Beyond Reward, Beyond Information Gain

Armando Vieira

Abstract

A growing line of criticism argues that mainstream AI relies on the wrong primitive: reward. Karl Friston's active inference framework replaces scalar reward maximization with minimization of expected free energy, where epistemic value—expected information gain—plays a central role in curiosity and exploration. The Experiential Coherence Framework (ECF) agrees that reward is too shallow, but argues that even information gain is not yet fundamental enough. ECF proposes that coherence, rather than inference, is the primitive organizational variable. This post explains where Friston's critique aligns with ECF, where ECF goes further, and how both views can be translated into a practical research program for AI.

Introduction

A useful way to pose the question is this: is current AI research *wrong*, or merely *incomplete*?

Friston's answer is that pure reward maximization is too crude. In active inference, agents do not merely optimize scalar return. They select policies by minimizing *expected free energy*, which contains both extrinsic value and epistemic value, the latter corresponding to expected information gain about hidden states [1, 2].

That already marks a serious departure from standard reinforcement learning. In classical RL, exploration is usually secondary: random action noise, entropy bonuses, count-based novelty, or explicit intrinsic rewards. A canonical example is the Intrinsic Curiosity Module, where intrinsic reward is given by forward-model prediction error in a learned feature space [3]. Large-scale follow-up work showed that such approaches can work well, but also exhibit familiar limitations, especially in stochastic environments [4].

ECF agrees with Friston's critique of reward-centric AI, but pushes the critique one level deeper. Its claim is that the problem is not only reward. The deeper problem is that *inference itself is still being treated as primary.* In ECF, coherence—not inference, not reward, not surprise—is the primitive organizational variable.

What Friston Gets Right

Friston's strongest point is that intelligence should not be built on externally shaped reward alone. In active inference, action is guided by expected free energy, which decomposes into epistemic and extrinsic terms. In the process-theory formulation, epistemic value is explicitly treated as expected information gain or mutual information, and is linked directly to curiosity and novelty seeking [1].

This already moves much closer to biological intelligence than standard reward optimization. It also connects naturally to embodiment and homeostatic structure: agents are partly defined by the states they expect, prefer, or must remain within. Friston's later review presents the free energy principle as a normative account of self-organization and optimal Bayesian design, not merely a replacement reward function [2].

So if the comparison is *standard RL versus active inference*, Friston's critique is powerful and important.

Where ECF Says Friston Still Stops Too Early

ECF's disagreement is not that information gain is useless. It is that information gain remains too *instrumental*.

The central ECF move is to treat cognition as organized by the relation between:

- **reach**: what a system opens toward,
- **yield**: what the world, body, or current constraint gives back,
- **presentation**: the temporary stabilization negotiated between them,
- **coherence**: the degree of fit or alignment across these dynamics.

On this view, what other frameworks describe in terms of priors, likelihoods, and posterior beliefs can sometimes be *mapped* onto reach, yield, and baseline reach. But that mapping does not mean the theories are identical. The March 2026 ECF manuscript is careful on this point: the relation to variational free energy is presented as an *interpretive equivalence under a mapping*, not as a derivation or formal reduction.

This yields a sharper criticism of AI:

Standard RL:	optimize reward
Active inference:	optimize expected free energy / information gain
ECF:	organize endogenous coherence under constraint

The difference matters most for curiosity. In active inference, curiosity is explained because information-seeking improves inference. In ECF, intrinsic curiosity should arise because unresolved tension between reach and yield drives reorganization. Curiosity is not a bonus. It is not merely a useful exploratory tactic. It is the expression of a system whose own organization remains open enough to become more coherent.

Why This Matters Technically

A large fraction of current machine-learning work on curiosity is still externally engineered. Pathak et al. reward the agent for prediction error in a learned feature space [3]. Burda et al. showed both the strengths and weaknesses of this strategy at scale [4]. In both cases, curiosity remains a designed signal.

ECF would say that this is the wrong level of abstraction. You do not get intrinsic curiosity merely by replacing one scalar objective with another. You need an architecture whose organization already contains:

1. differentiated functional roles,
2. recurrent mutual constraint,
3. endogenous pressure to reduce incoherence,
4. temporal thickness or persistence,
5. global integration.

This is why ECF is naturally suspicious of purely feedforward or stateless systems. A real curiosity mechanism, on this view, must be dynamic, recurrent, and history-bearing.

A Practical Research Program Combining Friston and ECF

The cleanest practical strategy is not to decide in advance whether active inference or ECF is "the winner." It is to compare three agent families directly.

1. Reward-First Baseline

A standard PPO, SAC, or related RL agent with sparse reward and ordinary exploration mechanisms.

2. Active-Inference Baseline

An agent that plans using expected free energy or a close approximation, with explicit epistemic value and prior preferences. There is already growing work on active-inference-style AI agents in sparse-reward and robotic settings, making this baseline feasible in practice [5].

