22 November 2021

Frontier Talk

Memo by Marina Sanchez Del Villar

Me and my Markov Blanket

Speaker: Professor Karl Friston (UCL) [1]

This event has been organised by the Technological Change and Society Interdisciplinary Research Cluster

This memo summarizes the first *Frontier Talk* of the Tech Cluster in the academic year 2021/2022. In these events, the Cluster invites leading academics in fields beyond the EUI's faculty to discuss about frontier research related to technology. A theoretical neuroscientist and a leading authority on brain imaging, Professor Karl Friston teaches neuroscience at the University College of London and he is the Scientific Director of the Wellcome Trust Centre for Neuroimaging for neuroimaging.

Professor Friston's talk focused on the free energy principle.

1. The statistics of life

1.1. The concept of a Markov blanket

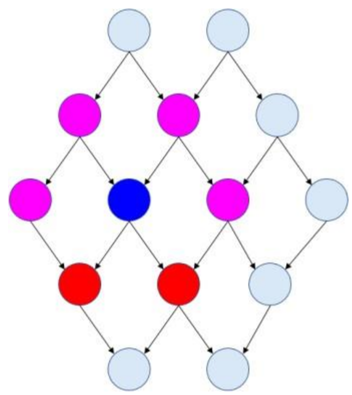
In order to talk about anything, say a particle, you need to be able to define the *thing* that separates the particle from the non-particle. This separation is itself a probabilistic construct. We refer to this boundary as a Markov blanket.

To better grasp the concept, consider a world that contains several states that influence each other in a hierarchical way. Figure 1 provides a schematic representation of such a world. A Markov blanket comprises all the states that surround a given state: the *parents* of the state, the *children* of the state, and the other *parents* of the *children* of the state.

This way we can divide all states of the world in three types: the internal states, inside the Markov blanket; the external states, outside the Markov blanket; and the Markov blanket itself. To know how the internal state changes, given the rest of the universe, one only needs to know how the blanket states change. All information of interest is contained in the blanket states.

The blanket states can be further separated in two types of states: sensory states influence but are not influenced by internal states, while active states are influenced but do not influence the internal state.

Figure 1: Schematic representation of a Markov blanket



Legend: external states, internal states, sensory states, and active states.

Source: Professor Karl Friston's presentation

This relationship structure maps nicely, for example, to the way the brain works: the activity and connectivity in the brain can be thought of as the internal states; the brain then sends signals to muscles or secretion organs, the active states; these in turn influence the outside, external states, whether they are your body or the environment; these external states provide different sensory signals (vision, touch organs) that then cause changes to your internal states. The key take-away from this example is that the mediation between the inside of the brain and the outside is done by the blanket states.

1.2. How do states evolve? The dynamics of self-organization

We can describe the evolution of any*thing* in terms of a trajectory between state spaces. The key insight about the kinds of systems we are interested in is that they always return to the same parts of the state space.

[1] The recording of this talk is available at: https://www.youtube.com/watch?reload=9&v=04F5xhx4ew0

The states that a particle keeps on revisiting define the kind of element the particle is. This also implies that the number of states for any system is limited.

We can use a probability distribution to express the probability of finding a system at any one particular state, at a random point in time. Since the particle always visits the same states, at some point the probability distribution stops changing. This interpretation is useful because we can use off-the-shelf equations from physics to describe the evolution of this probability distribution.[2] The equation that describes the movement of the system is called the equation of motion.

Since there are a number of states to which the particle continues to return, it must be the case that on average it will look as if the probability of moving to one state increases or decreases (randomly moving along the probability gradients). There is another component of the dynamics, which circulates around the possible states (a circular or solenoidal flow). The equation of motion, with its two components, constitutes what is called the dynamics of self-organization.

All systems that return to the same states are governed by these dynamics. Therefore, we can use this fundamental equation of motion to describe all of physics. For example, classical mechanics, used to explain the motion of the planets, focus on the circular flow instead of the random fluctuations; statistical mechanics focus on the random fluctuation and abstract from the circular flow; and complex quantum mechanics use both forces.

