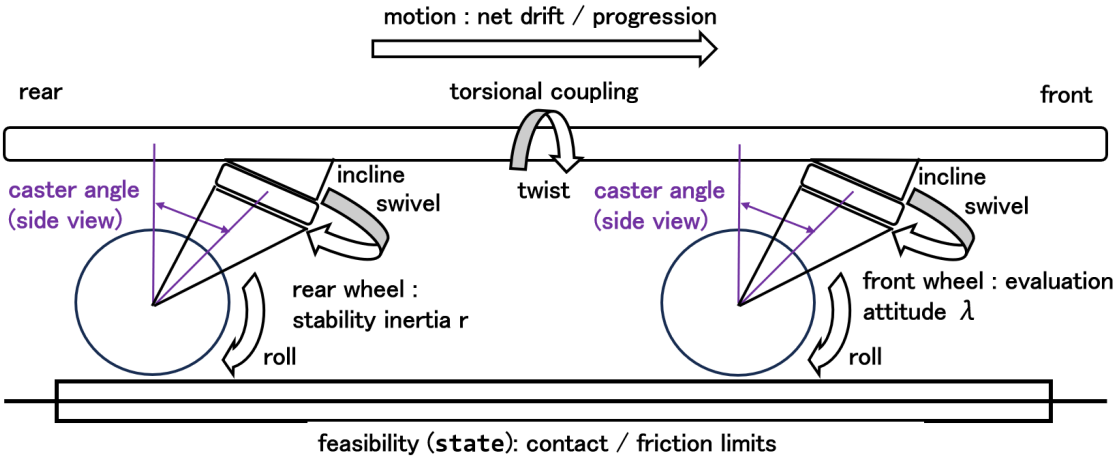


Figure 7: Caster-like mechanical instantiation (schematic; intentionally static). This figure fixes a *role correspondence* between abstract ingredients (λ, inertia, feasibility/`state`) and mechanical components (swivel/roll coupling and contact constraints). It is an instantiation of qualitative roles, *not* an equation-level identification of the U–S variables with mechanical coordinates.

9.1 What is being instantiated

The instantiation targets *roles*: (i) an attitude-like control parameter that changes how evaluation is performed (represented by λ), (ii) an inertia-like lag in the realized attitude (represented by r in the abstract model, and by finite adjustment in the mechanical system), and (iii) feasibility restrictions (represented by `state`). These roles are realized by swivel/roll coupling and contact constraints, but we do not identify the mechanical coordinates with the U–S plane.

9.2 Figure 7 as a role-level wiring diagram

Figure 7 is intentionally static. It should be read as a role-level wiring diagram (what implements λ, inertia, and feasibility), not as a claim that the pictured mechanical coordinates are identical to the abstract U–S variables.

