Active Inference agents increase their empowerment when taking explorative actions.

Summary:

- We aim to show that selecting actions that maximize information gain (active inference) increase the (subjective, potential) empowerment of an agent in the Tmaze.
- We also aim to redefine information gain in the language of Quantum information theory and show that if we remove the traps of the Tmaze, Left and Right actions also become informative.
- For both scenarios, we will first fix the generative model manually to prove the above statements (theoretical argument). Then, we will structure learn the model with a Tensor Network (TN) to show this empirically.

The structure is as follows:

1. Definition of Active Inference, Empowerment and its subdivisions (subjective vs objective, potential vs actual)
2. Definition of the Tmaze environment and its special cases.
3. Classic TMaze with traps in Left and Right arms.
    a. We fix the generative model to show that empowerment increases after gaining info.
    b. We train a Tensor Net to recover the same model.
4. Scenario without traps
    a. We fix the generative model to show that Left and Right are also informative now
    b. We train a TN to recover the same model.
5. Discussion on information gain as entanglement detection.

1 Empowerment as a metric

Empowerment is defined as the maximum Mutual information (MI) between actions and future observations over all action distributions. In our case, the agents will only plan for one time step so we only consider the next observation. We will remove the subscripts for clarity in subsequent equations.

$$\text{Empowerment}_t = \max_{p(a_t)} I(A_t; O_{t+1})$$

The MI can be written as a function of just two distributions, p(a) and p(o|a).

$$I(A;O) = \sum_a \sum_o p(a)\, p(o \mid a)\, \log \frac{p(o \mid a)}{\sum_{a'} p(a')\, p(o \mid a')}$$

They correspond to a choice of action distribution that maximizes the MI and a conditional distribution that represents the possible transitions in the environment.

1.1 Objective vs Subjective Empowerment:

If the transition probability p(o|a) is the true one, i.e. given by the environment, the empowerment that is computed with it shall be termed **objective empowerment**. If the transition probability is the belief of an agent, it shall be termed **subjective empowerment**.

Objective empowerment, on the one hand, could refer to the amount of accessible unique observations/actions available from a particular location, for a particular planning horizon length (the number of time steps into the future one looks ahead). On the other, it could refer to the empowerment of an agent whose beliefs are entirely accurate. That is, its subjective transition probability matches the objective one.

In plain words, objective empowerment is the degree of agency. Subjective empowerment is the perceived degree of agency, i.e. the degree of agency the agent believes it has.

1.2 Potential vs Actual Empowerment:

The formula for empowerment (irrespective of choice of subjective or objective transition) is computed over the maximum over all possible action distributions p(a). This is an assumption of free choice. The assumption is that the agent could consider all possible action distributions in order to select the one with highest MI.

What happens when this assumption does not hold? That is, when there is a constraint on the choice of action distribution. When action is determined by motivation other than maximizing MI of actions and future observations. The empowerment moves from being **potential** to **actual.** Potential empowerment is just another name for the vanilla empowerment formula to differentiate it from actual empowerment, which breaks the assumption of free will when the choice of action distribution is determined. Actual empowerment is just the MI of actions and future observations for a particular choice of action distribution.

Actual empowerment is not relevant to the main hypotheses but was part of Axels investigation into why Empowerment decreases when agents spend money or pass the football. Money represents Potential empowerment in the sense that increases the amount of actions available. However, a drug addict with lots of money may have very low Actual Empowerment, as his preferences constrain his action selection to only purchase drugs with it.

1.3 Empowerment of Active Inference agents.

The choice of action distribution in Reinforcement Learning (RL) is based on Reward Maximization, or Exploitation. The Free Energy Principle adds an Exploration term to the action selection known as Active Inference. We will investigate the Subjective Potential Empowerment of an Active Inference agent in the Tmaze. Potential empowerment is chosen because our main hypothesis is the general statement that Active Inference leads to higher Empowerment. General statements about Actual empowerment are difficult as it depends heavily on the choice of action distribution.

2 The T-maze environment.

We will work with an simplified environment to show our hypothesis and increase the chances that a TN structure learns it correctly. The environment is defined by 3 possible actions a mouse can take, *a* \in {Left, Right and Cue}. It is therefore also known as a three-armed bandit. Each action leads to a different room. When the environment is initialized, 50% of the time, a piece of cheese is placed in the right room and an electric shock in the left. The other 50% of the time the cheese is on the left and the shock on the right. The Cue contains an arrow pointing to the location of the cheese. The 4 possible observations are *o* \in {Cheese, Shock, Left, Right}. This environment is run for two time steps, so that one episode is composed of {a_1, o_1, a_2, o_2}.

If the mouse has a preference for seeing cheese, the optimal policy (set of actions) is going to the Cue in the first action and going to the room that was present at the cue location, o_1.

In our experiments we assume that the mouse has seen every possible episode and its generative model is accurate. That is, its p(o|a) is the true one.