1.3. Fitting the Markov blanket in the dynamics of self-organization

Since a *thing* has to have a Markov blanket, and any system of states can be explained by the equation of motion, we can put these two concepts together.[3] In particular, the equation of motion applies to the internal states, the sensory and the active states. Therefore, we can look at these states from the lenses of physics. This interpretation is called Bayesian mechanics, the forces of which are related to perception and action.

---

[2] Mainly, the Fokker-Planck equation and the Helmholtz decomposition.

[3] The point that there cannot be a physical thing that is not defined by its boundaries (ie, must have a Markov blanket) was stressed in the discussion. Professor Friston gave the example of the models that physicists use to teach classical thermodynamics, in which elements are *idealized*, and they do not have a Markov blanket. But as soon as the models introduce a heat reservoir, as is the case with the thermodynamics example, they are introducing the Markov blanket: an exchange of heat of moving from on-equilibrium to off-equilibrium states.

1.4. From self-organization to self-evidencing and generalized synchrony

Once we understand the laws that govern the states of the Markov blanket, the next logical step is to ask ourselves if the internal states appear to infer what causes changes in their sensory states. We refer to this awareness as self-evidencing.

For this purpose, Professor Friston simulates a small universe, called a "primordial soup" consisting of multiple particles, and uses the concept of Markov blanket to define one particle in this universe. To test for self-evidencing, one must look at how the internal states of a system behave. In particular, we look for correlations between changes in the internal states and the external states. In our brain example, this would be equivalent to testing whether the brain responds to visual inputs.

From simulations of the primordial soup, we can see that there exists a remarkable correlation between extreme changes in the external states and responses by the internal states. But the correlation does not provide the direction of the causality: do the external states cause the internal states or is it the other way around?

The answer is generalized synchrony:[4] they both cause each other. While the inside states are trying to infer the outside states, the outside states also try to infer the inside states. The blanket states act as a link between the outside and the inside states.

Because active states change, but are not changed by external states, they reduce the entropy of blanket states. This means that an action will appear to maintain the structural and functional integrity of the Markov blanket. Internal states appear to infer the hidden causes of sensory states by increasing the Bayesian evidence, and actively influence those causes (active inference).

2. The anatomy of inference

We will now use the same narrative but from the point of view of psychology. The question we will ask in this part is: *what must systems that conform to Bayesian mechanics look like?*

In our brain, our perceptions are constrained by what we expect to see and the hypotheses that can be called upon to explain our sensory input.

---

[4] Christiaan Huygens was the first to coin this term when he noticed that all pendula of clocks hanging from the same wall would end up swinging in synchrony.

This means that the brain has hypotheses or potential explanations of what causes the sensory inputs. Perception is simply hypothesis testing of what is generating the sensory data. It requires separating what we see into a prediction and an error (update).

Using the terms introduced in the first part, perception can be thought of as the impressions of the external space on the sensory part of the Markov blanket. We already know the functional form of the dynamics of the brain from the equation of motion. We can rewrite the equation of motion to separate the prediction, represented by the circular flow, and the prediction error, represented by the gradient flow.

When receiving a sensory input, the prediction error is determined by the sensations we expect the input to produce and the actual sensations we experience. The prediction error will drive the expectations until the brain has the best account possible of such sensory input. This update or message passing is done in a hierarchical way: the errors of one prediction are used to update the new prediction.

The reading of Bayesian mechanics applied to the human brain reduces to a minimization of prediction error. Humans have two ways to minimize prediction error: via actions, actively resampling the world by looking for sensations that are more similar to the predictions; or via perception, by literally changing the expectations to make the predictions more like the sensations. The action approach is the simplest way to minimize the prediction errors.

3. Action and perception

3.1 Active inference and planning

The point in this section is that humans do not just minimize surprise in terms of maximizing the probability of encountering the outcomes with the lowest prediction error. Humans also minimize the expected surprise or uncertainty: they explore to resolve the future uncertainty about the world in which they operate.

