Thesis Proposal

Studying emergence of agency in cellular automata within the FEP framework

Michele Vannucci, Vrije Universiteit Amsterdam

m.vannucci@student.vu.nl

[**Updated PDF Document**]

March 21, 2025

1 Introduction

I propose to study the emergence of agency and self-organization within an enclosed **complex adaptive system**, defining an agent's "brain", in a **feedback loop** with its environment through a **perception-action interface**, defining the agent "body." This can be studied using one of the currently most powerful frameworks in cognitive sciences, the **variational free energy** principle.

The mathematical formulation of the latter is derived with a Bayesian approach as the (negative) Evidence Lower Bound (ELBO) on the log probability of the perceived inputs, and it is analogous to the formulation of Variational Autoencoders (VAEs).

While VAEs are effective for generative tasks, they are not built to represent an agent with internal states and actions. Conversely, the general category of complex systems, to which the human brain belongs, seems natural to model an intelligent agent's internal system.

In detail, among the simplest, yet surprisingly powerful, computational models used to study the chaotic and non-linear nature of complex systems are **cellular automata**. Which can also display adaptive properties. This is why I considered starting my study from them, combining a bottom-up numerical and ML-inspired approach with a top-down mathematical discussion.

> 🏆 **The intended goal of this work** is to investigate and implement minimal (chaotic) complex systems that can effectively display adaptive behavior within their environment, while reinforcing the links between machine learning, free energy, chaos theory and self-organization, highlighting the connections and isomorphisms across different areas of study.

🔮 Finally, my **main hypothesis** could be that the exact definition of the complex system doesn't matter to achieve the emergence of a minimal form of agency, as long as it possesses a set of *mixing* properties. Additionally, I hypothesize that the exact implementation of the perception-action interface does not have a central role either, as long as it effectively isolates the system from its environment (e.g. it's a Markov Blanket). Along these lines, a Turing-complete CA coupled in a feedback loop with an environment—where *surprise* on sensory states can be minimized—could effectively model a minimal "intelligent" agent.

Below, I motivate my interest in this topic by highlighting its relevance within current research. I outline the structure of the proposed study, describe its key focus areas, and detail the methodologies.

2 Passive vs active inference

While the majority of current research efforts in machine learning is devoted to improving and studying differentiable DL models, the cognitive science community, adopting the enactivist view, has already highlighted the potential limitations of *passive* generative models that disregard agency. Pezzulo *et al.* [1] argue that we could proceed on the path of understanding intelligent behavior by studying computational models that actively interact with the world. They claim this is essential for an agent to build a grounded generative model of the environment and its hidden states. Bender *et al.* [2] also refer to the limits of modern models in capturing meaning in the context of NLU .

This is why I opted for the enactivist setting for this work, which entails the agent's model being coupled with an environment through a feedback-loop. This is analogous to the RL setting, but in the latter both the *reward function* and the learning algorithm are pre-defined.

3 The free energy principle

Conversely, the power of the free energy principle is that it can model the agent's "reward-seeking" and adaptive behavior as minimizing the upper bound, or negative ELBO (here corresponding to variational free energy F), on the **surprise** of its sensory input: $-\ln p(s)$. Friston [3] has shown how this can be equivalent to performing a gradient ascent on Value, defined by the Bellman equation.[1]

Additionally, the free energy principle entails that the minimization of surprise over the sensory states s is sufficient to bring about the generation of meaningful internal representation/states μ of the hidden causes ϑ of the external world. The former are optimized to represent the latter in a *Bayes-optimal* fashion, minimizing the KL divergence $D_{\mathrm{KL}}(q(\vartheta|\mu) \,\|\, p(\vartheta|s)$.

[1] Free energy minimization leads to a fixed point on the states distribution, as well as a convergence on policy and value, if defined as inversely proportional to surprise. This implies that the optimal policy has been reached, as the rate of change in free energy is minimal and "according to the principle of optimality, cost is the rate of change in value". See [4] for more on how free energy connects to RL.

