Task-4

★ Propose and describe some technical update(s) in the ML-related component of the "method", and justify why your proposed update is expected to perform better than the one presented in the paper in around 900-1000 words. It should not be based on using more complex technique(s) and/or features.

➔ The paper investigated the role of finite levels of iterated reasoning and non-selfish utility functions in a Partially Observable Markov Decision Process model that incorporates game theoretic notions of interactivity. The generative process for a recognition model that is used in the paper is to classify 200 subjects playing this game against randomly matched opponents.

➔ We know that in practice, POMDPs are often computationally intractable to solve exactly, so computer scientists have developed methods that approximate solutions for POMDPs. Grid-based algorithms comprise one approximate solution technique. In this approach, the value function is computed for a set of points in the belief space, and interpolation is used to determine the optimal action to take for other belief states that are encountered which are not in the set of grid points. We can also make use of sampling techniques, generalization techniques and exploitation of problem structure, and these can extend POMDP solving into large domains with millions of states. For example, adaptive grids and point-based methods sample random reachable belief points to constrain the planning to relevant areas in the belief space.

➔ Dimensionality reduction using PCA has to be explored because it removes correlated features, improves algorithm performance, reduces overfitting and improves visualization. The features in our case are (crudely envy, parameterized by α) and (guilt, parameterized by β). Here we can notice that these social factors are related to each other for an individual. There is also a need to incorporate some of the other parameters for our model, such as the Investor's envy and the temperature parameter of the softmax distribution in order to capture the nuances in the interactions.

- ➔ In general we know that each player has k levels of strategic thinking as in the Cognitive Hierarchy framework. There was an assumption in the paper that each k-level player assumes that his opponent is a k−1 level player, But this may not yield better predictions because players are randomly matched and we have to consider that opponent can also be a k-2 level player.

- ➔ In the paper we observe that the recognition model tends to misclassify Trustees with low β_T as having $k_T = 2$. This is because the Trustees with those characteristics will offer high amounts to coax the Investor. Investors are shown to be correctly classified in most cases. Overall the recognition model has a tendency to assign higher k_T to the Trustees than their true type, though the model correctly assigns the right cooperative/uncooperative type to the Trustee. A very wide range of patterns of dyadic interaction can thus be captured by varying just the limited collection of parameters of the model. There are few specific values in the paper like $\beta_T^{low} = 0.3$ and $\beta_T^{high} = 0.7$, To test the robustness of the recognition model authors also generated behaviours with different values of β_T (β_T^{low} = [0, 0.1, 0.2, 0.3, 0.4] and β_T^{high} = [0.6, 0.7, 0.8, 0.9, 1.0]). But what we need to know is that in practice we can't discretize a parameter like β, so we can consider a large number of values lying between 0 to 1 like around 100 or 1000 to approximate practicality.

- ➔ The problems resulting from a high-dimensional belief space can be reduced through the use of compression. One such approach is Roy's AMDP algorithm, which bears a strong relation to the algorithm presented in the paper. AMDP compresses distributions by representing them using their mean and entropy. The important difference between AMDP and the work described in this paper is that AMDP relies on an underlying discretisation.

- ➔ I also propose dynamic-programming updates for POMDPs that can be interpreted as the improvement of a finite-state controller. This interpretation can be applied to both value iteration and policy iteration. It provides no computational speedup for value iteration, but for policy iteration it results in substantial speedup by making policy evaluation straightforward and easy to implement. This representation also has the advantage that it makes a policy easier to understand and execute than representation as a mapping from regions of information space to actions. In particular, a policy can be executed without maintaining an information state at run-time.

➔ For POMDPs, policy evaluation has low order polynomial complexity. Therefore, policy iteration appears to have a clearer advantage over value iteration for POMDPs. Preliminary testing bears this out and suggests that policy iteration significantly outperforms value iteration as an approach to solving infinite-horizon POMDPs.

➔ Point-based algorithms may be surprisingly successful in computing approximately optimal solutions for partially observable Markov decision processes (POMDPs) in high dimensional belief spaces. In this work, we can seek to understand the belief-space properties that allow some POMDP problems to be approximated efficiently and thus help to explain the point-based algorithms' success often observed in the experiments. We can show that an approximately optimal POMDP solution can be computed in time polynomial in the covering number of a reachable belief space, which is the subset of the belief space reachable from a given belief point.

Therefore by adapting few of the above updates might probably result in increasing accuracy of predictions or decreasing computational time but in few cases we need to trade-off between efficiency and time complexity.