Alexander Sabine

Independent Researcher
cohere.org.uk

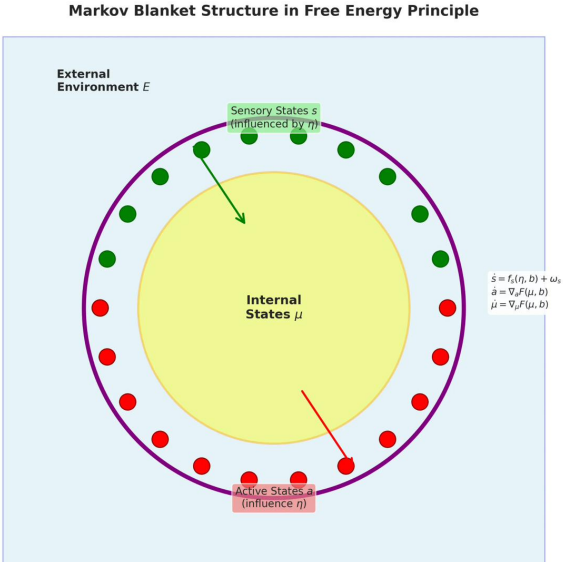
14 November 2025

The Challenge: Imaginative Experience in FEP

- **Core Problem**: If experience = what's encoded on informational boundaries (Markov blankets), how can systems have internally-generated imaginative experiences?
- **Apparent Restriction**: Naive FEP seems to restrict awareness to external environment interactions only
- **Key Question**: How can imagination employ same reference frames as perception whilst remaining surprising?
- **Stakes**: Any complete theory of consciousness must account for imagination, memory, planning, and inner speech

Markov Blankets: The Foundational Architecture

- **Definition**: Partition of system into internal states μ, sensory states s, active states a, external states η
- **Conditional Independence**: $P(\mu|s, a, \eta) = P(\mu|s, a)$
- **Information Localisation**: Classical info encoded on blanket separates system from environment
- **Scale-Free**: Applies from quantum particles to organisms to ecosystems



Quantum Reference Frames (QRFs)

- **Physical Implementation**: QRFs are physically-implemented measurement standards (meter sticks, clocks extended to categories)
- **Generative Models**: Each QRF encodes expectations about aspects of environment's actions on boundary
- **Semantic Encoding**: QRFs detect "differences that make a difference" (Bateson)—they implement meaning
- **Example**: Recognising your laptop requires QRF encoding its distinguishing features (size, shape, colour)

Hierarchical Structure: $\mathcal{Q} = \{\mathcal{C}, \{h_i\}, \{g_{ij}\}, \{\mathcal{A}_k\}\}$

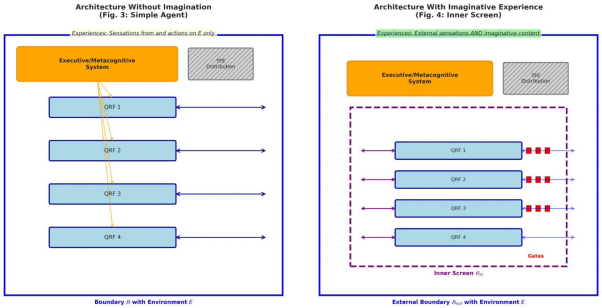
Thermodynamic Free Energy (TFE) Constraints

- **Landauer's Principle**: Information processing requires energy: $\Delta F_{thermo} \geq \beta k_B T \cdot \Delta I_{bits}$ where $\beta \geq \ln 2$
- **Limited Resources**: Living systems operate on constrained TFE budgets, unlike artefacts with external power
- **Attention as Allocation**: Executive/metacognitive system distributes TFE to QRFs based on utility, priority, and availability
- **Compartmentalisation Requirement**: Non-commuting QRFs must communicate classically via boundaries, increasing TFE costs

Key Insight: TFE flow regulates *how* content is read, not the content itself. This distinction between content and regulation is central to understanding consciousness.