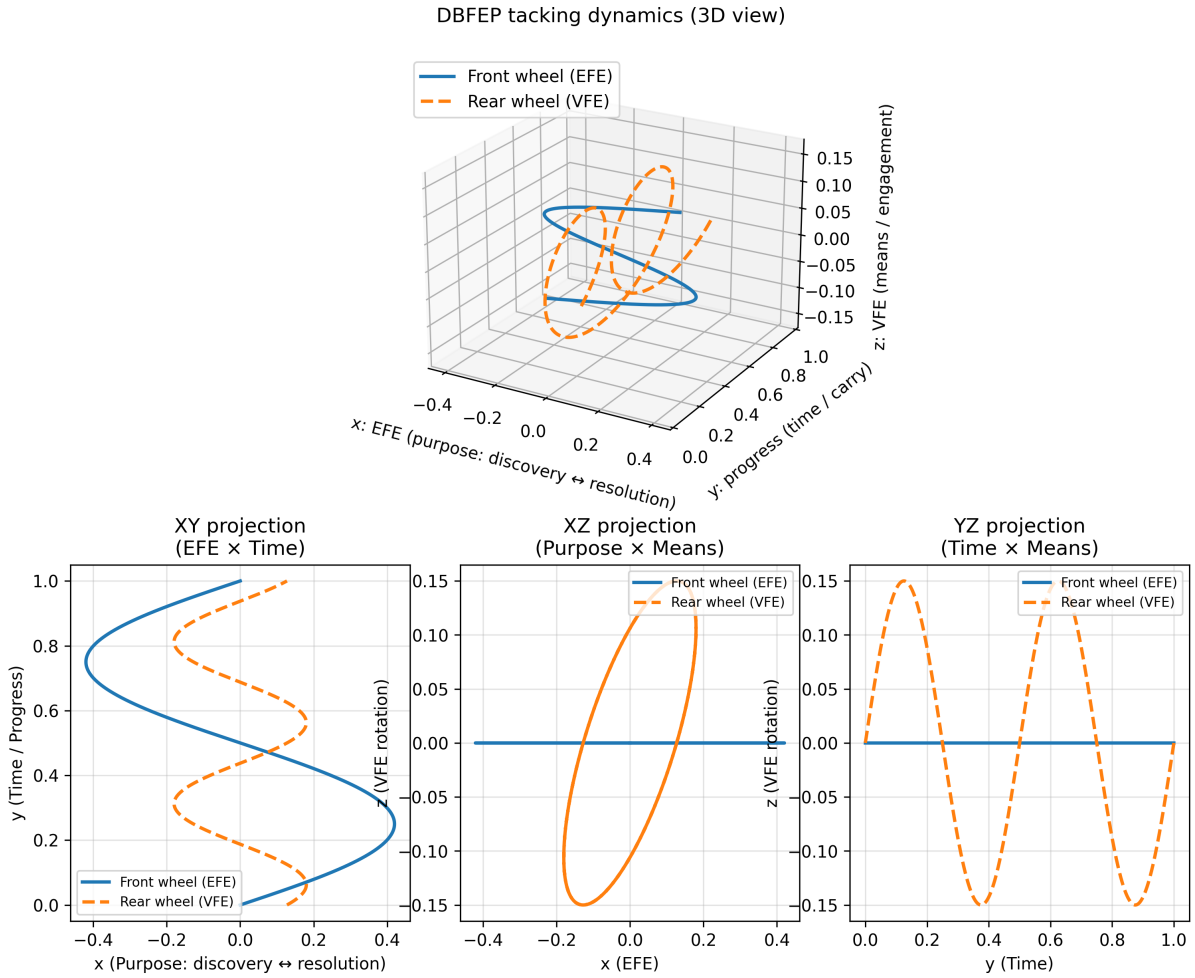


Figure 8: Tacking dynamics (schematic existence proof). (a) Phase-shifted bounded cycling in two orthogonal modes can yield net progression (drift). (b–d) The panels are *projections* of a higher-dimensional process: they isolate complementary signatures (alternation, phase-lag/loop geometry, and hysteresis under constraints) and therefore do not display the full entanglement of evaluation, feasibility, and switching.

9.3 Tacking as an existence proof (Figure 8)

Figure 8 provides a schematic "existence proof": a bounded coupled oscillation can yield net progression (drift) when orthogonal modes are phase-shifted, without any assumption of monotonic descent of a global objective. [9] Accordingly, the empirical content is the presence of qualitative signatures (phase-shifted bounded cycling and resulting drift), not a fit of device-specific parameters.

9.4 How to read the projections (Figure 8)

The panels in Figure 8 are *projections* of a higher-dimensional process. Each projection isolates a single relation that can be checked visually while suppressing others. We therefore interpret the projections as complementary diagnostics: (a) alternation of engagement between modes, (b) phase-lag/loop geometry induced by inertia-like delay, and (c) history dependence (hysteresis) in regime switching under constraints. No single panel is a complete representation of the underlying entanglement among feasibility, evaluation, and switching.

9.5 Bridge to the U–S selection rule (Figure 9)

Figure 9 provides the bridge between abstract evaluation geometry and a mechanical realization in a minimal, operational form. The bridge is a preservation of *switching logic* (rotation of evaluative rulers → first feasible contact → regime change) across representations, not an identification of coordinates between the U–S plane (left) and the physical track (right).

