Opinion Formation Dynamics in Multi-Agent Swarms

Nitish Gudapati* and Shashvat Jayakrishnan†
Group #108 - AA228/CS238, Stanford University

Studying information spread (or opinion formation dynamics) can give us various insights, from understanding how some trends go viral on TikTok, and knowing how rumours spread in your highschool, to even predicting who might win the next presidential elections. In this paper we study how opinions form and propagate in a network of multiple agents, and thereby how decisions emerge in a population. To simulate the social dynamics of decision making, we consider the problem of leader-less decision making in swarm robotic systems and culminate in the idea of consensus through a Markov Decision Process (MDP) formulation.

I. Introduction

Spreading of information over social networks is vastly explored in the literature from a graph theoretical perspective [1–5]; graphs are typically used to represent people/agents as nodes, and their interactions as edges. Similar to this idea of information spread is the spread of epidemics, which is a commonly used mathematical model to study 'virality' beyond the scope of just epidemics and also extended to study the spread of ideas, rumors or political perceptions [6]. Standard voter models specifically emerged from the study of formation and spread of political views and opinions in a population [7, 8]. These models highlight the social factors that affect convergence, majority opinion and biases in a population.

What do these models have in common? They all deal with multiple agents, study the spread of some information that one or more agents possess (be it ideas, trends, opinions, preferences, gossip or even COVID-19), and ultimately aim to observe the prevalence of that information over the entire population, understand step-by-step how the population got there, abstract details about high-level factors that influence society and finally attempt to control these factors to achieve desired collective behavior [9, 10].

Another approach to studying information spread and opinion formation dynamics is through modelling the problem as an MDP [11]. This is preferred over traditional graph theoretical approaches in order to make the computations more tractable and the modelling more feasible as complex graph based approaches often become impractical to work with. In this work, we follow this MDP approach. MDPs are an attractive option owing to their ability to capture nonlinear and stochastic dynamics of the system effectively.

In order to deal with the independence of multiple agents in such a formulation, Multi-agent MDPs (MMDPs) are an attractive alternative [12, 13], especially when it comes to dealing with very large population sizes. However, in our work, we have reframed the multi-agent problem into a single-agent problem by combining individual (corresponding to each agent) states, actions, rewards and transitions into a unified state and action space, as well as correspondingly redefined rewards and transitions. This is primarily because our model deals with smaller population sizes which do not have a lot of reward or utility interdependence between agents. Thus, a regular MDP formulation was simpler to model as opposed to an MMDP formulation.

II. Approach

Throughout this work, the term 'opinion' is used as an umbrella term that encompasses several similar concepts from social dynamics like ideas, interests, fads, political views, etc.

A. Markov Decision Process (MDP)

As mentioned in the Introduction, in this work we use an MDP formulation to model the problem of opinion formation dynamics in multi-agent swarms of robots. An MDP is a sequential decision making formulation. Although in reality, decision making in a population of multiple agents may be simultaneous within smaller clusters, we assume

*First-year Master's student, Department of Mechanical Engineering, Stanford University (gnitish9@stanford.edu)
†First-year Master's student, Department of Electrical Engineering, Stanford University (shashvat@stanford.edu)

that each of these smaller decision making gatherings occur at an infinitesimally sequential order, and that at a particular time instant no only one interaction/gathering happens in the population. This assumption allows us to reframe the decision making within a population as a sequential problem and therefore allows us to implement an MDP to model the same.

The MDP is a tuple $\langle \mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R} \rangle$ where $\mathcal{S}$ is the state space combining the states of all agents and $\mathcal{A}$ is the action space combining all possible actions taken by all agents. $\mathcal{T}$ is a vector of matrices of transition probabilities associated with each state action next-state combination for the combined state and action space model. The probability for transitioning from state s to s' given a action a is given by $\mathcal{T}(s, a, s')$. Lastly, $\mathcal{R}$ is typically the set of rewards associated with each state-action pair, however in our combined or state and action space model the rewards depend not only on current state and action but also on the next transitioning state. The immediate reward for taking an action a from state s is given by $\mathcal{R}(s, a, s')$.