3 Empowerment increases after inferring Actively

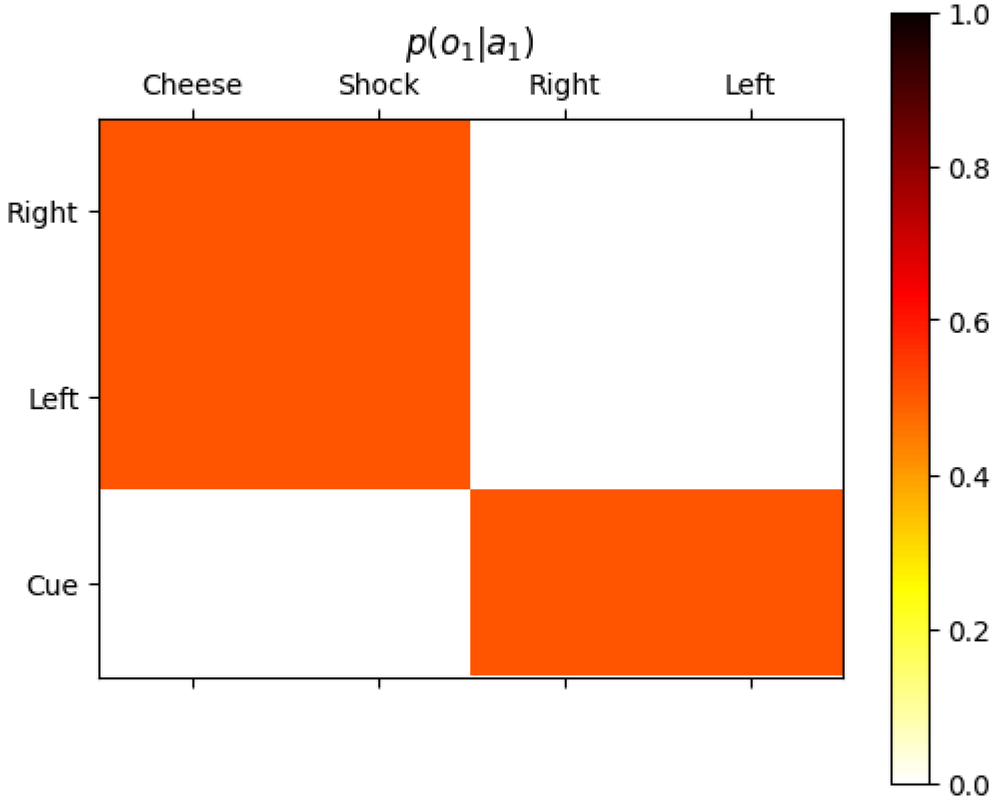
We first explore the classic scenario where the mouse is stuck if it chooses Left or Right with a_1 to show that empowerment increases after going to the Cue.

If we compute the Potential empowerment at time step 1 and then have the agent choose an action that expects to gain some information (Active Inference), we expect the empowerment at time step 2 to be higher. Intuitively, the better one understands the environment, the more control one has over it, or Knowledge is Power.

3.1.1 Empowerment at t_1 is 0.9 bits

The belief of how its first action leads to an observation is shown here. The p(a) that maximizes MI is uniform and the Potential empowerment is 0.9 bits.
The maximum MI would be 1.5 bits as bottlenecked by the action variable being the smallest with 3 states and log (3) = 1.5.

Maximum MI = $\min\left(\log|\mathcal{A}|, \log|\mathcal{O}|\right)$



The reason why the Empowerment is not maximal is that the Right and Left actions are redundant. They result in the same observations and are therefore indistinguishable. The agent effectively believes it only has 2 actions (Right/Left and Cue), or it has 3 actions but is uncertain if Right or Left leads to the cheese/shock.

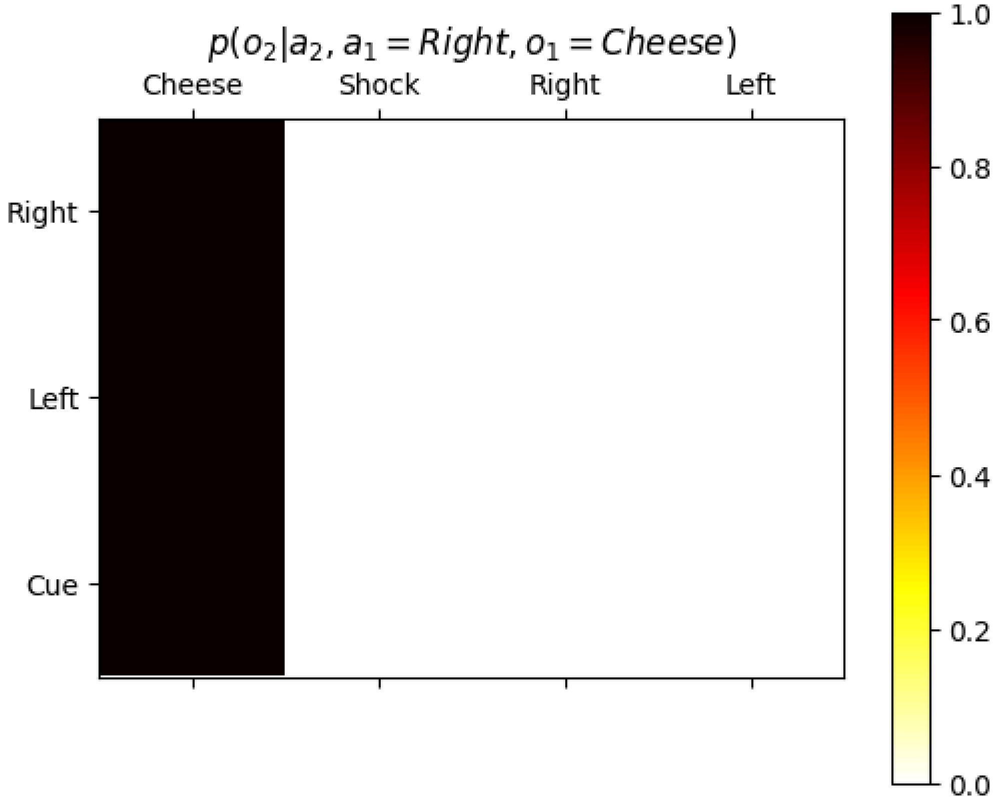
3.1.2 Only the cue action expects to gain 1 bit of information.