3. ECF-Inspired Coherence Agent

A recurrent or continuous-time agent with explicit internal variables for:

- **reach** π_t: anticipated viable continuations,
- **yield** y_t: current embodied and environmental constraint,
- **memory** m_t: sedimented history of successful coherence,
- **coherence / incoherence** I_t: alignment or tension between reach and yield.

A minimal sketch is:

$$I_t = D_{\mathrm{KL}}(\pi_t \,\|\, y_t),$$
$$\pi_{t+1} \leftarrow \text{mirror\_flow}(\pi_t, y_t, m_t),$$
$$y_{t+1} \leftarrow f(y_t, o_t, a_t),$$
$$m_{t+1} \leftarrow \Sigma(m_t, \pi_t, y_t).$$

The point is not simply to drive I_t to zero. The real question is whether the system can sustain a useful regime of *nonzero tension*: enough incoherence to remain exploratory, not so much that it fragments, and not so little that it collapses into habit. This is the heart of the ECF intuition.

What the Benchmark Should Actually Test

To distinguish these approaches meaningfully, the environments should stress exactly the cases where ordinary curiosity bonuses or shallow exploration struggle.

Useful testbeds include:

- sparse-reward tasks,
- deceptive local optima,
- changing or partially observable environments,
- tasks that require maintaining multiple live hypotheses before settling.

Evaluation should measure more than return:

- exploration depth,

- resilience to distribution shift,
- recovery after perturbation,
- basin escape or attractor reorganization,
- persistence of exploratory behavior after rewards are removed.

If an ECF-style agent only matches a curiosity-bonus PPO, then ECF has not yet purchased much. If it keeps exploring in a more structured, stable, and open-ended way, especially under deceptive or nonstationary conditions, then the framework starts to matter.

Where Humans Fit

Friston's developmental picture of "curious baby AGIs" learning with humans in the loop remains valuable. But ECF shifts how that role is understood.

- In a reward-first framework, humans act primarily as supervisors or labelers.
- In an active-inference framework, humans help shape priors, preferences, and model structure.
- In an ECF framework, humans help shape the *coherence landscape* itself.

That means:

- broadening viable reach rather than only correcting errors,
- stabilizing useful basins without over-rigidifying the system,
- providing rich, structured resistance rather than scalar rewards alone.

This is closer to parenting, teaching, and cultural scaffolding than to RLHF in its usual form.

So Is Current AI Research Wrong?

The most accurate answer is:

> *Not wrong, but aimed at the wrong primitive.*

Reward optimization is too shallow. Information gain is better, but still often treats curiosity as an instrument for improving inference. ECF argues that the next step is not merely replacing reward with a more sophisticated objective, but building agents whose organization makes curiosity endogenous.

That is a much harder program than adding another intrinsic reward term. But it is also more interesting, and potentially closer to what biological intelligence actually is.

Conclusion

Friston is right that scalar reward is too crude a foundation for intelligence. Active inference is a major step forward because it builds epistemic value and embodied constraint into the objective. But ECF suggests that even this is not yet deep enough. If intelligence is fundamentally about coherence dynamics under constraint, then the future of AI may depend less on better reward shaping or better information gain objectives, and more on building systems whose internal organization can remain open, tension-bearing, and transformable.

That would be a very different kind of machine.

References

[1] K. J. Friston, T. FitzGerald, F. Rigoli, P. Schwartenbeck, J. O'Doherty, and G. Pezzulo. Active Inference: A Process Theory. *Neural Computation*, 29(1):1–49, 2017. Available at: https://activeinference.github.io/papers/process_theory.pdf

[2] K. Friston, L. Da Costa, N. Sajid, C. Heins, K. Ueltzhöffer, G. A. Pavliotis, and T. Parr. The free energy principle made simpler but not too simple. *Physics Reports*, 1024:1–29, 2023. Available at: https://www.sciencedirect.com/science/article/pii/S037015732300203X

[3] D. Pathak, P. Agrawal, A. A. Efros, and T. Darrell. Curiosity-driven Exploration by Self-supervised Prediction. In *Proceedings of ICML*, 2017. Available at: https://arxiv.org/pdf/1705.05363

[4] Y. Burda, H. Edwards, D. Pathak, A. Storkey, T. Darrell, and A. A. Efros. Large-Scale Study of Curiosity-Driven Learning. 2018. Available at: https://arxiv.org/abs/1808.04355

[5] Anonymous placeholder for a recent active-inference AI / robotics benchmark paper discussed in the blog draft. Replace this entry with the exact paper you want to cite in the final version. Example source referenced previously: https://arxiv.org/html/2409.14216v1

[6] A. Vieira. The Experiential Coherence Framework: A Structural-Monist Proposal for Consciousness, Perception, and Learning. Revised manuscript, March 2026. Internal / unpublished draft cited conceptually in this post.