Here q is the **recognition density** which is a probabilistic representation, modeled through μ, on the hidden causes ϑ that generated the current sensations s. At the same time, actions optimize the prediction error $-\ln p(s(a) \mid \vartheta, m)$, where m is the system's generative model of how the latent states ϑ generate s, and a is the chosen action to minimize this quantity [3]. The dependence on m can be equally defined parametrizing p as $p_\phi(s(a) \mid \vartheta)$ (as done for VAEs), given that the parameters ϕ define m.[2]

Friston [3] also suggests how different theories in the biological domain (as neural darwinism and predictive processing) and in the physical domain (as dynamical systems and information theory) relate to the free energy as they define the optimization of equivalent or analogous quantities. An example is the Infomax principle which states that the brain maximizes mutual information between internal and sensory states ($I(s, \mu)$), analogous to the KL divergence defined above and the bottle-neck method formulation (if we see s as Y and μ as $\tilde{X}$) [5]. Furthermore, if we look at the formulation of free energy from the point of view of non-equilibrium thermodynamics, we see that it is formed by an **energy** term on the left, while the time-average of $-\ln p(s)$ can be interpreted as the **entropy** of sensory states:

$$F = \underbrace{-\big\langle \ln p(\tilde{s},\, \vartheta \mid m) \big\rangle_q}_{\text{energy}} + \big\langle \ln q(\vartheta \mid \mu) \big\rangle_q \;\; \geq \;\; -\ln p(s) \;=\; \text{surprise}$$

This gives a prescriptive view on what properties a physical system that *exists* has to possess to define its boundary with respect to the environment. Firstly, it has to minimize entropy of its sensory state, going against the natural tendency to disorder by the fluctuation theorem, and at the same time minimize energy or the surprise over the joint distribution of sensory, s, and latent states, ϑ [3]. Finally, another fascinating perspective considers $p(s)$—which is maximized by minimizing free energy—as the *evidence* of the model's own existence [6].

4 Self-organization and the Markov blanket

The concepts of **entropy** and **energy** minimization and self-evidence maximization reinforce the link between free-energy minimization and the self-organization (or autopoiesis in the biological domain) of the internal system. Importantly, it is theorized that any steady-state non-equilibrium system isolated by a **Markov blanket** (constituted by the sensory and action-inducing states) from the environment, defining its boundary (and therefore its *existence*, as it allows one to distinguish it from the surrounding environment), will minimize free energy with time [7]. The blanket creates a causal separation between the internal states of the agent and the external states of the environment, which become causally independent when conditioned on the blanket. Friston [8] considers random dynamical subsystems (particles) with local dynamical interactions and proves analytically how the internal and active states (on the Markov blanket) dynamics (or flow) can be described as a gradient descent on variational free energy. In detail, the internal states "will appear to have solved the problem of Bayesian inference by encoding posterior beliefs about hidden (external) states," under the given generative model. At the same time, active states are shown to place a bound

[2]Alternatively, it would be interesting to model m and a as strictly dependent on the internal states μ, to reduce the parameters to be optimized to only one (multi-dimensional) variable.

on the entropy of the joint distribution $p(s, a, \mu)$. Maintaining the self-organization of the agent. A fundamental assumption of this proof is for the system to be **ergodic**, meaning that the time average of its states converges, after a sufficient amount of time, to its random dynamical attractor corresponding to the variational free energy optima.

5 A computational implementation

While this is considered a too-strong assumption to generalize over every biological system [8, 9], we can still use it to artificially build an agent that effectively minimizes free energy. Simulations in this direction are rare within current literature, motivating the subject of my thesis. In fact, it could be insightful to analyze what properties an actual engineered computational internal system and its barrier (the Markov blanket) could have in order to display intelligent behavior, while taking advantage of the general conditions already prescribed by the FEP (e.g. Markov blanket and Ergodicity).

Friston [8] simulates 128 interacting particles, and then analyzes their behavior given a Markov blanket found *a posteriori* within a subset of these particles (an emergent cell-like structure). In contrast, the latter can be defined *a priori* to equip a computational system with an artificial resilient boundary from its environment. The optimal Markov blanket is speculated in [8] to have low structural and dynamical entropy or, in other words, it should change slower than the internal and the external states it separates. This again can be artificially enforced while designing the blanket.

6 Complex adaptive systems

In the context of designing an artificial agent, the blanket can be defined as the perception-action interface of the internal system and its environment in which we want an arbitrary task to be solved. Since we outlined already some of the properties of the blanket, our focus now falls inward on the internal machinery of the agent and its states. To preserve generality this system can be modeled as a complex system, a category under which any biological or cognitive system could fall, as this is not the case for the already mentioned dynamical systems (which define interactions by differential equations). Finally, there are many computational models that constitute or could model a complex system: message-passing networks, VAEs, reservoir networks, attractor networks, generalized cellular automata, and so on.

7 Cellular automata

For simplicity, I plan to begin my study from (generalized) cellular automata for which ergodic and **mixing** properties have been outlined in literature [10, 11]. Moreover, it has been shown that they can exhibit emergence and self-organization as in the well-known Conway's game of life or as demonstrated by Hamon *et al.* [12], where rules are learned to allow the emergence of entities with sensorimotor agency in the Lenia continuous-state CA. Another attractive property of some cellular automata is that they can emulate a universal

Turing machine [13] in polynomial time [14]. From this, we can speculate that if there is an optimal behavioral algorithm, π^*, within an environment (e.g. which minimizes surprise of the agent), this will be computable by a Turing-complete CA[3] and it could be the dynamic attractor of the latter, if enclosed in a Markov blanket, given the free energy principle.

