ENS Paris-Saclay

Modelisation in neuroscience - and elsewhere (MVA 2024)

Active Inference, Curiosity and Insight

BARBERTEGUY Victor
Teacher: NADAL Jean-Pierre

-March 2024-

Contents

1 Introduction

Over the last decades, a part of the neuroscientific community has been developing models to explain how us humans - and thus human brains - can learn so fast with so few data. Among those, Professor Karl Friston, an expert that has spearheaded research since the end of the 90s and invented widely used methods (Statistical Parametric mapping, Voxel-Based Morphometry), proposed the theory of Active Inference that is able to explain human acquisition of knowledge in a unknown environment with a surprising accuracy. The article *Active Inference, Curiosity and Insight* [2], published in 2017, provides a thorough account of this theory as well as a simulation of an agent learning an abstract rule. In this report, we aim at explaining the main mathematical foundations of the theory presented in this work, and to reproduce some features of the numerical simulations with slight changes.

The whole point of the article is to show that *"resolving uncertainties about the world, through active inference, entails curious behavior and consequent 'aha' or eureka moments"* [2]. In other words, they try to explain how resolving uncertainties (by learning, decision making, inference) about the world can explain complex phenomena such as intrinsic motivation (curiosity) or insight ('aha' moments). Let's first birefly define those terms:

Curiosity. a curious, or intrinsically motivated agent, explores its environment simply to learn more contingencies about the world. This results in enhancing the agent's predictive power. Indeed, if it knows the world better, it is more likely to predict the outcome of a given action with accuracy.

Eureka moments. Eureka moments, or gaining insight, is an emerging behaviour in the setting of an agent acquiring knowledge, in which the agent understands something about the environment that was not undestood before. This concept was introduced in 1979, in a study that compared recall for sentences that first left one complex, but that were then understood perfectly.

Active Inference (AI). Active Inference is a concept derived from that of the **free energy principle** (FE) which tries to express action and perception as a minimization of variational FE (VFE).This VFE principle states that any agent must minimize its entropy (i.e., its uncertainty about the world, or expected surprise). In other words, AI is a concept that assumes that optimizing FE is our only prior belief.

What is demonstrated in this article, is that AI results not only in the acquisition of knowledge but also builds a knowledge structure (that we call a **generative model**) that memorizes (through Bayesian model selection) the most probable model of the world. This means that there are uncertainties to be resolved at different steps of the learning and memorizing process. Friston et al. identify three sources of entropy: about the state of the world and the context (**seeking epistemic behavior**), about the policies (**goal-seeking pragmatic behavior**), and the contingencies between them (structure learning, **curious novelty-seeking policies**).

The original article is composed of three sections, about AI and the underlying objective functions followed by a concrete example of abstract rule learning, and finally structure learning and Bayesian model reduction (the third level of uncertainty). In this work, we are only going to cover the first two parts of the article. In our first section 2, we establish the links between AI and common concepts in neuroscience such as mutual information, and define the generative model introduced in the article. In the second section 3, we explain the abstract rule learning setting, and run additional experiments after discussing the original one.

2 Active Inference: defining a model

The framework of AI is by nature very general, and can be applied in a wide spectrum of applications: games, 2-step maze (a classical example in intrinsic motivation), fMRI, saccadic eye movements, ... It is based on a **generative model** (GM) that tries to represent the real world (the **generative process**) as faithfully as possible. The term generative is used to emphasized that in the world, there is an underlying process that generates world states, and the generative model produces outcomes of different modalities (e.g, a visual or auditory feedback) based on the hidden state of the agent. The aim is to learn a generative model that has strong beliefs about how future observed outcomes are produced. In other words, the GM infers the most probable causes (states) of outcomes. We call these states hidden or latents, because they are inferred thanks to (partial) observations. Here is where the FE is used: after evaluating the expected FE of each policy (set of future states) knowing the outcome, the most likely action can be selected, and produces outcomes. Then the loop starts again. In this section, we detail and explain how the generative model is built, which parameters (initial distributions) are chosen and under which assumptions (Markovian), and then derive the free energy to show its link to classical theories in Neuroscience.

2.1 Generative model

The thorough description of the generative model is quite complex, thus we will only detail the main features here. However, much of the concepts come from the theory of Markv Decision Processes (MDP). The generative process is depicted by the Fig. 1.