The architecture of active inference consists of the outside world supplying sensory signals, which are then used to minimize the expectations or beliefs about the external state of the world.[5] Once the agent has the right expectations given the sensory inputs, he can then generate predictions about feelings or sensations in terms of his actions. The agent will then minimize those prediction errors by acting upon the world. This reflexive kind of behavior is engaged in synchrony with the outside world.

There is a more sophisticated way of prescribing action, which involves equipping these models with a model of consequences of behavior through planning. Using the same Bayesian mechanics, we can add a temporal depth to the model so that agents predict into the future the consequences of actions in the world, providing the machinery to plan.[6]

As before, the agent takes sensory inputs that build states of the world based on the minimization of the prediction error. These states of the world are subsequently rolled out into the future around different courses of action, and the agent evaluates the prediction error that he would expect if he committed to such courses of action. Rolled into the future, prediction error becomes uncertainty. At the final stage, the agent uses that uncertainty to take the action that has a minimum expected uncertainty, which will then be executed as part of a plan.

There is a symmetry between the prediction error that underwrites our sense-making or perception, and the uncertainty that one would expect of taking a given action in the future. We can refer to the prediction error as **free energy**, as it comprises an entropy term and an energy term. We can interpret the free energy as the agent trying to find the most accurate explanation of what he perceives, in the simplest possible way. As such, there will always be a lower bound on the simplicity that a system can reach. When moving to the future, the expected complexity becomes risk, and the expected inaccuracy becomes ambiguity.

3.2 Epistemic affordance

Consider an example related to visual inputs. When humans are faced with a visual scene, they only sample a small part of it. We think we see the whole picture because we create a construction through the active inputs, when in fact, we only see points of information at any one time.

---

5 The free energy principle uses the terminology of beliefs to refer to a Bayesian probability distribution, rather than propositional or intentional stance in a psychological sense. Mathematically, changes in beliefs are changes in a probability distribution that induce information geometry. Belief updating hence consists of movements in a physical space, whose states encode a probability distribution. The free energy provides the gradients generating these movements.

6 This active inference and planning are also being applied in some machine learning or artificial intelligence tools.

We build a picture by deploying *visual palpation* to the right parts of the visual scene, to build up a confident expectation of the element we are looking at.

One can explain this behavior in terms of the expected free energy or the salience of centering the visual sampling in different areas of the input. This behavior builds a mental map that in the visual neurosciences is called the salience. Empirically, our eyes move in a way that resolves the most amount of uncertainty, deploying our active vision in the key areas of a figure.

We can also simulate this behavior with a hierarchical model that employs the Bayesian mechanics described before. This example shows the important of the epistemics of being curious creatures, which is prescribed by the physics of self-organization.

4. Questions

Applying the free energy principle to policy making

As social scientists, we are often involved in policy making (improvements of existing or developments of new policy). The audience was interested in understanding how to move from the descriptive, positive side of the talk to a more policy oriented, normative version of this talk. The focus was placed on the new draft of the artificial intelligence regulation.

For Professor Friston curiosity is currently missing in much of artificial intelligence research and machine learning generally. People have committed to a focus on utility reward functions that does not come along with the information seeking that would support epistemic foraging. Additionally, most of human behavior comes from curiosity, and not so much of it is driven by extrinsic rewards.

In terms of policy making, the Bayesian mechanics, which comes from the physics part of the free energy principle, changes the way questions are asked. Currently a policymaker may ask: *what kind of policies do good systems that flourish possess?* Bayesian mechanics turns that formulation on its head. It starts by deriving the mathematical definition of a system that flourishes and then asks: *what properties must it possess?* This is where the physics of self-organizations is important. It does not state *how* to behave, but rather the behaviors that *are* evidenced in systems that persevere.

For policy makers, the interest lies in minimizing the risk and ambiguity of the future policies. The whole point of deriving the expected free energy is to build a probability distribution over policies. The free energy principle is hence the mechanics that underlines the policy making, in the sense of selecting the most probable policy, but it does not tell you what the best policy *is*. In this sense, an interesting scientific enquiry would be to understand the anatomy of message passing and belief updating in the policy arena.