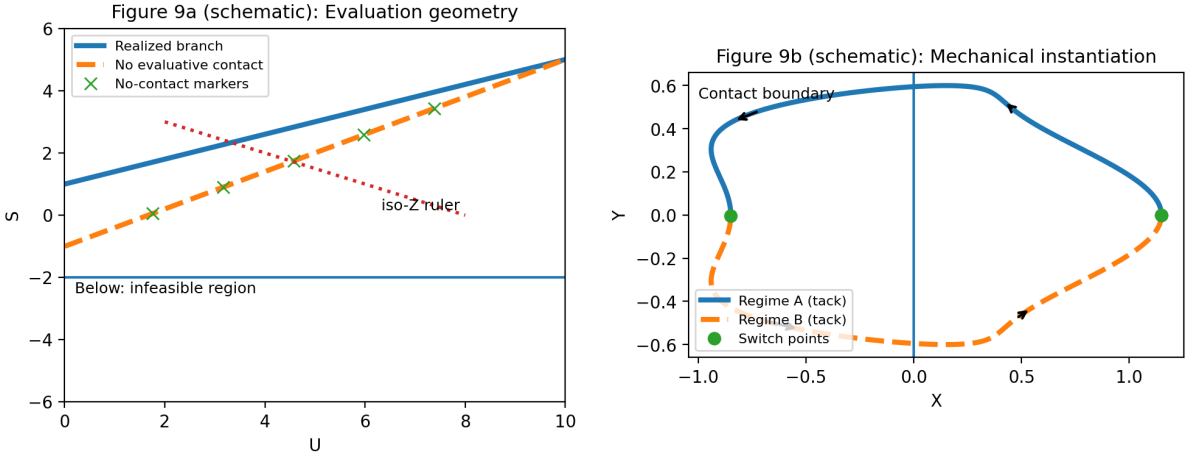
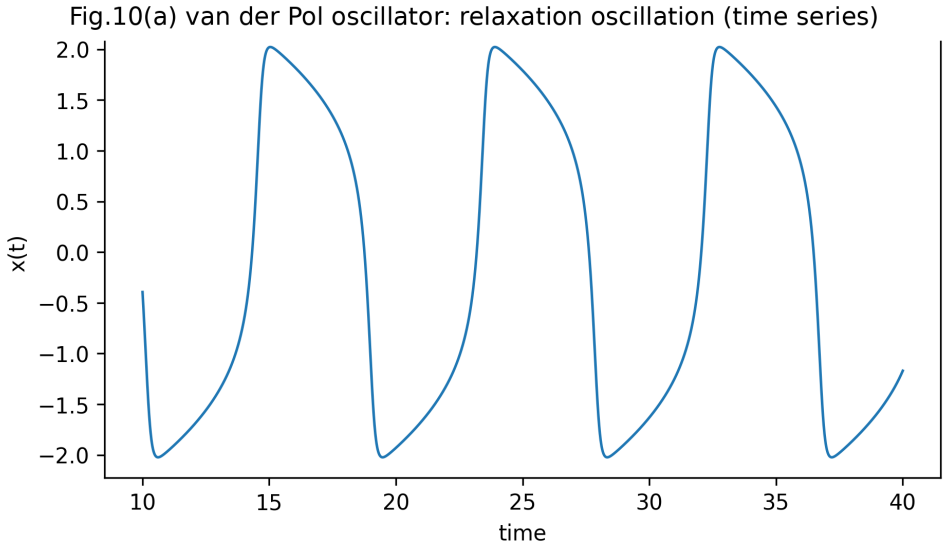


Figure 9: Role-preserving bridge between abstract evaluation geometry and mechanical instantiation. Left (evaluation geometry): Policy–horizon branches are embedded in the shared U–S space together with an example iso-Z evaluative ruler. The solid branch denotes a trajectory that admits a first feasible contact with the ruler and is therefore realized under the switching dynamics. The lighter branch, marked with $\times$, denotes regions where the iso-Z ruler cannot make feasible contact. These regions indicate loss of evaluative contact and predict constraint-induced breakdown, even though the geometry is drawn continuously. The shaded area indicates states that are infeasible due to explicit constraints. Right (mechanical instantiation): The same rotation–contact–switching logic is realized as a tacking-like physical trajectory in ordinary spatial coordinates. Alternating motion regimes (solid and dashed) and discrete switching points (circles) illustrate how continuous changes in attitude λ and inertia r can induce discrete regime changes and hysteresis in a concrete dynamical system. No coordinate-wise identification between the two panels is assumed; the mapping preserves selection roles rather than variables.