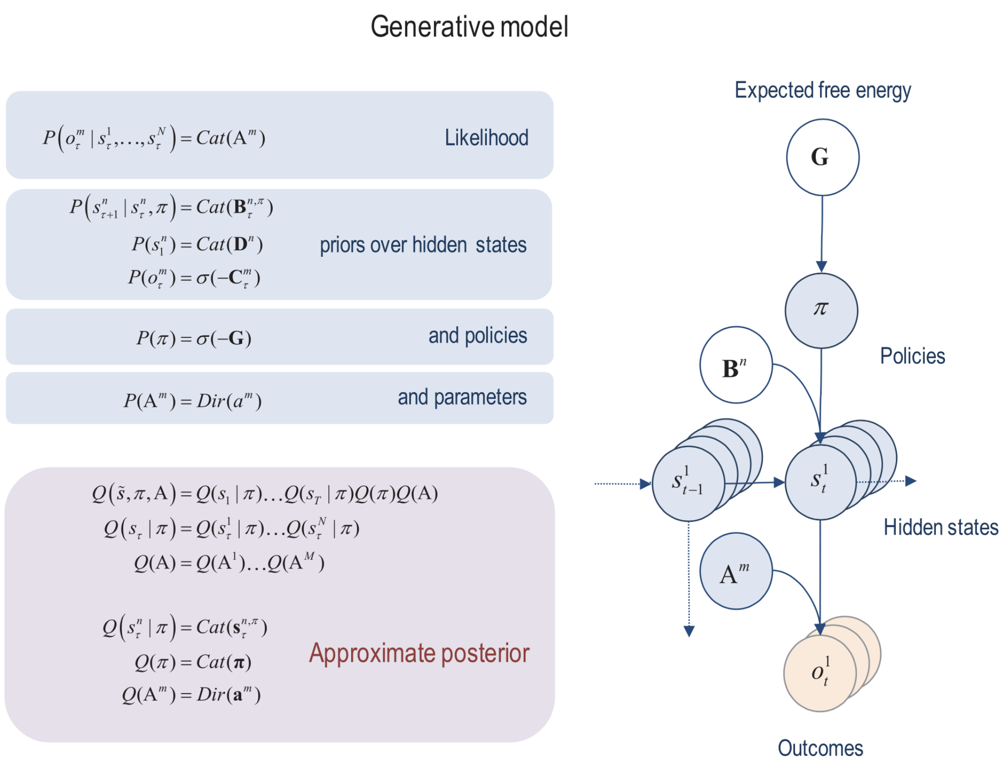

Figure 1: The generative model and approximate posterior that are at play in the AI framework. The approximate posterior expresses the joint probability of outcomes and their latent causes (hidden states). The matrix **A** is the likelihood of the model, mapping hidden states to outcomes, thus **A** is filled with probabilities of having a certain outcome for every combination of hidden states. The matrix **B** is the hidden states transition matrix, and depends on π the policy (the sequence of actions). The gist of this generative model is that the policies are more probable if they optimize the FE **G**. The forms of the posterior needed for variational Bayesian model inversion are specified in the low-left panel. Here, τ represents all possible timesteps and m the diffent modalities (the dimension of the outcomes). Figure taken from [2]

The crucial parameters to be specified for the generative model are the A^m matrix which is the likelihood matrix (the probability of the consequence given the cause) mapping the set of hidden states to the outcome in the m^{th} modality (such as reward feedback, visual feedback, ...). The matrix B^n is the transition matrix which specifies the transitions between the n^{th} latent factor (factors are some features of the world that are observable or not: color, position of objects, speed, ...). This notation for the B matrix has an intrinsic hypothesis that simplifies the problem: the new state for a given factor does only depend on the previous state of this factor, and is independent from others. For example, if my goal is to find something cheap to eat, the visual feedback of the food is independent of its price. To be able to run updates on these parameters, priors about the out outcomes and the initial states must be precised. This is respectively the role of C^m and D^n. As seen 1, those priors are expressed as Dirichlet distributions for which concentration parameters are the only needed. Dirichlet distributions make sense as they can be seen as the number of times a combination of states and outcomes have been observed (and thus how likely the combination is).