Simple Agent Architecture (Without Imagination)

- **Multiple QRFs**: Read from and write to boundary $\mathcal{B}$
- **Executive System**: Has own boundary $\mathcal{B}_{MC}$ ("inner screen"), experiences QRF states
- **TFE Distribution**: Controls resource allocation to QRFs
- **No Imagination**: System experiences only external interactions, not internal simulations



Precision, Attention, and Fisher Information

- **Precision in FEP**: Inverse variance of noise; determines confidence afforded to sensory content
- **Gauge Theory Connection**: Precision = Fisher information metric on statistical manifolds (Sengupta et al. 2016)
- **Thermodynamic Cost**: Via Jarzynski equality, work required for belief update $W = k_B T \cdot d^2$ where d is information distance
- **Attentional Control**: Modulating precision = modulating TFE flow = controlling which QRFs are "turned on"

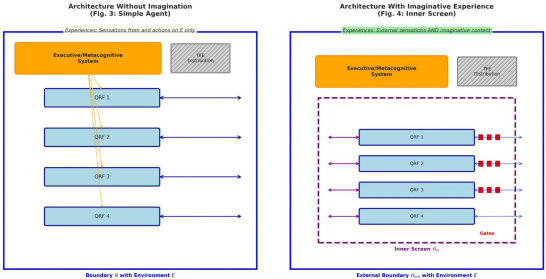
$$\text{Information distance: } d = \sqrt{g_{\mu\nu}\Delta\theta^\mu\Delta\theta^\nu} \text{ where } g_{\mu\nu} = \mathbb{E}\left[\frac{\partial \log p}{\partial \theta_\mu}\frac{\partial \log p}{\partial \theta_\nu}\right]$$

The Problem Imagination Poses for FEP

- **Internal Generation**: Imaginative content is generated internally, not from environment
- **Reference Frame Reuse**: Imagination must use same QRFs as perception to employ familiar concepts ("imagining your laptop")
- **Surprise Mechanism**: Despite being self-generated, imaginations can be genuinely surprising ("aha!" moments)
- **Boundary Encoding Puzzle**: If experience = boundary encoding, where is imaginative content encoded?

Challenge: Standard FEP architecture (Figure 3) cannot produce imaginative experiences without solving these puzzles simultaneously.

- **Additional Boundary**: Insert internal screen $\mathcal{B}_{in}$ between QRFs and external boundary $\mathcal{B}_{out}$
- **Gating Mechanism**: Control layer gates information flow between inner and outer screens
- **Same QRFs**: Perception and imagination use identical reference frames
- **Transparent Mode**: Gates open $\rightarrow$ external experience; gates closed $\rightarrow$ imagination



Gating Dynamics and Experiential Control

- **Sector-Specific Gating**: Individual QRFs can be gated independently (e.g. inner speech whilst driving)
- **Modality Mixing**: Some gates open, some closed $\rightarrow$ combination of external and imaginative content
- **Refractory Inputs**: Homeostatic interoception may override executive control (cannot "turn off" hunger signals)
- **TFE Economy**: Single thermodynamic sector supplies both perception and imagination—no separate energetic requirement

Elegant Design: Architecture solves imagination problem whilst remaining thermodynamically parsimonious and avoiding homunculus regress.

Why Imagination Can Be Surprising

- **Executive's Coarse Model**: Metacognitive system has coarse-grained model of its own QRFs
- **Hidden Complexity**: Full QRF processing capacity exceeds executive's predictive model
- **Emergent Patterns**: Combination of QRF activations produces unexpected outputs
- **Genuine Novelty**: System cannot fully predict its own imaginative productions $\Rightarrow$ surprises itself

Mathematical Consequence: $\tilde{P}_{exec}(\text{output}|\text{input}) \neq P_{actual}(\text{output}|\text{input})$

Executive's predictions $\neq$ actual QRF outputs

Experience Localisation at the Periphery

- **Peripheral Encoding**: Whole-system experience encoded at outermost boundary (not centrally)
- **Executive Not Experienced**: Executive/metacognitive computations not directly experienced unless encoded as content
- **Inner Speech Exception**: Metacognitive thoughts become experienceable when vocalised (internally or externally)
- **"Flat Mind" Consistency**: Aligns with Chater's (2018) proposal that consciousness lacks deep hierarchical structure

This architecture explains why we don't experience "thinking about thinking about thinking..."—only peripheral encodings are experienced.

