Post-decision wagering has been used as method for measuring participants’ awareness of how likely their decisions are correct. Human studies have shown that people are more likely to choose a safe bet option when they sensed a high uncertainty in deciding the correct action to avoid task penalty and maximize the reward. Model-free reinforcement learning technique, such as actor-critic method, has been shown to be able to exhibit similar performance in terms of action selection in a post-decision wagering task, giving potential insights of the computational mechanisms of human decision making. An interesting extension would be to apply post-decision wagering to examine the agent’s learning of an internal model of environmental state transitions. Here, we propose a modified two-step task in which the agent is asked to report the predicted environmental state. The agent could make a choice among the possible environmental states and receives large reward or penalty depending on the correctness of the answer or make a safe bet option to receive a tiny reward. We hypothesize that a successfully learnt agent is more likely to make a safe bet option when there is a high uncertainty associated with the underlying transition of the environmental state.

[need to add project significance and how we could probe the model to understand brain mechanism and how is this novel compared with the model based reinforcement learning with learnt probabilistic model in computer science literature. Need to discuss details of the model. If we just use actor critic, then we would be using model free learning and it does not explicitly learn the underlying model? ]