The ultimate goal is to learn the likelihood matrix *A* (how outcomes are generated from the hidden states), thus the transition matrix and initial parameters are fixed and known. To learn this likelihood matrix, the generative model sequentially:

- Selects a policy π based on the minimization of the free energy **G**
- Produces a sequence of hidden states thanks to the *B* matrix and the action specified by π
- Generates outcomes thanks to the *A* matrix

In this setting, inferring the hidden states (perception) can be done by inverting the model given a sequence of outcomes, and learning amounts to update the parameters of the model. In mathematical terms, we maximize the expectation of hidden states and policies knowing a sequence of outcomes with respect to VFE, while learning is about updating the concentration parameters. Eventually, we get posterior beliefs denoted as $Q(\tilde{s}, \pi, A)$. The first line of the low-left panel of 1 is using the mean-field theory to get the isolated posterior beliefs of each factors. The main idea of the mean-field theory is to partition the unknown variables and assume each partition is independent (strong assumption). It has been shown that we can then obtain the Q of a given partition by using the previous values of the partition.

2.2 Deriving the expected free energy

In the Bayesian framework, optimizing the VFE with respect to the relevant parameters ensures that expectations represent posterior beliefs given the observed outcomes. As said in the previous paragraph, the beliefs we obtain are not exactly the true beliefs, as we assume they can be marginalized (mean-field assumption). We denote F the VFE of the hidden states, that is the sum between the expected energy and the entropy. Here, we will explain how unpacking these equations in different shapes allows to recognize some known quantities in other fields of neuroscience.

First, we can minimize the free energy with respect to posterior beliefs:

$$\begin{aligned} Q(x) &= argmin_{Q(x)} F \\ &\approx P(x|\tilde{o}) \\ F &= \mathbb{E}_Q[\ln Q(x) - \ln P(\tilde{o}, x)] \\ &= \underbrace{KL[Q(x)||P(x|\tilde{o})]}_{\text{divergence},\geq 0} - \underbrace{\ln P(\tilde{o})}_{\text{log evidence}} \end{aligned} \tag{1}$$

Here, KL denotes the Kullback-Leibler divergence, and $\tilde{o}$ are the aggregated observations up to the current time. As the KL is positive, it means that F is minimized when the posterior is the true posterior. Up to this point, we know that given observed outcomes, minimizing the free energy F entails that expectations corresponds to posterior beliefs. But as explained in the first section, the best policy π minimizes the uncertainty about the world. What is crucial to understand is that beliefs about policies that we encode with expectations also depend on future outcomes. In other words, the cost of a policy is the expected free energy expected in the future. The true distribution over policies P is approximated by the expected VFE for future outcomes. We have:

$$\begin{aligned} P(\pi) &= \sigma(-\mathbf{G}(\pi)) \\ &= \sigma(-\sum_{\tau} G(\pi, \tau)) \end{aligned} \tag{2}$$

We must now minimize the expected free energy G with respect to Q, using the posterior predictive distribution $\tilde(Q)$ over hidden states and their outcomes under a particular policy (obtained in 1). Also, the main contribution of this article is to introduce a new term in the expected free energy, which is the likelihood between the outcomes and the hidden states (the matrix **A**):

$$\begin{aligned} G(\pi,\tau) &= \mathbb{E}_{\tilde{Q}(o^t,s^t|\pi)}[\ln Q(\mathbf{A},s_\tau|\pi) - \ln P(o^t,s_\tau,\mathbf{A}|\tilde{o},\pi)] \\ &= \underbrace{\mathbb{E}_{\tilde{Q}}[\ln Q(\mathbf{A}) - \ln Q(\mathbf{A}|s_\tau,o_\tau,\pi)]}_{\text{(negative) novelty}} \\ &\quad + \underbrace{\mathbb{E}_{\tilde{Q}}[\ln Q(o_\tau|\pi) - \ln Q(o_\tau|s_\tau,\pi)]}_{\text{intrisic or epistemic value}} - \underbrace{\mathbb{E}_{\tilde{Q}}[\ln P(o_\tau)]}_{\text{extrinsic or expected value}} \end{aligned} \tag{3}$$

This last form is quite interesting. The last term, the extrinsic or expected value, results from the minimization of F, and states that the prior over future outcomes is the real value P. The intrinsic or epistemic value can be very well interpreted as it can be directly linked to **Bayesian surprise** or **Mutual Information**. We can use works about such principles of Maximum Mutual Information [1] to interpret the behavior of our agent. If the agent is confident about s_τ the state of the world (there is no uncertainty), then we maximize the mutual information, and there is no epistemic value anymore: the extrinsic value (exploitation behavior) takes over. On the contrary, if the observations provided do nor bring more information, the agent will eagerly look for new observations that informs it about the state of the world (to maximize the mutual information): it's exploration.

Finally, the last term means that policies will be more likely if the bring new information about how outcomes and hidden states are connected. This term can also be seen as mutual information, and we want to update the matrix A so that it corresponds to the matrix A when knowing the future combination of hidden states and outcomes. Minimizing the expectation between these two posteriors mean the agent seeks to expose itself to novel combinations of states and outcomes.