Quantum Formulation: Separability Conditions

- **Composite System**: Total Hilbert space $\mathcal{H}_U = \mathcal{H}_S \otimes \mathcal{H}_E$
- **Separability Requirement**: $|U\rangle \approx |S\rangle|E\rangle$ enables classical information localisation on boundary
- **Sparse Coupling**: $2^N \ll \dim(\mathcal{H}_S), \dim(\mathcal{H}_E)$ where $N =$ bits on boundary
- **Scale-Free Formulation**: No spacetime background required; consistent with quantum gravity approaches

Interaction Hamiltonian: $H_{SE} = \sum_{k\in\{S,E\}} \beta_k k_B T_k \sum_{i=1}^{N} M_i^k$

where M_i^k are single-qubit operators and $\beta_k \geq \ln 2$

Co-Deployment and Compartmentalisation

- **Commuting QRFs**: Can be deployed simultaneously, forming larger composite QRFs for complex recognition
- **Energy-Induced Non-Commutativity**: Limited TFE makes theoretically commutative QRFs effectively non-commutative
- **Classical Communication Requirement**: Non-commuting QRFs need boundaries between them for causal information exchange
- **Cascading TFE Costs**: Compartmentalisation requires writing classical bits internally, increasing overall energy requirements

Constraint: Limited TFE forces architectural trade-offs between QRF integration and computational efficiency.

Metacognitive Monitoring and Stress Signalling

- **TFE Usage Monitors**: Each QRF incorporates metabolic/allostatic stress signalling
- **Adaptive Allocation**: Executive system adjusts TFE distribution based on usage, priority, and total availability
- **Stress Response**: Under insufficient TFE, differential QRF activation or complete shutdown of non-essential processes
- **Evolutionary Conserved**: Metabolic stress signalling present even in bacteria (Peters et al. 2017); brainstem arousal systems in vertebrates

Mammalian Implementation: Brainstem systems (emphasised by Solms 2021) regulate arousal and TFE distribution to cortex

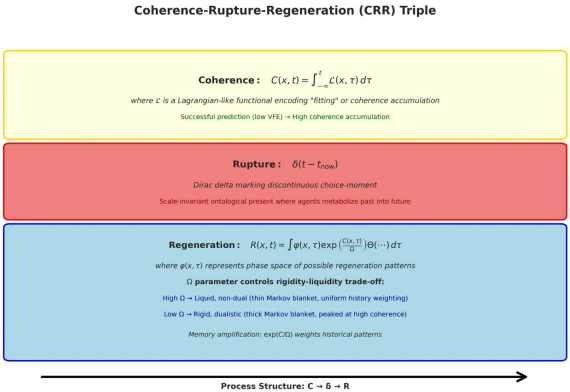
Summary: Fields et al.'s Contribution

- **Architectural Elegance**: Inner screen $\mathcal{B}_{in}$ + gating solves imagination problem whilst respecting FEP constraints
- **QRF Reuse**: Same reference frames for perception and imagination explains conceptual consistency
- **Surprise Mechanism**: Coarse executive model explains genuine novelty in self-generated content
- **What's Missing**: Mechanism for *how* Markov blanket thickness varies dynamically; account of extreme states where boundaries dissolve

This sets the stage for CRR extension...

