# Chemical autonomous agents: reverse-engineering artificial life under the free-energy principle

Johan Medrano[1]

[1]Wellcome Centre for Human Neuroimaging, UCL Queen Square Institute of Neurology, London WC1N 3AR, UK
johan.medrano@ucl.ac.uk

## Abstract

This work examines *chemical autonomous agents* — minimal chemical reaction networks exhibiting dissipation, autocatalysis, homeostasis, and associative learning — under the lens of the *free-energy principle* — a normative account of adaptive systems. The free-energy principle allows us to 1) identify the partition of system states belonging to the agent, 2) uncover how that agent synchronises its internal states to its environment, 3) understand the agent's environmental fitness from the reaction rate constants. This initial work suggests that the free energy principle can provide a systematic approach to decompose, analyse, and understand complex adaptive systems and artificial life.

## Motivating example

(Bartlett and Louapre, 2022) introduced a minimal chemical reaction network embodying basic properties of living systems: dissipation, autocatalysis, learning, and homeostasis (Bartlett and Wong, 2020). The network comprises three key species: a compound (B), a toxin (T), and an antidote (N). Compound B undergoes autocatalytic growth but with a finite capacity, following logistic growth. The toxin, delivered by bolus injection with a concentration profile governed by $u(t)$, rapidly degrades B. The antidote rapidly neutralizes the toxin and slowly degrades B. Chemical kinetics equations describe the concentration evolution of each substance in a zero-dimensional, well-stirred solution.

(Bartlett and Louapre, 2022) used the system to compare three defence mechanisms, i.e., reaction networks generating the antidote to counteract of the toxin. (Bartlett and Louapre, 2022). We focus on their associative learning (AL) network, which introduces three reactions and three compounds, $S$, $M$, $L$. The associative network shows associative learning: the *stimulus* $S$ is delivered shortly before each bolus and catalyses the *short-term memory* $M$, that is transformed in a *long-term memory* $L$ in the presence of the toxin $T$. When the long-term memory encounters the stimulus, it anticipates the toxin and releases the antidote (Fig. 1b).

When the AL network is configured within a specific range of reaction rates, the agent survives by anticipating the toxin bolus and rapidly generating antidote. Outside the range, the agent fails to adapt and dies. Understanding the behaviour within the survival range and associated defence mechanism failures is crucial to understand agent survival. We believe insight can be gained by reverse-engineering the system's dynamics under an information-processing perspective. In this study, we explore reframing and analysing the behaviour of the chemical artificial agent using the free-energy principle (FEP).

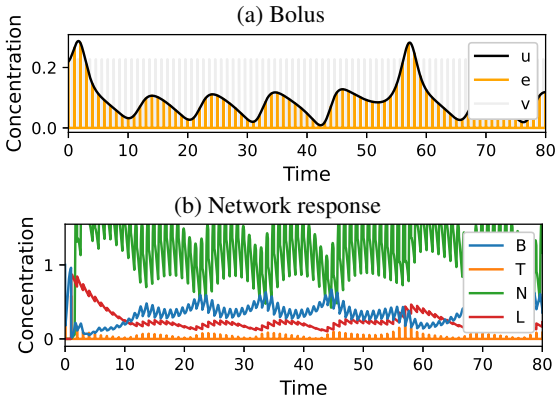


Figure 1: Simulation of the chemical autonomous agent. The amplitude of toxin delivery is driven by the sum of the 3 variables of a Lorenz system, normalised between 0 and 0.2. Parameters from the original work have been adapted for a step delivery of the toxin and stimulus.

## Chemical autonomous agents under the free-energy principle

The FEP is a postulate about the dynamics an open subsystem has to exhibit to adapt to environmental fluctuations and resist thermodynamic dissipation (Friston, 2013; Friston et al., 2023). Here, we present the essential theoretical aspects of the FEP and its application to the chemical autonomous agent.

**State partition** The FEP examines open subsystems,e.g., living organisms, that interact with their environment. Subsystem quantities are categorised as either directly interacting with the environment or not — the latter are the *internal states*, denoted as $\mu$ (Friston, 2019). Mediating states are known as *blanket states*, $b$, as they form a *Markov blanket* for the internal states, i.e., completely determine the internal states. Blanket states are further divided into *sensory states* $s$ and *active states* $a$. Under the FEP, adaptive systems necessitate this partition as it defines the agent and each quantity's role the adaptive behaviour. The partition can be derived from the system's Jacobian matrix (Friston, 2019). For the chemical autonomous agents, short- and long-term memory compounds are internal states, the antidote is an active state, the toxin and stimulus are sensory states, and the bolus amplitudes are external states (Fig. 2a).

**Synchronisation map** Under the FEP, an agent updating its internal model and achieving homeostasis is explained by its the internal and active states flowing down its *variational free-energy* (VFE). The VFE quantifies how well the internal model, stipulatively encoded by the internal states, approximates the *posterior distribution* of the external states given the blanket states. In other words, internal states encode beliefs about external states and update through variational Bayesian inference, syncing with external states (Da Costa et al., 2021). Indeed, regression of external states against internal ones shows the long-term memory synchronising with the toxin bolus amplitude (Fig. 2b). This allows catalysing the antidote proportionally to the toxin bolus amplitude, as seen through the influence graph (Fig. 2a).