Now that we have our objective function, it is possible to update the beliefs. These update rule need quite complex derivations that will not be developed here. 2 summarizes these updates, and show a schematic of how the learning process takes place in the brain. The main point to understand is that the most resource-consuming part (the part that brings the most cognitive load) is the learning of the parameters of the likelihood matrix A. Friston et al. then interpret this feature from a neurobiological perspective, saying that the mapping between observations (outcomes) and hidden states, and inferring those same hidden states, heavily relies on the context and the goal at stake.

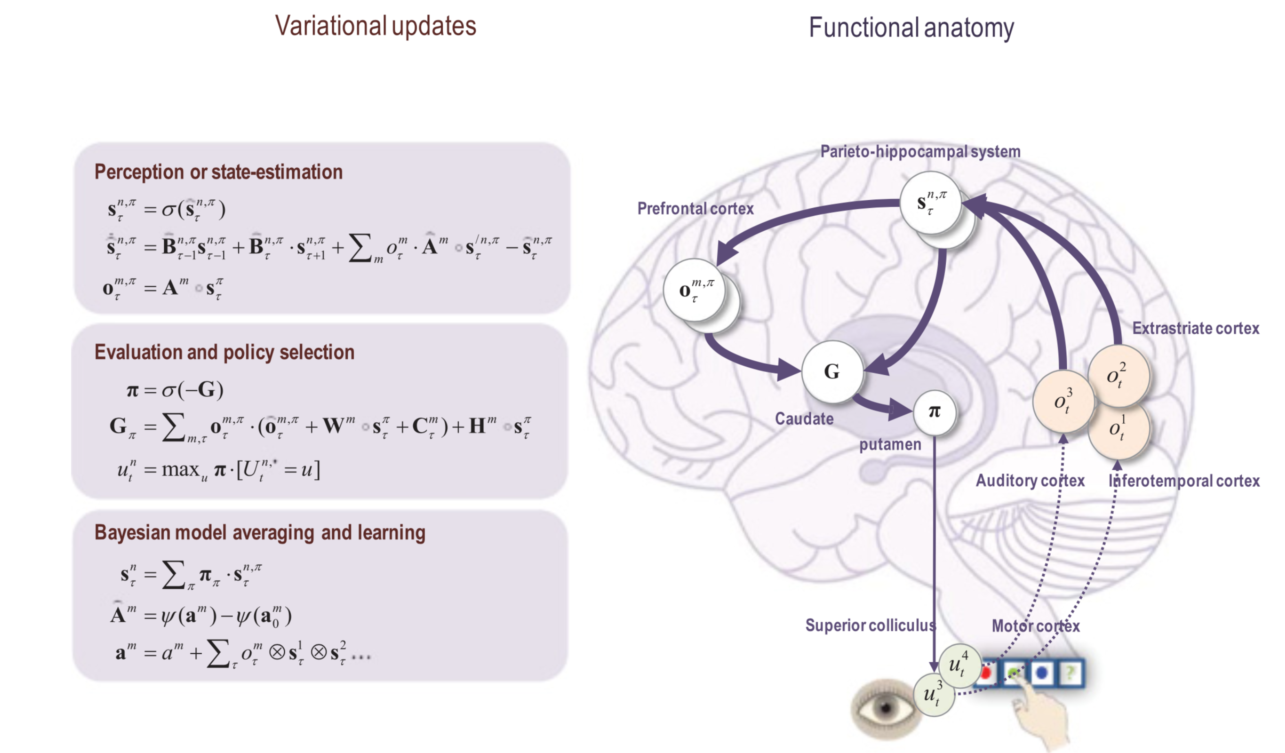

Figure 2: Schematic representation of the belief updating. The equations on the left show the results of the derivation of the belief update, thus giving a practical way to compute variables at time t. The right panel show i a very simplified way one possible configuration of how information flows between the brain. It is more to understand the dependencies between the different variables rather than giving a formal account of brain activity. Taken from [2]

3 Curiosity and Learning: a practical example

The model described in this article is then applied to a concrete example of **abstract rule learning**. Put simple, the agent faces a game, and has to infer the correct rule by simply playing. In this section, we describe the game rule to then explain and criticize the results of the article in terms of curiosity and learning (we will skip the parts about the simulated neural firing rates ad the Bayesian model reduction that simulates sleep or reflection). Finally, we come up with a numerical simulation that mimics this game, and play this game in a setting that could pave the way for more realistic results.