Our goal is to find a policy π that maps every state $s \in \mathcal{S}$ to an action $a \in \mathcal{A}$, such that the expected utility associated with implementing the policy π is maximized while the objective of consensus is reached.

$$U_\pi(s) = \mathbf{E}\left[\sum_{t=0}^{\infty} \gamma^t \mathcal{R}(s_t, \pi(s_t), s_{t+1}) \;\middle|\; s_0 = s\right] \tag{1}$$

$$\pi(s) = \underset{a}{\operatorname{argmax}}\left[\sum_{s'} \mathcal{R}(s, a, s') + \gamma \mathcal{T}(s, a, s') U(s')\right] \tag{2}$$

B. Agents

The number of agents (N_{ag}) corresponds to the population size. At a particular time instant each agent can have a particular opinion (state) and can choose to interact (action) with others in the population. To make the modelling dynamic, the number of agents can be modified and scaled as required.

C. State Space

As mentioned previously, each agent has the independent ability to opine. The opinion corresponding to each agent is the individual state of that particular agent. Each agent can have one of many opinions and the total number of opinions per agent is given by N_{op}. In our model, each state $s \in \mathcal{S}$ contains the combined information of all the individual states of every agent in the population at any given time instant.

Let $[O^{(1)}, O^{(2)}, O^{(3)}, \ldots, O^{(N_{ag})}]_t$ be vector of the individual opinions corresponding to each agent in the population at any given time instant t, where $O^{(j)} \in \{o_1, o_2, o_3, \ldots, o_{N_{op}}\} \; \forall \; j = 1, \ldots, N_{ag}$.

We have encoded this combined information about agent opinions onto one state $s \in \mathcal{S}$ as follows:

$$s = \left[\sum_{j=1}^{N_{ag}} O^{(j)} * (N_{op})^{N_{ag}-j+1}\right] + 1 \tag{3}$$

Thus for N_{ag} agents and N_{op} opinions per agent, we have $|\mathcal{S}| = (N_{op})^{N_{ag}}$ states in $\mathcal{S}$.

D. Action Space

In addition to having opinions, each agent can independently choose to perform one of two actions - interact with others in the population (i_1) or ignore/avoid interactions (i_2). Therefore, the number of allowable actions per agent $N_{ac} = 2$.

Let $[I^{(1)}, I^{(2)}, I^{(3)}, \ldots, I^{(N_{ag})}]_t$ be vector of the individual actions corresponding to each agent in the population at any given time instant t, where $I^{(j)} \in \{i_1, i_2\} \; \forall \; j = 1, \ldots, N_{ag}$.

We have encoded this combined information about agent interactions onto one action value $a \in \mathcal{A}$ as follows:

$$a = \left[\sum_{j=1}^{N_{ag}} I^{(j)} * (N_{ac})^{N_{ag}-j+1}\right] + 1 = \left[\sum_{j=1}^{N_{ag}} I^{(j)} * 2^{N_{ag}-j+1}\right] + 1 \tag{4}$$

Thus for N_{ag} agents and $N_{ac}(= 2)$ actions per agent, we have $|\mathcal{A}| = (N_{ac})^{N_{ag}} = 2^{N_{ag}}$ states in $\mathcal{A}$.

E. Transition Model

The transition model is defined for the combined state and action space for all agents in the population. The probability for transitioning from state s to s' given an action a is given by $\mathcal{T}(s, a, s')\ \forall\ s, s' \in \mathcal{S},\ a \in \mathcal{A}$. In our implementation, the transition model is a $|\mathcal{A}|$ length vector of matrices of size $|\mathcal{S}| \times |\mathcal{S}|$. The transition probability $\mathcal{T}(s, a, s')$ corresponds to the value in row s, column s' of matrix a in the vector $\mathcal{T}$. Therefore, $\mathcal{T}(s, a, s') =$ `T[a][s,s']` in Julia.

The transition probabilities are set randomly for agents that *interact* while the self-transitions are set to 1 for agents that *ignore*. The joint transitions of all agents are the product of individual transitions. For action a corresponding to $[I^{(1)}, I^{(2)}, I^{(3)}, \ldots, I^{(N_{ag})}]$ where all $I^{(j)} = i_2$ for $j = 1, \ldots, N_{ag}$, $\mathcal{T}(a)$ is an identity matrix. This means that none of the agents interact with each other in the population and thus neither of them change their opinion.

For actions $a \in \mathcal{A}$ which correspond to only one agent interacting and rest all ignoring, we assume the agents are self-interacting which means independent knowledge gathering or enlightenment. Thus, in this case the transition matrix $\mathcal{T}(a)$ may not be identity.