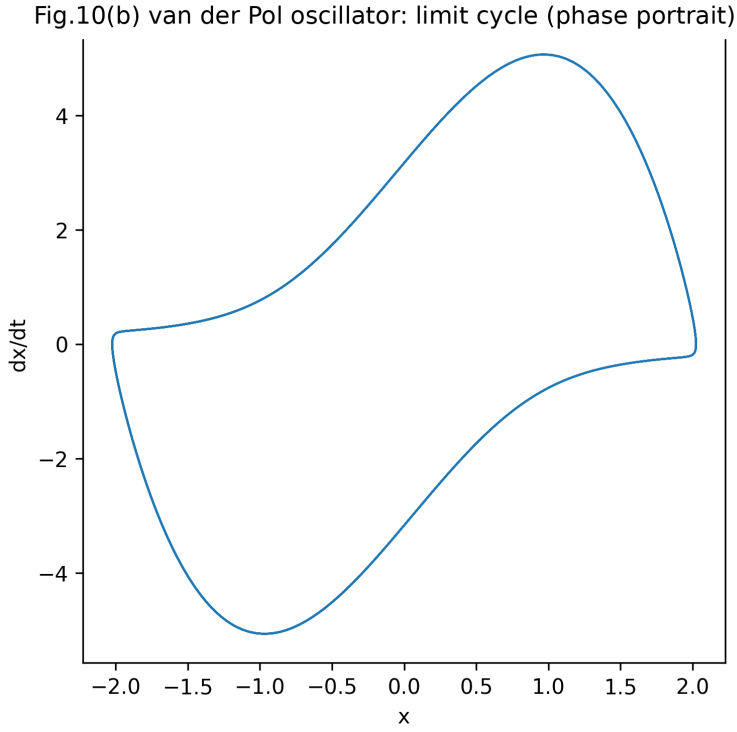
van der Pol oscillator as a phenomenological reference. To make the notion of bounded non-equilibrium cycling concrete, we include the standard van der Pol oscillator,

$$\ddot{x} - \mu(1 - x^2)\dot{x} + x = 0, \tag{3}$$

which exhibits a stable limit cycle for $\mu > 0$.[10, 11] Here it serves only as a compact *phenomenological* reference for bounded cycling (sustained periodic motion under nonlinear dynamics). We do *not* claim mechanistic identity between the oscillator and the U–S embedding. Instead, the oscillator provides a vocabulary for waveform features that are often measurable in behavior (e.g., dwell times and abruptness of transitions).[12] Heuristically, low μ corresponds to smoother oscillations, while high μ yields relaxation-like cycles with long dwell times and sharp transitions. Distinguishing inertia-driven versus feasibility-driven switching failure should be done by the intervention logic in Section 11, rather than by parameter fitting alone.



(a) Time series $x(t)$ (relaxation oscillation).



(b) Phase portrait $(x, \dot{x})$ (limit cycle).

Figure 10: van der Pol oscillator as a canonical example of bounded non-equilibrium cycling. For $\mu > 0$ the system converges to a stable limit cycle. (a) Time series $x(t)$ showing relaxation oscillations. (b) Phase portrait $(x, \dot{x})$ showing the limit cycle. The figure is included as a minimal dynamical reference for bounded cycling; no parameter fitting or identification with the U–S coordinates is intended. [10, 11]

10 Projections and Psychological Readings

Having demonstrated a concrete mechanical instantiation in the previous section, it is crucial to clarify the relationship between the underlying structural motif and its various observational interpretations. In DBFEP, we treat the multiplicity of descriptive levels—ranging from normative evaluation to psychological and neurobiological narratives—explicitly in terms of *projections*. The necessity of this projection-based approach arises from the fact that the underlying dynamics live in a higher-dimensional space coupling evaluation, constraints, and internal states. Any single behavioral readout or psychological category is inherently an information-losing slice of this fuller process. By formalizing these as projections, we provide a methodological safeguard against category errors, ensuring that divergent psychological descriptions (e.g., exploration vs. exploitation, or perseveration) are understood as complementary diagnostics of the same unified constrained switching dynamics rather than as distinct, competing mechanisms.

This section explains how to read the paper's figures and why no single panel should be interpreted as "the mechanism." The underlying dynamics live in a higher-dimensional space (evaluation, constraints, and internal adjustment states). Any two-dimensional plot is therefore a *projection*: it makes one consequence legible while necessarily hiding other degrees of freedom.

10.1 Projection-invariant claim versus projection-dependent signature

The projection-invariant claim of DBFEP is the selection logic: at each moment, an iso-Z ruler ($Z = U - \lambda S$) meets a feasible set and selection occurs at the first feasible contact. The projection-dependent content of a figure is which signature of this logic becomes visible after marginalizing over hidden variables. Agreement with a single projected signature is therefore suggestive but not decisive; the intended evidential standard is a pattern family across complementary projections.

10.2 How to read Figure 11

As shown in Fig. 11(a), projection onto the XY plane can mask discrete switching as an apparently smooth trade-off. By contrast, Fig. 11(b,c) reveal delayed responses and hysteresis that become visible under alternative projections.