3.1 The three dots game

Let's first explain the game for us humans. The game is quite simple. For one trial, you will be shown three colour dots, that are whether green, blue or red (there ca be several or all dots of the same colour), and one is the correct dot. There is a hidden rule that is at play, and the only information you have is that the color of the correct dot depends only on the color of the central dot. However, you can only see one dot at a time (right, left, or middle), and can only, for each trial, see three dots. Then you must look at the starting dot (grey one) and tell your choice. The game stops when after a certain number of trials, you have found the abstract rule that governs the game.

For example, we will set the rules as follows:

- If the central dot is red, the correct color is the color of the dot on the left
- If the central dot green, the correct color is the color of the central dot
- If the central dot is blue, the correct color is the color of the dot on the right

The Fig. 3 shows one possible configuration of a game.

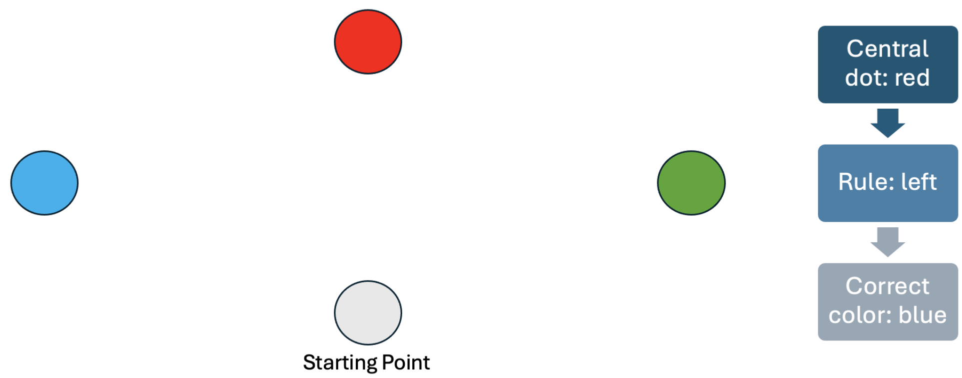

Figure 3: Example setting of a trial. The agent, based on past experiences ad subsequent belief updates, must infer the chain of thoughts indicated on the right as fast as possible.

The main difficulty of simulating the agent is to know how to convey to the agent the rules of the game without language. For example, how to have the agent knowing from the beginning that the correct color is determined by the central dot ? In fact, all these prior beliefs about the world are encoded in the likelihood matrix **A** and the transition matrix **B**.

First, for this game we have **three different modalities**, that's to say that the outcomes comeunder three different ways. There is a **visual** feedback (the color that the agent currently sees), a **proprioceptive** outcome (where it sees: right, left or middle) and **auditory** feedback (if the agent has made the correct choice or not). These three outcomes can be mapped to **four sets of hidden states**, two that are called contextual (i.e., they don't change or a given trial): the rule at play (left, right or middle) and the correct color (red, green or blue), and the two others that depend on the agent: where it is looking at, ad the choice it currently has in mind (red, green, blue or undecided). To play the game, the agent, based on its observations, can choose between 6 actions: looking in one of the three directions, or coming back at the starting point and choose one of the three colours.

Although the ensuing transition matrices are quite simple (for the contextual states, it is identity as the rule and the correct color don't change for a given trial, and for the others, the new state of where and choice do only depend on the action that the agent chose), building the matrix A is not so simple, that is why we will detail it here. Practically speaking, we create a different A matrix for each modality called A^m. The $where$ and $feedback$ likelihood matrices are deemed to be known by the agent. Indeed, it knows that the feedback depends on the correct color, and that it gazes at the direction determined by the current state of the world. In other, words, we fill the 5 dimensional matrices A^{where} and $A_{feedback}$ with high concentration parameters for the correct contingencies, and zero elsewhere. The A_{what} matrix is a little more tricky to fill. The learned A_{what} matrix and the initial one are shown in Fig. **??**, and are really the important part that links the hidden states of the world to what color the agent will see. It shows how each correct color is linked to the seen color for each direction of gaze (columns) and rules (right). In the perfect A_{what}, the agent knows that when it looks at the center (center column), the rule at play (rows) are uniquely determined by the color they are seeing (there is no uncertainty about the rule). Here, we reasoned in an inverse way: we fetched the hidden states for a given outcome. A example of "downstream thinking" would be: if I look at the left, and the rule is left, then I am sure that what I see is the correct color (that is why there are identity matrices for rule: left (resp. right) and sample: left (resp. right). The initial A_{what} matrix is full of uniform priors, except that it knows that the rule is entirely determined by the central cue. Implicitly, the agent knows that there are three rules at play (which is a direct consequence of having three possible colors), but does not know the rules at play, nor the correct colotr: only where it is looking at and its current choice. Finally, it is also worth noting that for the feedback modality (which would correspond to more or less the negative reward in classical Reinforcement Learning), we force the agent to make a choice on the third timestep if it has not done it before by setting highly unlikely values for $undecided$, and a high likelihood for the choice it has in mind, even if it is the wrong one.