The transition probabilities corresponding to each agent can give interesting insights about the agent. For an agent who interacts, a higher self-transitioning probability upon interaction means the agent is *stubborn*, while a higher transitioning probability to a specific other state may imply a *bias*. An interacting agent with nearly similar probabilities of transitioning to other opinions that are higher than the self-transition probabilities may be a rather *gullible* agent. The combination of these transition probabilities in the joint transition matrices in $\mathcal{T}$ give us information about the extent of stubbornness and biases of a given population.

F. Reward Function

The rewards associated with each agent in the population are tabulated below:

Conditions	Cost	Reward	Net Reward
State Transition	1000	0	-1000
No Interaction	500	0	-500
Enlightenment (Self Interaction)	0	50	50
Interaction Consensus	0	100	100
Final Consensus	0	10000	10000

Table 1 Rewards corresponding to each agent

These individual rewards corresponding to each agent are accumulated in our combined state and action space formulation, thereby giving rise to a cumulative reward which is a summation of all individual rewards of agents. This way we ensure independent decision making and individual reward maximization while also incentivizing combined actions of agents in the population for the greater good. Note that these rewards are a function of current state s, current action a, and next state s'.

G. Value Iteration

Value iteration is a type of an exact solution method that ensures computing an optimal policy given an MDP model. In value iteration we directly update and improve the value function by performing a *Bellman update*:

$$U_{k+1}(s) = \max_a \left[\sum_{s'} \mathcal{R}(s \mid a, s') + \gamma \mathcal{T}(s \mid a, s') U_k(s')\right] \tag{5}$$

Performing this update for several iterations converges into an optimal value function $U^*(s)$ which satisfies the *Bellman optimality equation*:

$$U^*(s) = \max_a \left[\sum_{s'} \mathcal{R}(s \mid a, s') + \gamma \mathcal{T}(s \mid a, s') U^*(s') \right] \tag{6}$$

The optimal policy $\pi^*(s)$ corresponding to this optimal value function is given by:

$$\pi^*(s) = \underset{a}{\operatorname{argmax}}\, U^*(s) \tag{7}$$

The algorithm is implemented using QuickMDP model, DiscreteValueIteration and POMDP.jl libraries. [14]

III. Results and Analysis

A. Simulation

The optimal policy generated by the MDP value iteration method is simulated using rollouts. A new state is sampled from the current state taking the corresponding action given by the policy based on the transition model of the MDP.

In the visualization, each agent is represented by a circle or marker in the simulation and each action of the corresponding agent is represented by the colours of the circles or markers. The agents interacting in each time step are highlighted using a black circle around the corresponding markers. Fig.1 shows the simulation for 5 agents with 3 Opinions (States) each. The consensus is reached in 6 iterations in this case. For the same model and initial state, the random policy took 71 iterations for consensus.