This optimal behavior can be described as the optimal policy within an RL setting. We already mentioned the inverse relationship between surprise and value above, but we have to point out that minimizing the first maximizes the second only if there is a relationship between reward (or cost), which is usually arbitrarily defined, and surprise.

For example, if we model a biological environment with reward as the food gathered and cost as the energy used by the agent to move around, we can intuitively hypothesize that the agent will be able to gather resources effectively and remain alive only if it correctly minimizes surprise. This means modeling the environment and its hidden causes. In turn, it can minimize surprise only if it gathers resources to move around to collect more evidence.
In this case cost (or reward) and surprise are intertwined. But if we want to solve an arbitrary task, like the mountain-car example, priors based on the cost function have to be artificially implemented. This is done in [4] where they establish a prior on the generative model of the agent, defined as a random dynamical system, so that its equations of motion have stable sinks at the given goal position.
This instructs us on the types of environment and tasks that we could consider in this work. The "goal" behavior has to align with minimal surprise.

In fact, free energy seems ideal (being extremely general) to model chaotic or biological systems in a complex and intricate environment such as the physical world, where no explicit task or quantity has to be optimized beyond *adaptive fitness*, for which free energy could account if we define it as the time-average of surprise [3]. On the other hand, simulating such systems is out of reach and does not allow one to perform ablative experimentation on which minimal properties of the mechanics, of the environment, the blanket, or of the internal system are necessary and how they allow self-organization and emergence of agency.

Therefore, I would start from the simplest experiments that still respect the conditions of the free energy principle, to see how they, combined with the computational power of cellular automata, could allow an agent to efficiently learn nearly optimal policies in a human-designed environment.

8 Example experiments

Here I briefly outline some possible experiments and research methods for this study; the latter can be divided into an empirical and a mathematical approach.

[3]For a CA to emulate any possible Turing machine program it must have an unbounded number of states, but we can assume that a rich amount of the computations can yet be carried out with a finite automaton.

8.1 Numerical simulations

8.1.1 A two-armed bandit

The simplest experiment I could think of involves using a chaotic (possibly ergodic) and Turing-complete mono-dimensional cellular automaton such as Rule 110, with periodic boundaries. We can then model the environment with a Multi-armed bandit setting with only two arms. The hidden causes are formalized as $\vartheta = \langle x, \theta, \gamma \rangle$. x represents the hidden states (the different arms), and $x(t)$ the current one. θ are the parameters of the arms and the transition functions between them, while γ controls the stochasticity of these transitions and of the sensory recordings, defining the *noise* [3]. With $\gamma = 1$ all the update functions become deterministic.

The automaton will then have a single *sensory cell*, $s(t)$, with update function:

$$s(t+1) = \begin{cases} X_i & \text{with probability } \gamma, \\ \neg X_i & \text{otherwise} \end{cases}$$

where i is the index of the selected arm, $i = x(t)$, and X_i is the random variable describing its output. We can define one arm with a very *predictable* outcome $X_1 = \mathit{Bernoulli}(p_1)$ with $p_1 = 1$ and another one with a completely random outcome $X_2 = \mathit{Bernoulli}(p_2)$ with $p_2 = 0.5$.[4] The sensory state s will not be affected by the automaton's local rules, only the states around it. We also need one cell state $a(t)$ encoding the action; the Markov Blanket will then be $\{s(t), a(t)\}$,[5] and these two cells could be placed anywhere in the CA, which contains a total number N of cells. Finally, $a(t)$ is normally updated by the CA rule via neighboring states, so the automaton can only perform two different actions encoded by the possible states of this elementary CA: $\{0, 1\}$. The effect of the action on the hidden states is then:

$$x(t+1) = \begin{cases} 1 + a(t) & \text{with probability } \gamma \\ 1 + \neg a(t) & \text{otherwise} \end{cases}$$

Meaning that action $a(t) = 0$ will select arm 1 and action $a(t) = 1$ will select arm 2 (with probability γ).

I then expect the long term average of $a(t)$ to be $E[a] = 0$, since the optimal policy to minimize sensory surprise would be to pick the arm with lower variance (which is the first one, as $\mathrm{Var}[X_1] = 0$ and $\mathrm{Var}[X_2] = 0.25$).