The core message to draw from this part is that we were able to abstract the problem from human-understandable instructions into a set of prior beliefs for the different parameters at play in the Generative Model.

3.2 Solving the game

The article indeed shows that the agent, using belief updates according to the model defined in 2, is able to find the correct answer with high confidence about 14 trials. The Fig. 5 indeed exposes this achievement. However, this is not the

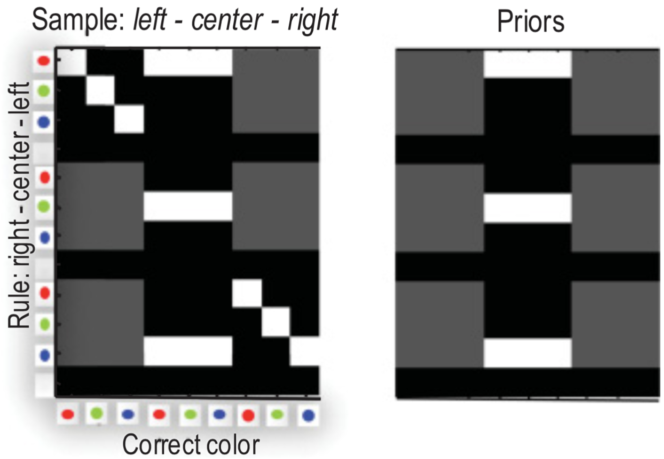

Figure 4: This is a convenient display used in the original article to show the important parameters of the high-dimensional A_{what} matrix. There are 3x3 matrix (rules and where), each subdivided in a 4x3 (what and correct) matrix. On the right, we see the learned A_{what} with the correct priors, i.e. when the agent has found out the rule. Looking the cue at the center entirely determines the rule at play, and looking at right (resp. left) when the rule is right (resp. left) ensures the outcome is the correct color. However, when a rule is right (resp. left) and the agent samples the left (resp.right) direction, it has a uniform prior (so it has no information) about what color it sees, or about the correct color. On the left, this is the initial A matrix for the agent. It knows the paraddigm (the rule is determined by the center cue), but nothing else.

main focus of these results. Here, we can observe how acting with intrinsic motivation to seek exploration reduces the uncertainty in the mapping between the hidden states and the observed outcomes.

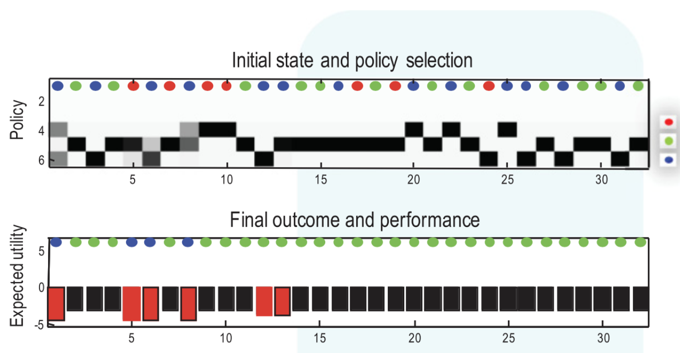

Figure 5: Results of the learning process of the agent. The blue area shows the moment from which the agent does not make anymore mistake. The upper part shows the initial state, and the likelihood to sample the three locations. On the lower part, the final choice and the feedback outcomes are shown. The x axis represents the number of trials. The blue area is the zone in which the agent behaves perfectly. Thus, the agent has learned the rule in 14 trials. Taken from [2].