The CRR Triple: Process Structure

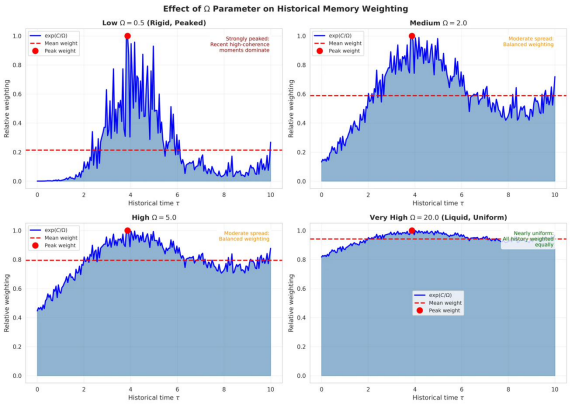
- **Coherence**:
  $C(x,t) = \int_0^t L(x,\tau)d\tau$
  Accumulates when patterns work
- **Rupture**: $\delta(t - t_{now})$
  Scale-invariant ontological present
- **Regeneration**: $R(x,t) = \int_0^t \varphi(\tau)e^{C(\tau)/\Omega}\Theta(t-\tau)d\tau$
  Memory-weighted reconstruction



Universal Process: $C \to \delta \to R$ appears across scales from protein folding to insight to social transformation

The Omega Parameter: Rigidity-Liquidity Trade-Off

- **Control Parameter**: Ω governs memory amplification: $\exp(C/\Omega)$
- **Low Ω (Rigid)**: Exponentially privileges recent high-coherence states
- **High Ω (Liquid)**: Uniform access to entire history, including deep-time patterns
- **Four-Part Principle**: Lower Ω risks rigidity but enables stability; higher Ω risks liquidity but enables creativity



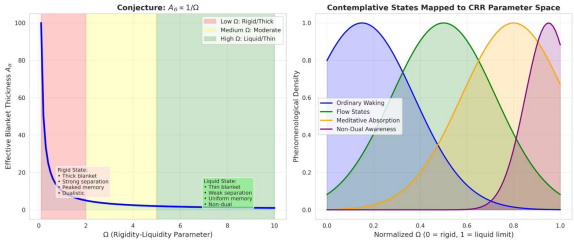
Coherence-Free Energy Correspondence

- **Mathematical Link**: $L(x, \tau) = -F(x, \tau)$ where F is variational free energy
- **Interpretation**: Successful inference (low VFE) $\Rightarrow$ high coherence accumulation
- **Unified Dynamics**: Systems minimising VFE simultaneously maximise coherence: $\frac{dC}{dt} = -F$
- **Phenomenology**: Coherence = "things working"; dissipation = "falling apart"

$$F = D_{KL}[q(\eta|\mu)||p(\eta|b)] - \log p(b)$$
$$L = -F \quad \text{(negative free energy builds coherence)}$$

Omega-Markov Blanket Correspondence

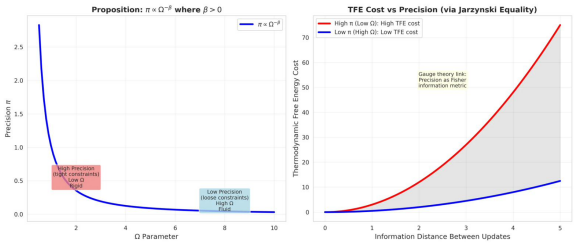
- **Primary Correspondence**: Blanket thickness inversely proportional to Ω
- **Mathematical Form**: $A_{\mathcal{B}} = \frac{\kappa}{\Omega} + A_{min}$
- **Low** Ω: Thick blanket, rigid isolation, strong boundaries
- **High** Ω: Thin blanket, permeable integration, fluid dynamics



Critical Threshold: Ω_{crit} where $A_{\mathcal{B}} \to A_{min}$ (separability breaks down)