<u>The boundary between the internal and external states</u>

The audience was interested in the nature of the boundary between "internal" and "external", and whether such boundaries are given in the natural world or they are projected by the application of the theory.

The answer from the physicist point of view is that the Markov blanket is not a concept that one would use in a theological sense to make sense of a behavior. It must be there definitionally before you can define anything. If a thing exists, it must have a Markov blanket, and given that there is a Markov blanket, it must possess certain physics. The Markov blanket is not constructed by application to something, rather it defines the thing.

For every level of description there is a Markov blanket, but a thing can comprise a *subthing* (Markov blankets of Markov blankets ad infinitum). There is no unique Markov blanket. Every element contains a Markov blanket, and the equation of motion applies similarly, with modifications for temporal scales. For example, the smaller the unit of analysis, the more important the random component (small particles), while the bigger the unit of analysis the more important the circular flow (big celestial bodies). However, there is no privileged level of analysis or temporal scale at which these Markov blankets are organized. One can always contextualize one Markov blanket with additional Bayesian mechanics.

<u>Interactions between multiple agents and the connection with Economics</u>

Members in the audience asked about the application of the free energy principle to economic dynamics, with an emphasis on multiple agents interacting.

The main idea of the talk can be transposed to a setting with multiple internal states and blankets. An abstract state space can have an ensemble of Markov blankets, which surround internal states, call them institutions. In such a system, an outside state does not need to know anything about the mechanics of the inside of the institutions. Thanks to conditional independence, they only need to understand the dynamics of the Markov blankets that surround them. This simplification of the system greatly reduces the computational complexity of any simulation.

In fact, the free energy principle is mainly useful when we try to simulate dynamics of self-organization under the constraints that apply to Markov blankets. The free energy principle takes a descriptive approach, as it translates dynamics that must exist into equations of motion, which can then be put into a computer to simulate behaviors present, for example in the financial markets.

Back to the system with multiple Markov blankets, the distinction of the different sensors between external and internal is blurred: the active states for one blanket are the sensory inputs for another and vice versa. As much as an agent is trying to learn about the environment, the environment is also learning about the agent. The free energy of such a system is minimized when all blankets are in generalized synchrony.

The connection between uncertainty and life, reproduction, and death

The audience asked about the connection between uncertainty, free energy, life, and death.

Professor Friston pointed out that there are Markov blankets that have almost attained an equilibrium, and hence have almost stopped changing. This is the case for some particles, like stones, where their relevant states are small (for a stone, the states could be simply the temperature). But there needs to be uncertainty to resolve or else the states stop evolving. For all particles, stones included, there is a cyclical nature, and so there is always a next state to which to go.

Applied to biotic creatures, this reading implies that the uncertainty is fully resolved when an organism dies, and it becomes one with the ambient surrounding it. A world without living things would be conducive to one where uncertainty would be almost fully resolved. From the free energy principle's point of view, life is hence connected to uncertainty.

Life, reproduction and evolution can be understood as a natural process of Bayesian model selection. Evolution selects the model that is better adapted to the econiche. Natural selection works by finding the phenotypes that are the best fit to the environment and hence will have a high marginal likelihood of being part of a population.

Could we ever stop natural selection? Consider the science fiction scenario of a static environment, where humans halter evolution by populating the world with non-ageing intelligent artifacts that do not change. The problem is that, if these artifacts learn and are curious, from the point of view of any one artifact there is an inevitable change in their external states. This is because the external states are constituted by creatures that are also learning, and hence, changing.

An artifact that has the right kind of curiosity and lived forever would mean that it has lost the opportunity to do Bayesian model selection, and hence would eventually become obsolete and no longer fit its initial purpose. As we consider more adaptative robots, the more we qualify the adaptative notion, the closer we are to human species and hence natural selection (applied to the intelligent artifacts). This means that death is part of any life cycle.