- **Panel (a): time vs. regime (choice sequence).** Discrete switching between regimes is visible. This is compatible with both constraint-driven switching (loss/gain of feasible contact) and inertia-driven delay; by itself it does not identify which hidden variable changed.
- **Panel (b): time vs. attitude (delay / inertia signature).** A realized stance can lag a commanded stance $\lambda(t)$ when adjustment has inertia. Similar lag-like traces can also be induced by hidden feasibility changes or other latent states; the panel motivates, but does not by itself prove, an inertia interpretation.
- **Panel (c): attitude vs. regime (hysteresis signature).** Increasing and decreasing sweeps of λ yield different switching thresholds. The loop width is informative, but this projection suppresses the constraint geometry; it cannot, by itself, identify which feasibility boundaries are responsible.

10.3 Clinical projection examples (hypotheses, not identifications)

The U–S geometry can be used as a compact language for formulating clinical hypotheses. The following are *projections* (illustrative mappings), not mechanistic identifications; clinical validation would require dedicated measurements and interventions.

Inertia and perseveration (Parkinsonian set-shifting as an example). Difficulty in set-shifting and perseverative dwell in a prior action regime can be projected as increased effective inertia: the realized attitude λ_{eff} lags the commanded signal $\lambda(t)$, w2dening hysteresis loops and prolonging dwell times. This is a mapping hypothesis: it should be assessed by interventions that separate attitude sweeps from feasibility covariation.

Feasibility contraction and fatigue-like collapse. Functional collapse under fatigue or resource depletion can be projected as a contraction of the feasible region: the iso-Z ruler may fail to make contact with any admissible branch ("loss of contact"), yielding abrupt breakdown that is a geometric consequence of constraints rather than an optimization failure.

How to make these mappings discriminative. DBFEP recommends intervention-based discrimination rather than relying on a single projection: the decisive test is whether hysteresis and lag persist when feasibility is held approximately constant (supporting inertia) versus whether they disappear when feasibility is stabilized (supporting feasibility covariation).[4, 8]

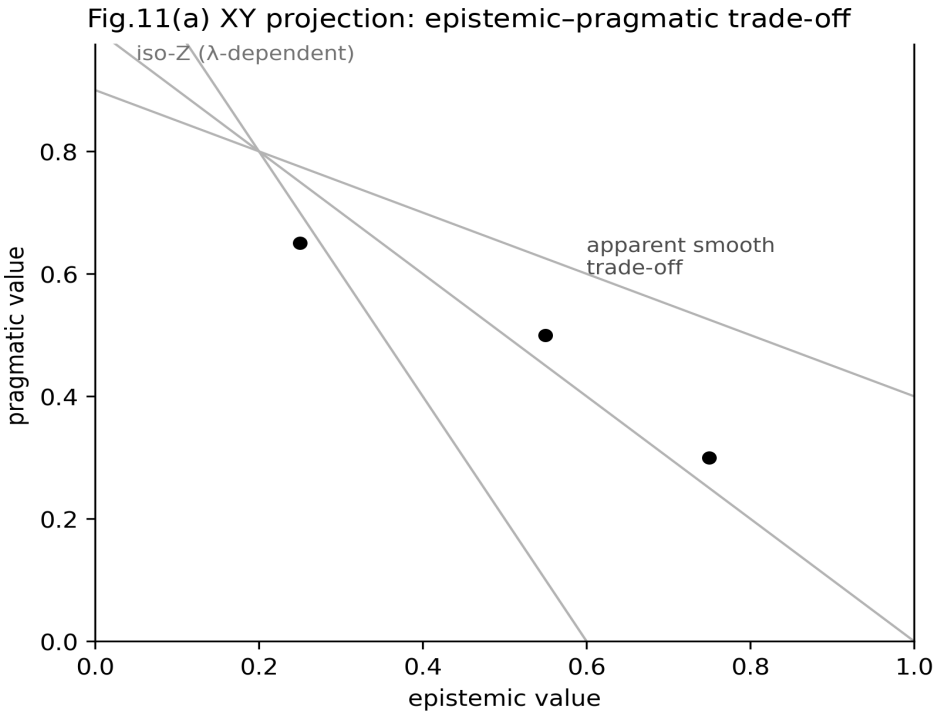
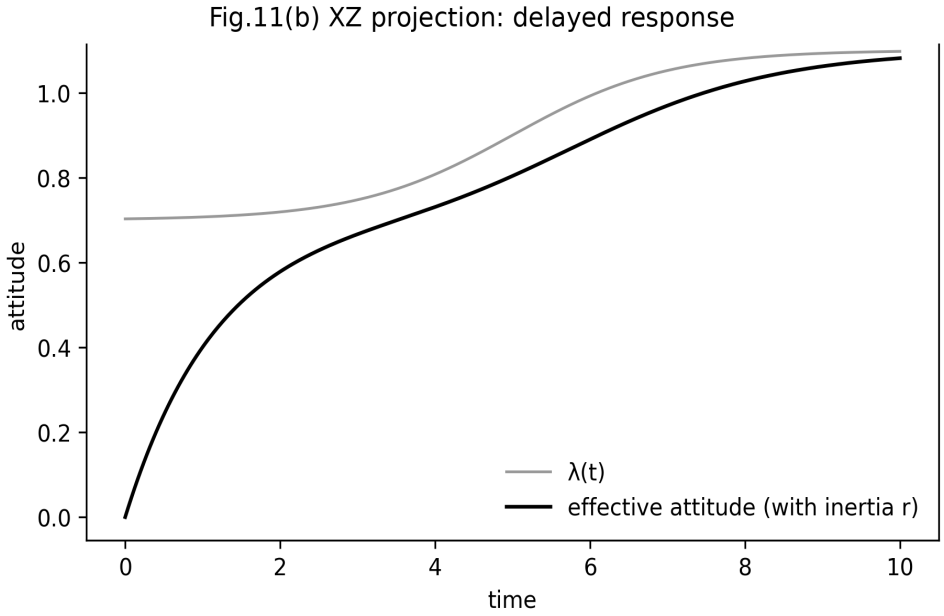