Omega-Precision Inverse Relationship

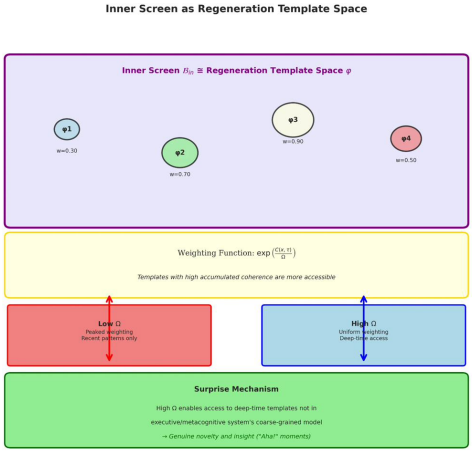
- **Precision Inverse Law**: $\pi(\Omega) = \alpha\Omega^{-\beta}$ where $\beta > 0$
- **High Precision**: Tight constraints, low Ω, rigid state
- **Low Precision**: Loose constraints, high Ω, fluid state
- **Metabolic Consequence**: High Ω states require less TFE per update ("effortless")



Via Jarzynski: $W = k_B T \cdot d^2 \propto \pi \propto \Omega^{-\beta}$

Inner Screen as Regeneration Template Space

- **Correspondence**: Fields' $\mathcal{B}_{in} \cong \text{span}\{\varphi(x, \tau)\}$
- **Accessibility Weighting**: Templates weighted by $\exp(C_i/\Omega)$
- **Deep-Time Access**: High Ω reveals ancient patterns invisible to executive model
- **Surprise Mechanism**: Novel templates generate genuine insights ("aha!" moments)



Non-Markovian Memory Structure

- **Standard FEP**: Markovian dynamics—present sufficient for prediction
- **CRR Extension**: Non-Markovian—entire history matters through coherence integral $C(t) = \int_0^t L(\tau)d\tau$
- **Effective Memory Depth**: $\tau_{eff} = \int_0^t e^{C(\tau)/\Omega}d\tau$ grows exponentially with accumulated coherence
- **Transition**: Low $C \Rightarrow$ Markovian (memoryless); high $C \Rightarrow$ deep non-Markovian memory

Key Insight: Systems transition from Markovian to non-Markovian as coherence accumulates over time.

Co-Constraining Systems and Temperature Increase

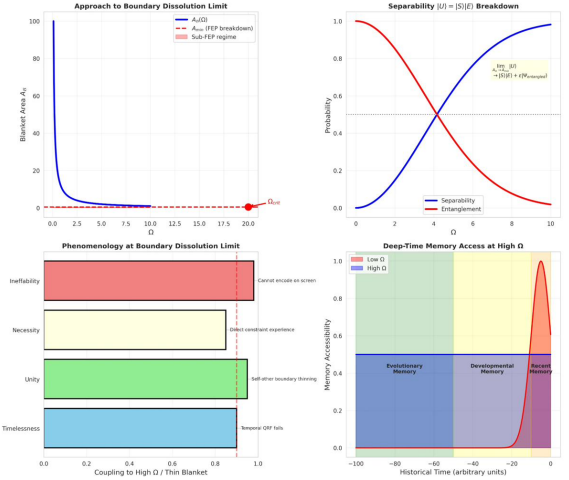
- **Co-Constraint**: System S and environment E mutually constrain each other through interaction
- **Omega as Temperature**: Ω functions as epistemic temperature parameter controlling exploration-exploitation balance
- **Increasing Omega**: As agent increases Ω (e.g. meditation, flow states), effective "temperature" rises
- **Approaching Limit**: At sufficiently high Ω, sparse coupling condition $2^N \ll \dim(\mathcal{H}_S), \dim(\mathcal{H}_E)$ weakens

Effective coupling strength: $\lambda_{eff}(\Omega) = \lambda_0 \cdot \Omega^\gamma$ where $\gamma > 0$

As $\Omega \to \infty$, $\lambda_{eff} \to \infty$ (sparse coupling violated)