Further analysis could vary the parameters of ϑ, analyzing the policy adopted by the automaton as well as the trajectories of the internal states, measuring how and if they correlate to the hidden and sensory states (similarly what is done by Friston [8] for the particles simulation), to verify that they actually perform Bayesian inference. Finally, this could allow us to encode the implicit distributions p and q to calculate variational free energy, F, either numerically or with Monte Carlo estimation using the formula defined above.

[4] Crucially, contrary to the usual multi-armed bandit setting, the arm's output value itself does not encode the reward as no reward function is defined. I expect the automata to be influenced primarily by their variance as detailed below.

[5] Figure 1 displays the causal graph of the interactions, we can see that the set $\{s, a\}$ is a Markov blanket for the internal states μ as it blocks all possible paths between μ and x, effectively d-separating the two.

8.1.2 More experiments

Further experiments could be carried out by adding more arms or modeling each arm as an MDP state with stochastic transitions, so that the automaton could not select directly the best arm but would have to go through other arms every time it falls out of an optimal hidden state. The resulting policy could then be compared to one obtained through value iteration.

For the CA to encode multiple actions, we could use more cells as active states (to increase the number of binary-string representations possible), or increase the number of states of a single cell. This means that we would have to change the rules of the automaton, too. Investigating different types of rules and automata is fundamental either way, since we want to isolate the properties of the internal system that lead to adaptive behavior. I could also perform the same experiments with completely different computational models, as enumerated above, or with generalized CA that can have different spatial properties deviating from the classic grid structure.

Furthermore, we could also have more *naturalistic* experiments, for instance placing the automaton in a game environment, where predicting sensory input (e.g. the next player's move) requires planning ahead and modeling the game environment and the opponent, similarly to how humans or MCTS methods do. A less RL-inspired yet more difficult task could be represented by the ARC dataset implemented within an environment to test the generalization capabilities of the model. Nonetheless, this is probably beyond the scope of a single study.

Finally, the same tasks and experiments outlined above can be carried out with a classical RL model, while controlling for the computational budget and information about the environment to match the CA setting. These active-inference-based methods can also be compared with the standard *passive* learning approach with labeled data.

8.2 A mathematical approach

An additional and probably necessary methodological approach for this research involves using mathematical formalism to study the evolution of complex systems such as cellular automata. With the aim of examining the properties of rules that yield stable attractors coinciding with an optimum on free energy and policy. This draws inspiration from the mathematical approach adopted with random dynamical systems to derive an ascent on free energy and Bayesian inference from the dynamics or "flow" of the system [8].

I recognize that for complex systems as cellular automata this is trickier, since we likely cannot have a precise analytical solution as we do with vector calculus in random dynamical systems. However, we could gain precious insights from the current state-of-the-art in the mathematical theory on CA, self-organization, complex systems, and chaos theory to inform our numerical experiments. These might remain the only way to predict the final behavior of such systems, given principles such as computational irreducibility or Kolmogorov complexity.

9 Conclusions

I think chaotic systems and cellular atomata in particular are a fascinating medium to study the nature of self-organization, and how they could give rise to emergent properties such as agency or intelligence. The power of CA and their "learning" capabilities has been highlighted in the already mentioned literature, and in more recent research too [15, 16, 17]. Additionally, [18] shows how they can be used to emulate open-ended evolution, aligning with the evolution-like properties that the brain possesses as theorized by Neural Darwinism.

Nonetheless, there is little, if any, research at the intersection of all the concepts and approaches outlined above, namely Free Energy Principle, Reinforcement Learning, Self-organization, and Complex Systems. This is why I think that a study in this direction could be extremely insightful, and I hope I demonstrated clearly its structure and feasibility. I look forward to any suggestions on how it could be improved.

My interest in these topics started by looking at the life-imitating emerging patterns in CA, and my curiosity was further prompted by reading about the theories in the domains here covered. I'm awestruck by their existential value in trying to explain seemingly unexplainable phenomena such as the emergence of life and intelligence, and I would love to write a thesis about it.

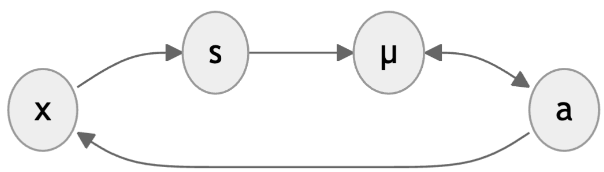

Figure 1: Causal relationships between μ (the *internal* states of the automata), x, s, and a. Given two arbitrary positions for the sensory and active state, $j, k \leq N$, The mono-dimensional CA defined in section 8.1.1 can be formulated as an array of cells: $[\mu_0, \mu_1, .., a_k, \mu_{k+1}, \mu_{k+2}.., s_j, \mu_{j+1}, .., \mu_N]$ with periodic boundaries.