Missing section showing that only the Cue expects to gain info

3.1.3 Empowerment at t_2

3.1.3a Getting trapped decreases empowerment

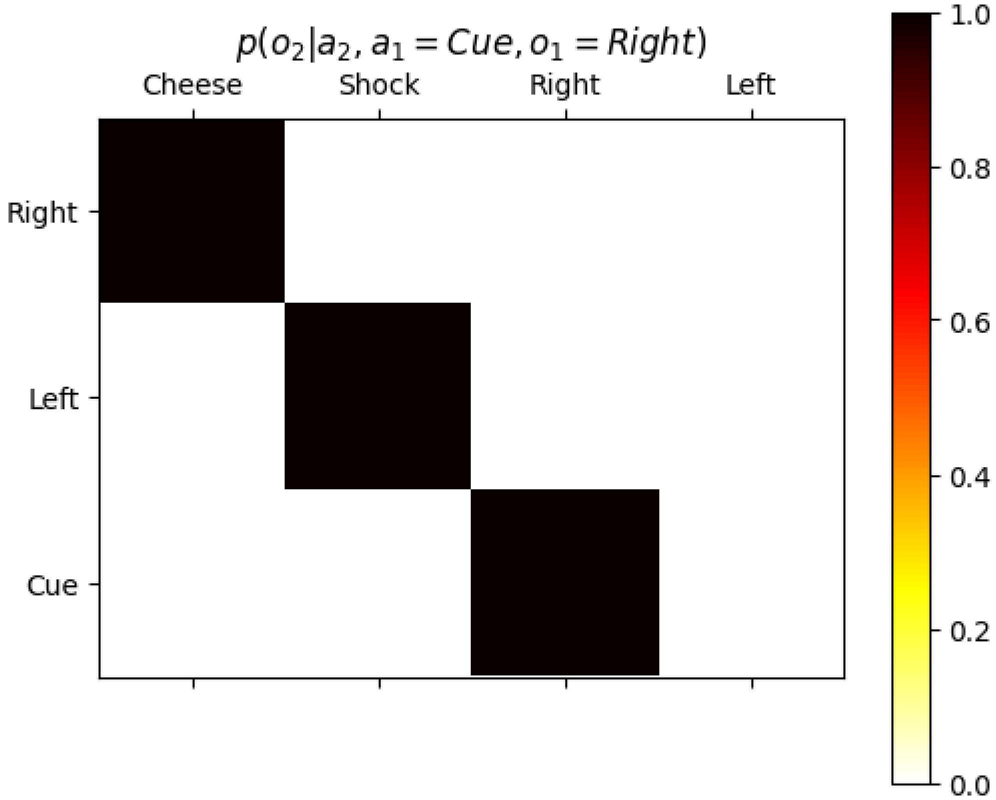
If a_1 was Right or Left the expected observations will be the same regardless of action as the mouse is trapped.



The choice of p(a) is irrelevant as the potential empowerment for this kind of transition belief is always 0 bits. Its actions are all redundant, so it effectively only has one action and log_2(1) = 0. This is intuitive. A trapped agent has no agency.

3.1.3b State exploration increases empowerment

If a_1 was Cue, the transition belief would look like this if o_1 = Right :
(its analogous for o_1 = Left)



Uncertainty has been reduced about the observation at the Cue location which resolves the redundancy between Right and Left. The agent now has 3 unique actions instead of 2 and so its empowerment is 1.5 bits instead of 0.9 (at t_1).

3.2.1 Structure Learning with Tensor Networks

Missing section on recovering these transition beliefs by learning them from data.

4 No traps scenario. Showing that Right and Left give information if they are not traps. That is, positions in the environment give information if they enable movement to positions that are entangled with it.

Missing section on manually defining the beliefs for this alternative scenario and then recovering them by learning them from data.