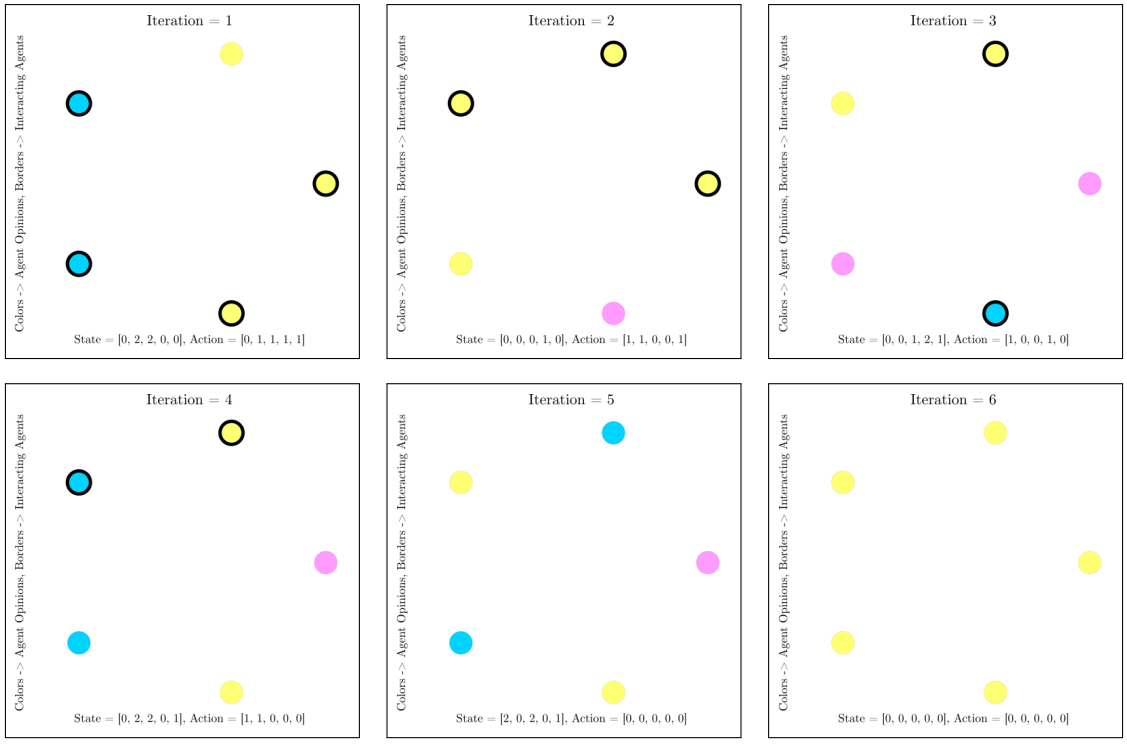

Fig. 1 Simulation using the Policy generated by Value Iteration

B. Performance Metrics

The value iteration method is compared with a randomly generated policy from the POMDP.jl library. To validate the performance of the algorithm, the simulation is run for 1000 times for each set of number of states and number of agents. A normal distribution is fitted for the cumulative rewards of the 1000 simulations in each case and plotted. Fig.2a shows the normal distribution of cumulative rewards for 3 states and 2 agents, and similarly Fig.2 b,c,d correspond to 3, 4 and 5 agents respectively.

Fig.2e shows the variation of the mean cumulative rewards as the number of agents increases for the value iteration policy and random policy. Fig.2f shows the variation of the standard deviations for the same cases. It can be observed from Fig.2e that the mean cumulative reward of the random policy exponentially decreases with increase in the number of agents. The mean cumulative reward of Value iteration policy gradually increases and is significantly greater than the random policy.

From Fig.2f, we can observe that the standard deviation significantly increases with number of agents in case of random policy while in thee case of Value iteration, there isn't much change. This shows that the Value iteration policy is consistent. It is worth noting that for very low number of agents, there might be some cases where the random policy might get better cumulative reward that the value iteration policy due to the stochastic nature of the transition model, where a random policy might transition to a state with very low probability due to randomness. However in general, the value iteration policy performs relatively better and significantly improves as the agents increase.