References

1. Pezzulo, G., Parr, T., Friston, K.: Active Inference as a Theory of Sentient Behavior. Biological Psychology **186**, 108741 (2024). `https://doi.org/10.1016/j.biopsycho.2023.108741`. (Visited on 03/10/2025)

2. Bender, E.M., Koller, A.: Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 5185–5198. Association for Computational Linguistics, Online (2020). `https://doi.org/10.18653/v1/2020.acl-main.463`. (Visited on 03/12/2025)

3. Friston, K.: The Free-Energy Principle: A Unified Brain Theory? Nature Reviews Neuroscience **11**(2), 127–138 (2010). `https://doi.org/10.1038/nrn2787`. (Visited on 03/05/2025)

4. Friston, K., Ao, P.: Free Energy, Value, and Attractors. Computational and Mathematical Methods in Medicine **2012**, 937860 (2012). `https://doi.org/10.1155/2012/937860`

5. Tishby, N., Pereira, F.C., Bialek, W.: The Information Bottleneck Method, (2000). https://doi.org/10.48550/arXiv.physics/0004057. arXiv: physics/0004057. (Visited on 03/05/2025).
6. Hohwy, J.: The Self-Evidencing Brain. Noûs **50**(2), 259–285 (2016). https://doi.org/10.1111/nous.12062. (Visited on 03/12/2025)
7. Kirchhoff, M. *et al.*: The Markov Blankets of Life: Autonomy, Active Inference and the Free Energy Principle. Journal of the Royal Society, Interface **15**(138), 20170792 (2018). https://doi.org/10.1098/rsif.2017.0792
8. Friston, K.: Life as We Know It. Journal of The Royal Society Interface **10**(86), 20130475 (2013). https://doi.org/10.1098/rsif.2013.0475. (Visited on 03/12/2025)
9. Colombo, M., Palacios, P.: Non-Equilibrium Thermodynamics and the Free Energy Principle in Biology. Biology & Philosophy **36**(5), 41 (2021). https://doi.org/10.1007/s10539-021-09818-x. (Visited on 03/12/2025)
10. (PDF) On Strong Mixing Property of Cellular Automata with Respect to Markov Measures. ResearchGate (2024). (Visited on 03/13/2025)
11. Pivato, M.: Ergodic Theory of Cellular Automata. In: Cellular Automata: A Volume in the Encyclopedia of Complexity and Systems Science, Second Edition. Ed. by A. Adamatzky, pp. 373–418. Springer US, New York, NY (2018). https://doi.org/10.1007/978-1-4939-8700-9_178. (Visited on 03/13/2025)
12. Hamon, G. *et al.*: Discovering Sensorimotor Agency in Cellular Automata Using Diversity Search, (2024). https://doi.org/10.48550/arXiv.2402.10236. arXiv: 2402.10236 [cs]. (Visited on 03/08/2025).
13. Department of Computation and Neural Systems, Caltech, Mail Stop 136-93, Pasadena, California 91125, USA, Cook, M.: Universality in Elementary Cellular Automata. Complex Systems **15**(1), 1–40 (2004). https://doi.org/10.25088/ComplexSystems.15.1.1. (Visited on 03/14/2025)
14. Neary, T., Woods, D.: P-Completeness of Cellular Automaton Rule 110. In: Automata, Languages and Programming. Ed. by D. Hutchison *et al.*, pp. 132–143. Springer Berlin Heidelberg, Berlin, Heidelberg (2006). https://doi.org/10.1007/11786986_13. (Visited on 03/10/2025)
15. Wolfram, S.: What's Really Going On in Machine Learning? Some Minimal Models. Stephen Wolfram Writings (2024). (Visited on 03/05/2025)
16. Differentiable Logic CA: From Game of Life to Pattern Generation, https://google-research.github.io/self-organising-systems/difflogic-ca/. (Visited on 03/14/2025).
17. Mordvintsev, A. *et al.*: Growing Neural Cellular Automata. Distill **5**(2), e23 (2020). https://doi.org/10.23915/distill.00023. (Visited on 03/15/2025)
18. Plantec, E. *et al.*: Flow-Lenia: Towards Open-Ended Evolution in Cellular Automata through Mass Conservation and Parameter Localization. In: ALIFE 2023: Ghost in the Machine: Proceedings of the 2023 Artificial Life Conference. MIT Press (2023). https://doi.org/10.1162/isal_a_00651. (Visited on 03/09/2025)