Figure 11: Projection-dependent readings as diagnostic windows. (a) XY projection: Illustrates how discrete switching among candidate policies can appear as a smooth epistemic–pragmatic trade-off.



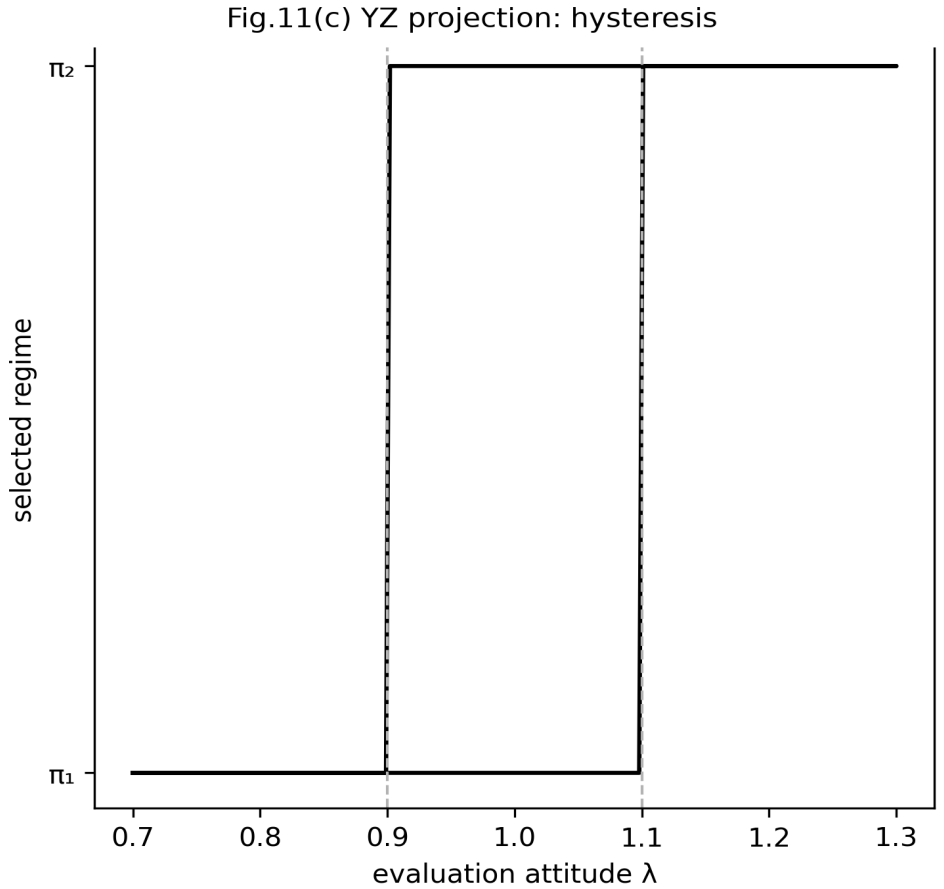


Figure 11: (b) XZ projection: Captures the phase lag between commanded and effective attitude, serving as a signature of inertia r. (c) YZ projection: Reveals hysteresis loops, where switching thresholds depend on the direction of attitude sweep.