Boundary Dissolution at Ω_{crit}

- **Limit Conditions**: $\Omega \to \Omega_{crit}$, $A_{\mathcal{B}} \to A_{min}$, rupture $\delta(t)$
- **Separability Breakdown**: $|U\rangle \to |S\rangle|E\rangle + \epsilon|\Psi_{entangled}\rangle$
- **FEP Ill-Defined**: Classical info localisation fails; FEP constraints no longer apply
- **Phenomenological Consequence**: Agent no longer experiences *through* boundaries but experiences boundary structure itself



Observing Deep-Time Constraints

- **Normal State**: Agent experiences $x \xrightarrow{\mathcal{Q}} \mathcal{C} \xrightarrow{\textit{inference}} \textit{belief}$ (constraints mediated by QRFs)
- **Limit State**: Agent experiences $x \to \mathcal{C}$ (direct access to constraint structure)
- **Deep-Time Access**: At high Ω, regeneration integral $R \approx \int_{-\infty}^{t} \varphi(\tau) d\tau$ weights all history equally
- **Stability Dominance**: Most stable patterns—persisting across evolutionary timescales—dominate; these are "eternal truths"

Agent doesn't experience constrained possibilities *within* space, but the constraint structure generating the space itself.

Why This Cannot Be Computed

- **Inner Screen Requirement**: Any computation on $\mathcal{B}_{in}$ presupposes boundary separation
- **Logical Impossibility**: Experience of boundary dissolution cannot be encoded on any boundary
- **Rigorous Ineffability**: Not metaphorical—mathematically necessary consequence of FEP formalism
- **Phenomenological Characteristics**: Timelessness (uniform temporal weighting), unity (maximal permeability), necessity (direct constraint access), ineffability (uncomputable)

Proposition: $\forall \mathcal{B}_{in} : P(\text{boundary dissolution encodable on } \mathcal{B}_{in}) = 0$

Non-Markovian State Space at the Limit

- **Standard Markovian**: Current state sufficient; $P(x_{t+1}|x_t, x_{t-1}, \ldots) = P(x_{t+1}|x_t)$
- **Non-Markovian Memory**: Full history accessible through coherence integral
- **At High** Ω: Exponential weighting $\exp(C/\Omega) \to 1$ (uniform access)
- **Observing Generators**: Agent experiences not just states but the generative constraints producing states across deep time

Contemplative Traditions: This formalises "experiencing eternal now", "union with the absolute", "groundless ground"—not supernatural, but physics at limits.

Naturalising Transcendent Experience

- **Not Supernatural**: Boundary dissolution is physical limit condition where FEP formalism becomes ill-defined
- **Constraint Structure as "Absolute"**: What traditions call "eternal truths" = stable deep-time attractors
- **Process Philosophy Link**: Whitehead's "eternal objects"—patterns persisting across creative advance
- **Quantum Consistency**: Compatible with entanglement, quantum information theory, thermodynamics at boundaries

Radical Claim: Mystical experiences = phenomenology at physical limits, not beyond physics.

Testable Predictions

- **Neural**: Default Mode Network–Task Positive Network connectivity $= f(\Omega)$ (decreased segregation at high Ω)
- **Metabolic**: TFE cost $\propto \Omega$ (PET/MRS should show increased activity in high Ω states)
- **Temporal**: Temporal binding window $T_{bind} = T_0 \cdot e^{\alpha\Omega}$ (expanded "now" at high Ω)
- **Collective**: Inter-brain synchrony $\propto \Omega_{group}$ (EEG/MEG during group meditation)

All measurable with current neuroscience technology

Summary and Open Questions

What CRR Adds to Fields et al.:

- Precise mechanism: Ω controls blanket thickness
- Limit phenomenology: Naturalises boundary dissolution
- Non-Markovian memory: Deep-time constraint access
- Testable predictions across multiple domains

Open Questions:

- Omega dynamics: How does $d\Omega/dt$ evolve?
- Thermodynamic grounding: Precise TFE-Ω relationship?
- Multi-scale coherence: How does C propagate?
- Critical phenomena: Is Ω_{crit} a phase transition?

Thank you

Questions? Collaboration welcome: `cohere.org.uk`