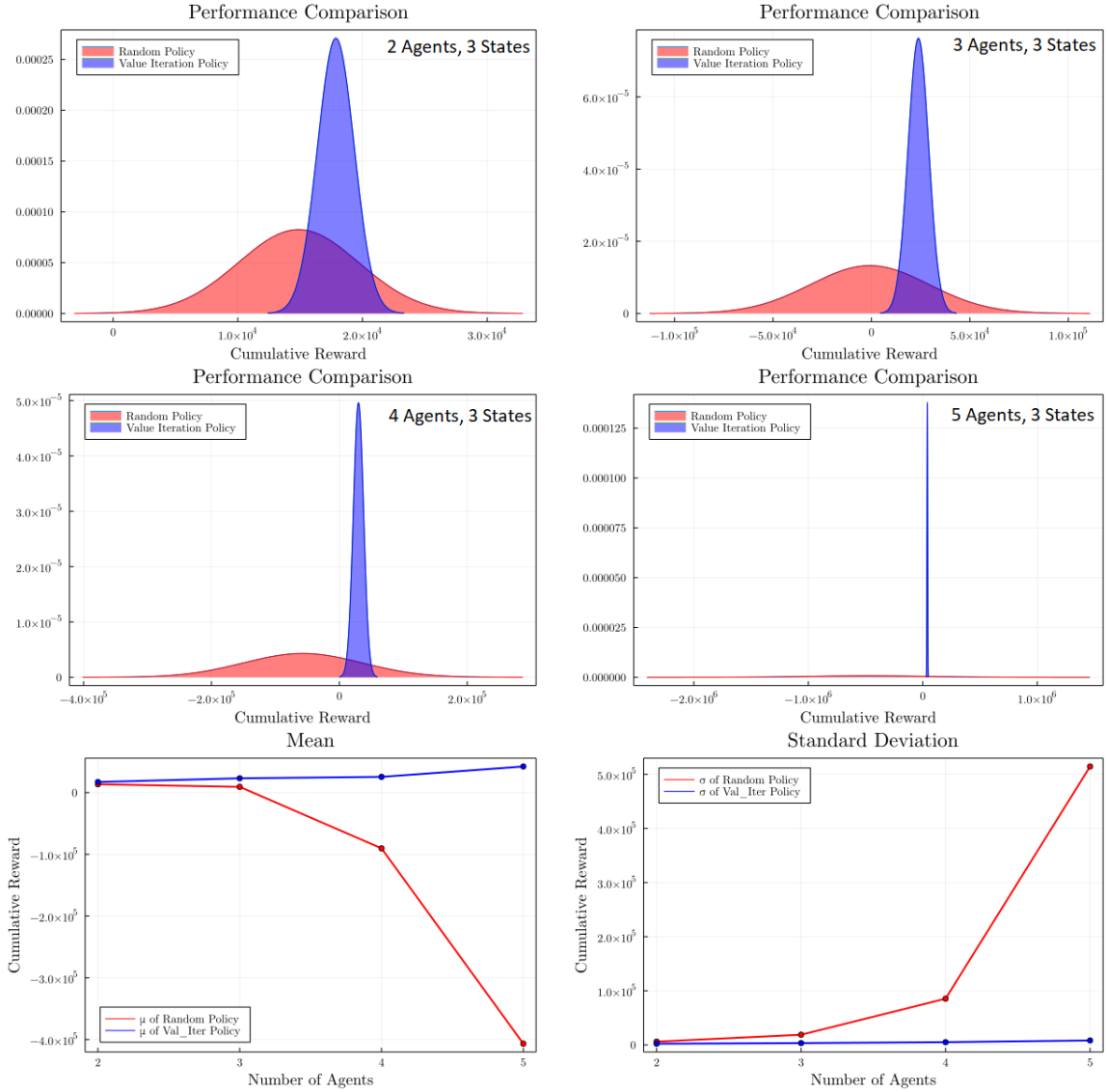

Fig. 2 Gaussian Distributions of Cumulative Rewards for 3 Opinions and 2 - 5 Agents

A similar trend is observed in the case of changing states from 2 to 5 keeping the number of agents as 3. Fig. 3 shows the corresponding variations of mean and standard deviations of cumulative rewards for different number of states and 3 agents. In a realistic scenario, the number of agents would be very large, while the number of opinions will be significantly smaller.