11 Testable Predictions and Falsifiability

DBFEP is designed to be falsifiable because it makes comparative predictions: if a pattern is driven primarily by feasibility (constraints) versus primarily by inertia (attitude adjustment), the joint signatures across projections should change in different ways. Below we state testable hypotheses that do not require committing to a particular neural implementation.[8, 13]

Table 1: Illustrative candidate projections (non-exclusive). These are not one-to-one anatomical mappings.

DBFEP quantity	**Candidate operationalization (examples)**
Attitude λ (in $Z = U - \lambda S$)	State-dependent reweighting between *cost* (U) and *uncertainty/openness* (S), operationalized via changes in gain/arousal, broad neuromodulatory tone, or task-dependent valuation biases (e.g., manipulations that shift effort sensitivity or exploration tendency).
Inertia r	Coarse-grained multi-timescale dynamics producing lag and history dependence, operationalized via measurable delays in switching (behavioral lags), slow state variables, and/or interoceptive–autonomic feedback timescales (e.g., the response of autonomic markers to changing volatility or constraint load).
Constraint state / feasible region	Manipulations and readouts of feasibility constraints (fatigue, time pressure, cognitive load, motor/energetic limits), indexed by capacity measures and optional autonomic proxies (e.g., heart-rate variability, skin conductance, pupil-linked arousal) as *candidate* correlates of constraint proximity.

11.1 Predictions tied to selection geometry (high-confidence)

1. **Feasibility shifts switching.** Manipulating feasibility (e.g., resource limits, task difficulty, or actuator bounds) should shift the switching threshold even if the commanded attitude $\lambda(t)$ is held fixed.

2. **Constraint-induced breakdown (loss of contact).** Abrupt failure of cyclic operation should align with violation of feasibility boundaries (no admissible contact), rather than with a smooth drift of evaluative quantities alone.

3. **Hysteresis under attitude sweeps.** Increasing versus decreasing sweeps of λ should yield distinct switching points whenever first-touch contact operates under constraints.

11.2 Predictions tied to inertia and hysteresis (medium-confidence)

1. **Phase lag.** The realized/estimated attitude λ_{eff} inferred from behavior should lag behind the commanded $\lambda(t)$ when adjustment rates are finite.

2. **Rate dependence.** Faster sweeps of $\lambda(t)$ should increase lag and widen hysteresis loops; sufficiently slow sweeps should shrink them toward a quasi-static limit.

11.3 Minimal falsification strategy (decoupling intervention)

The key falsification move is the decoupling intervention: vary $\lambda(t)$ while holding feasibility approximately constant (or vice versa). If hysteresis and lag persist when feasibility is controlled, this supports an internal-inertia interpretation. If the signatures vanish when feasibility is

controlled, the observed patterns were likely driven by hidden feasibility co-variation. When feasibility is expected to co-vary with physiological state, autonomic measures are plausible candidates for operational proxies of feasibility-relevant changes, but they should be treated as measurement hypotheses rather than as defining variables of the theory.

12 Relation to free-energy framework and Active Inference

DBFEP is intended as a geometric and operational complement to policy-based variational inference, not a replacement. Its purpose is to make three distinctions explicit and empirically usable: present-conditioned variational term as present-conditioned (point-like), future-oriented variational term as policy-indexed (trajectory-like), and feasibility as a separate constraint layer that can dominate switching.[7, 8, 13]

12.1 Compatibility

DBFEP does not modify the formal definitions of present-conditioned variational term or future-oriented variational term. It can be read as a representation choice: we embed present and future evaluation into a shared plane and study selection under rotating rulers and feasibility constraints. Readers who prefer alternative derivations can treat U, S, and λ as operational coordinates that summarize the relevant trade-offs.

12.2 What DBFEP adds

- **Type distinction:** point versus bundle-of-trajectories, which reduces "single objective" misreadings.
- **Selection rule:** an explicit first-feasible-contact procedure in place of implicit optimization narratives.
- **Constraint layer:** breakdown and clipping as geometric consequences of feasibility, not anomalies of inference.

12.3 What DBFEP does not claim

DBFEP does not claim thermodynamic identity of its coordinates, nor that λ is a unique internal scalar. The claims are at the level of signatures and interventions: if an agent exhibits regime switching under constraint-limited evaluation, the projected patterns described here should be observable and manipulable.

13 Limitations

Several limitations matter for correct interpretation and for future work.