**Kullback-Leibler divergence** Using the linear synchronisation map, we can evaluate the Kullback-Leibler (KL) divergence between the approximate and true posterior distribution under Gaussianity (Fig 2c). This KL divergence, part of VFE, gauges the distance between approximations and true posteriors, indicative of an agent's information-based fitness (Hidalgo et al., 2014). Examining KL divergence changes with chemical reaction rates reveals a plateau when a chemical agent dies, i.e., the concentration of $B$ reaches 0 (Fig. 2d). Thus, adaptive behaviour is reflected as nontrivial regimes of information processing in the internal states.

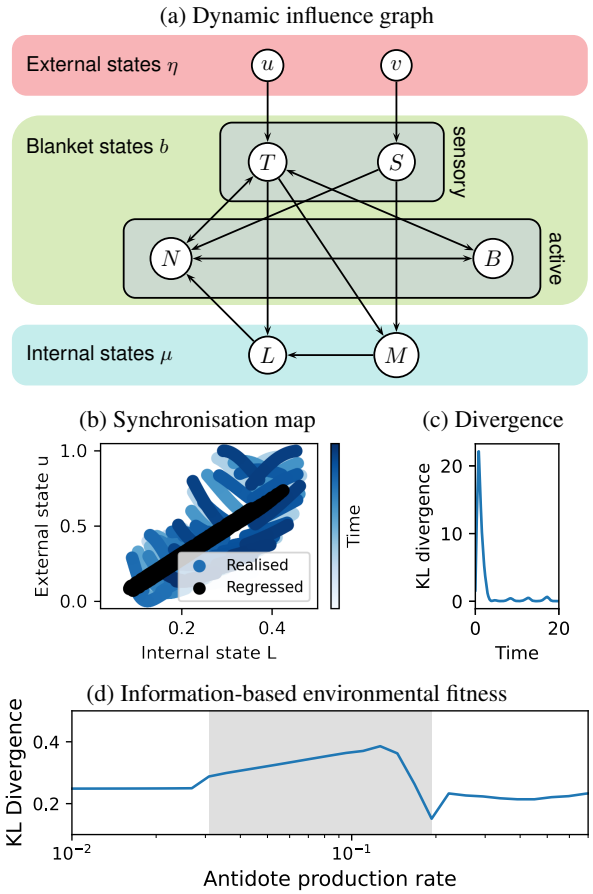


Figure 2: a) Dynamic influence graph for the reaction network. An arrow from $C_i$ to $C_j$ indicate that specie $C_i$ appears in the flow for specie of $C_i$. Coloured frames show a meaningful partition for this example, obtained by considering $L$ and $M$ as internal states. b) Trajectories of the system (blue, same parameters as Fig 1) show synchronisation between long-term memory and toxin amplitude. Using regression of external states against internal states, we create a linear synchronisation map (black). c) KL divergence between approximate and true posteriors, evaluated from the synchronisation map in 500-point rolling windows around the initial time. d) Information-based environmental fitness, measured as KL divergence over the entire trajectory, plotted against the antidote production rate. In the grey range, concentration of $B$ always satisfies $[B] > 10^{-16}$.

## Concluding remarks

This study applies the FEP to chemical autonomous agents. The state partition uncovered from the system's Jacobian allows us to identify an agent in a nonlinear system of chemical reactions. The state partition prescribes the existence of a synchronisation map, which reveals that the agent internally models the toxin bolus amplitude using the long-term memory compound, enabling proportionally catalysing the antidote to control its environment. Naturally, failure to sync internal states to environmental fluctuations causes the agent death, reflected by a lack of information processing. These findings suggest considering the FEP as a principled framework for developing methods to reverse-engineer complex adaptive systems and organisms. Future work will extend this approach to complex biochemical systems, aiming to establish an analytical and computational framework for analysing complex adaptive systems and organisms.

## References


Bartlett, S. and Louapre, D. (2022). Provenance of life: Chemical autonomous agents surviving through associative learning. *Physical Review E*, 106(3):034401.

Bartlett, S. and Wong, M. L. (2020). Defining lyfe in the universe: from three privileged functions to four pillars. *Life*, 10(4):42.

Da Costa, L., Friston, K., Heins, C., and Pavliotis, G. A. (2021). Bayesian mechanics for stationary processes. *Proceedings of the Royal Society A*, 477(2256):20210518.

Friston, K. (2013). Life as we know it. *Journal of the Royal Society Interface*, 10(86):20130475.

Friston, K. (2019). A free energy principle for a particular physics. *arXiv preprint arXiv:1906.10184*.

Friston, K., Da Costa, L., Sajid, N., Heins, C., Ueltzhöffer, K., Pavliotis, G. A., and Parr, T. (2023). The free energy principle made simpler but not too simple. *Physics Reports*, 1024:1–29.

Hidalgo, J., Grilli, J., Suweis, S., Munoz, M. A., Banavar, J. R., and Maritan, A. (2014). Information-based fitness and the emergence of criticality in living systems. *Proceedings of the National Academy of Sciences*, 111(28):10095–10100.