Recall, in Equation 3, a new term (compared to prior work by Friston et al. is introduced. We still have the uncertainty about the state of the world (intrinsic value) and our preferences for some observations (exploitative behaviour or extrinsic value), but there is a new set of unknowns which are the contingencies (parameters of the generative model) between hidden states ad outcomes (A matrix). There is a certain attractiveness to put yourself in siutations where you can learn the world's structure (reduce your ignorance). Entailing this behavior intrinsically in the equation amounts to search for novelty. And that is precisely what the agent has used to solve the game in a hand of trials. We see that during the first trial, given the initial state, the agent has an equal appetite to sample new locations: it is because each one of those will bring the same amount of novelty to reduce its uncertainty about the world. When it does not make errors anymore, it selects samples only the state of the world that will lead it to the correct answer. As it has learned the rule, it perfectly knows the state of the world, and has a preferred location to sample to reduce uncertainty about the correct color (intrinsic value). Put simply, whereas novelty drives the behaviour during the first tries, the extrinsic value (preferred observations: looking at the left) and the intrinsic value (knowing the state of the world, i.e the correct color (it is the only state it doesn't know)) dominate.

Although this example efficiently shows how the new term introduced by the articles ensures an attraction to novelty, one hypothesis done throughout the article might be questionable, or at least alleviated, to have a more reality-like game. Of course, there are assumptions that are rightful, such as the assumptions that the new states of the world, under a specified action, depend only on the state at the previous time for the same factor. That is what allows us to fill B matrices separately. However, in the results showed here, the agent, when it fails, has the right to start agai for three trials. That means that it can try any colour to get the good results. This is different from how human would play. If someone tells 'I think the right answer is the blue colour', we will only tell if he/she is right or wrong, then stating the correct colour. That means that the agent has the chance to come back to the game and and generate new outcomes to reduce its uncertainty about the A matrix. I think that this makes the learning for the agent faster (hypothetically 3 times) than playing the game in a setting where a new configuration is shown, whether the agent makes a mistake or not.

3.3 A simple example of novelty seeking behavior

Following the remarks done above, we first tried to reimplementing the three dots game, this time not allowing the agent to sample new locations after making a choice. The Active inference and belief update routines used in the original article are done using $matlab$ [1], but a python library has recently been created [2] (2022), with simple examples of how active inference could be implemented. The initial goal was to contribute to this library by implementing the generative model of the three dots game, and then to use the built-in routine functions to run training and active inference.

The first step is to create the different A and B matrices, which require a perfect understanding of how they are built. The main issue is that the descirption provided in the article seems somehow uncoherent, or at least information is missing. Indeed, it is clearly stated that the hidden state of where feedback is of dimension 4, yet is is plotted only in 3 dimensions in the display of the likelihood matrix A. Moreover, the classes that are encapsulated in the $pymdp$ library do not allow to deal with complex behaviours, such that switching temporarily the beliefs of the likelihood matrix to force the choice of the agent at timestep 3. Implementing this game would require a substantial amount of work and is thus not in the scope of this report.

To better exemplify active inference, and the role of the likelihood, we used the built-in environments and agents provided in the library. This simpler cases, albeit not showcasing how powerful the AI framework is, allow the user to better understand the learning loop and how the A matrix is updated. The **T-maze** problem is the basic example in active inference, yet it enables to see very quickly how exploitative behaviour takes over novelty seeking behaviour. See Fig 6.

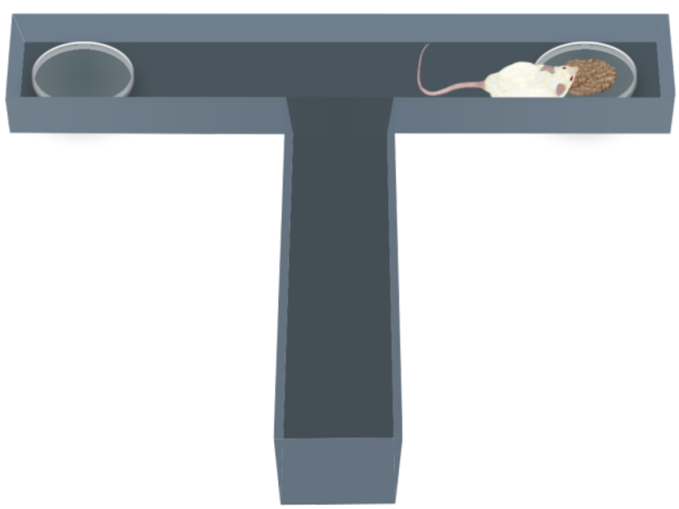

Figure 6: In this environment, the mouse starts at the center location, and can the whether go to the central arm, where there is an information about the location of the reward (right or left) or directly in one of those two arms.