13.1 Non-identifiability from projections

Because the signatures are defined through projections, different hidden-variable mechanisms can yield similar panels. This motivates complementary projections and intervention-based tests (Section 11). A single panel should not be treated as a unique fingerprint of the mechanism.

13.2 Proxy dependence

The U–S plane is an operational coordinate choice. The mapping from raw data to proxies for "cost" and "uncertainty" is application dependent, and empirical work must document proxy construction and robustness to alternative definitions.

13.3 Multiplicity of constraints and attitudes

Real systems may involve multiple constraint dimensions and multiple attitude-like variables. DBFEP collapses these to a single feasibility variable and a single attitude scalar to make the geometry explicit. Extending the framework to multi-dimensional feasibility and multi-parameter attitude is straightforward conceptually but may require richer data and stronger modeling assumptions.

13.4 Scope of psychological readings

The psychological interpretations are hypotheses meant to organize measurement and intervention. They should be evaluated against alternatives (e.g., reward learning, habit formation, control-theoretic hysteresis) using discriminative experiments rather than treated as settled identities.

14 Conclusion

We reformulated the interaction between present-oriented evaluation (present-conditioned variational term-like quantities) and future-oriented evaluation (future-oriented variational term-like quantities) as a constrained dynamical structure on a shared state space, rather than as a joint optimization problem. The key claim is geometric: adaptive switching can be described as evaluative rotation of an iso-Z ruler together with first-touch contact on a feasible set. Finite adjustment rates introduce lag (inertia), and contact geometry introduces history dependence (hysteresis).

The core unique merit of the Dynamic Balancing free-energy framework (DBFEP) is its *transferability*—the ability to provide a compressed, operational descriptive language that remains invariant across different physical substrates. A prime, intuitive example of this portability is the casterboard-like mechanical instantiation presented in Section 9. By showing that the abstract logic of "evaluative rotation and first-feasible-contact" can be realized in a physical entity with mass, friction, and mechanical constraints, we provide an existence proof that DBFEP bridges the gap between high-level policy-based variational inference formalisms and concrete physical implementation.

Methodologically, this paper emphasizes that the most visible patterns in behavioral data are *projections*: any given observational panel reveals only a partial slice of a higher-dimensional process. By treating psychological and biological descriptions as complementary diagnostics of this unified motif, we offer a rigorous methodology for cross-level model comparisons in biological cybernetics, anchored in physically realizable dynamics. Within this scope, DBFEP provides a portable descriptive language that enhances the interpretability of present-conditioned variational term-future-oriented variational term coupling while keeping thermodynamic interpretations and strong neurobiological claims explicitly out of scope unless supported by additional measurements.

Funding

This research did not receive any specific funding.

References

[1] Toshio Sato. Dynamic balancing free energy in non-equilibrium: Dual j–f mapping, diagonal policy, and s-drift. *Research Square (Preprint)*, 2025. Initial theoretical formulation and mathematical grounding.

[2] Toshio Sato. Dbfep: An extension of free energy balancing stabilization and exploration (u–s cycle and phase diagram validation). Technical report, Zenodo, 2025.

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

[4] Karl Friston, Thomas FitzGerald, Francesco Rigoli, Philipp Schwartenbeck, John O'Doherty, and Giovanni Pezzulo. Active inference: A process theory. *Neural Computation*, 29(1):1–49, 2017.

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

[6] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley, 2 edition, 2006.

[7] Lancelot Da Costa, Thomas Parr, Naveed Sajid, Sasa Veselic, Vasilisa Neacsu, and Karl Friston. Active inference on discrete state-spaces: A synthesis. *Journal of Mathematical Psychology*, 99:102447, 2020.

[8] Thomas Parr and Karl Friston. The anatomy of inference: Generative models and brain structure. *Frontiers in Computational Neuroscience*, 12:90, 2018.

[9] Dalton A. R. Sakthivadivel. Regarding flows under the free energy principle: A comment on "how particular is the physics of the free energy principle?". *Physics of Life Reviews*, 42:25–28, 2022.

[10] B. van der Pol. On relaxation-oscillations. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):978–992, 1926.

[11] Steven H. Strogatz. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*. Westview Press, 2 edition, 2018.

[12] J. A. Scott Kelso. *Dynamic Patterns: The Self-Organization of Brain and Behavior*. MIT Press, 1995.

[13] Beren Millidge, Alexander Tschantz, and Christopher L. Buckley. Whence the expected free energy? *Neural Computation*, 34(3):447–482, 2022.