Here, what is hidden is that the reward is always located at the same place (right or left). The mouse doesn't know this, and thus samples the cue location at the first trial, because it knows that (prior belief) that it will provide information

[1]https://fr.mathworks.com/products/matlab.html
[2]https://github.com/infer-actively/pymdp

about the hidden state (reward). After the first timestep, the extrinsic value dominates, and the mouse goes straight to the right. As it finds the reward, it increases even more its confidence, thus leading it to choose the right location even more. From timestep 2, we deem that the mouse has a complete understanding of the world's structure (it has learned the matrix A). These results are shown in 7.

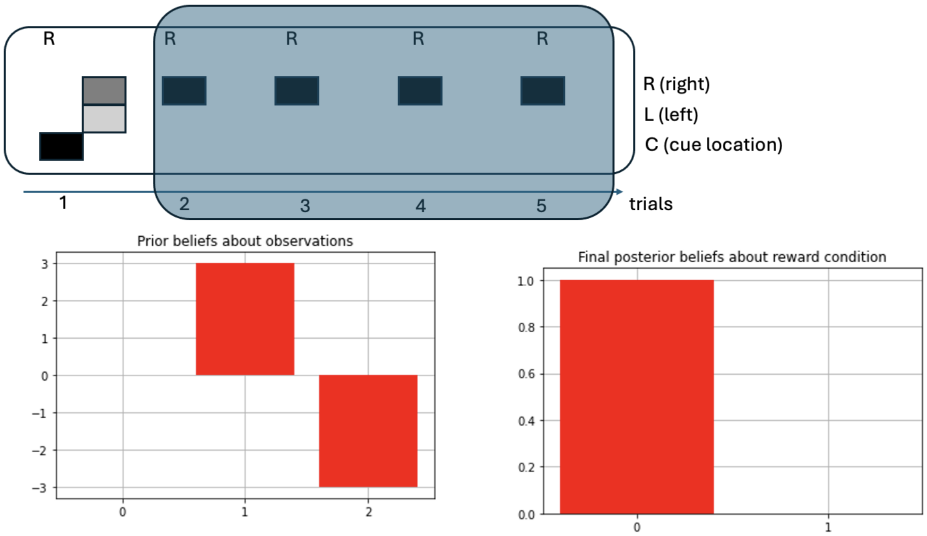

Figure 7: The matrices show that the initial beliefs about reward location are uniform, but in the end the mouse knows the correct mapping between the observations and the state of the world. On the upper panel, we see that it first goes to the cue location to then exploit its information. As the reward is always at the same place, the mouse gains increasing confidence (preferred outcomes) and goes straightly to the right location.

4 Conclusion

Throughout this report, we delved into the theory of Active Inference, that proposes a model to explain why human brains learns so much from few data. The main contribution of the work by Friston et al. is to introduce a new term in the expected free variational energy, following the insight that uncertainty unpacks in three different categories: from the hidden state of the world, from the policy itself and from the mapping between the observed outcomes and the hidden states of the world. Whereas the previous active inference equations accounted for the uncertainty about the hidden state of the world (intrinsic value) and the policies (extrinsic values), the newly introduced term in the FVE is the novelty. This novelty characterizes how much do one knows about the world's structure. It entails that minimizing free energy leads to novelty seeking behaviour, i.e discovering new combinations of observed outcomes to infer new combinations of hidden states.

Whether it is on simple environment (Tmaze) or more complex (three dots problems), this generative model belief updates yields very good results, yet assuming a had of hypotheses for simplification. Please note, however, that a whole part of the article has not been explained in this report for the sake of concision. In a nutshell, this framework can be used to simulate neuron firing rates and reaction times when the agent is confident or seeking novelty. What is shown is that when the agent is confident about the state of the world, it makes a choice quicker than when it seeks novelty. Finally, at the end of the article, a whole section on Bayesian model reduction is used to remove the unnecessary parameters. This model reduction is supposed to represent sleep or reflection, and allows the agent to infer the abstract rule in the three dots problem in a number of trials similar to humans (between 5 and 7).

References

[1] Horace Barlow. “Possible Principles Underlying the Transformations of Sensory Messages”. In: *Sensory Communication* 1 (Jan. 1961). doi: 10.7551/mitpress/9780262518420.003.0013.

[2] Karl J. Friston et al. “Active Inference, Curiosity and Insight”. en. In: *Neural Computation* 29.10 (Oct. 2017), pp. 2633–2683. issn: 0899-7667, 1530-888X. doi: 10.1162/neco_a_00999. url: https://direct.mit.edu/neco/article/29/10/2633-2683/8300 (visited on 03/13/2024).