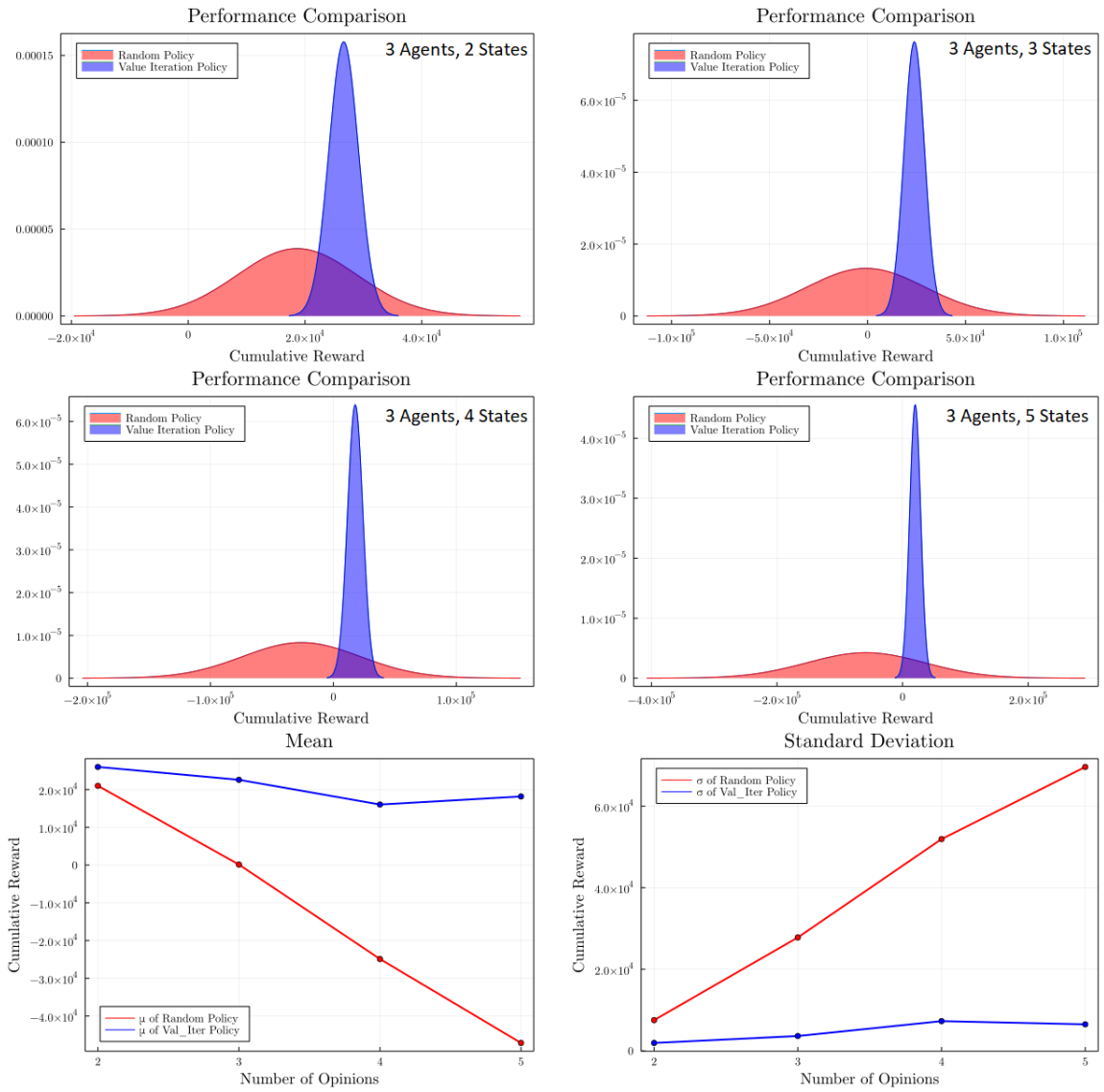

Fig. 3 Gaussian Distributions of Cumulative Rewards for 3 Agents and 2 - 5 Opinions

IV. Conclusions and Future Work

In this paper, we modelled opinion formation dynamics in a multi-agent swarm using a Markov Decision Process (MDP) to model the system followed by value iteration to compute the optimal policy. The results from our policy were benchmarked against random policies and the comparisons were discussed in the previous section. In conclusion, value iteration with an MDP-based modelling results in an exact optimal policy evaluation which accumulates high cumulative rewards relative to a random policy. The relative cumulative reward accumulated from value iteration increases with an increase in number of agents and number of opinions.

As for the future work, although our simulation yields an exact solution for our modelling using value iteration, the trade-off is the large size of the state space. With an increase in population size, the size of the state space increases exponentially and for large populations this might become computationally intensive. An alternative to dealing with large population sizes is either an MMDP formulation or a graph-based MDP approach (where each node in the graph is an MDP). This would deal with every agent in the population independently and thereby bring down the size of the state space. In case of MMDPs, we will need to define joint policies which will be a tuple of individual policies corresponding to each agent and associated to these join policies would be joint rewards and utilities.

In the problem we are dealing with, possible candidates for observations would either be the vibe of an interaction among agents, or perhaps an observation of other agents' personalities, or even prior knowledge about other agents' belief systems. Incorporating observations in our combined state and action space model would not accommodate independent belief systems of each agent. In order to retain independence in beliefs just like independence in decision making, we wish to explore Partially Observable Markov Games (POMGs) as a viable alternative. In a POMG formulation, we define joint policies, joint utilities and joint rewards while maintaining the individuality of all agents.

Appendix

The Source code and the Gif Visualitions for this paper can be accessed from the following GitHub Repository: https://github.com/gnitish18/Consensus_of_Multi-Agent_Systems [15]

Contributions

The members of the team had regular discussions and have equally contributed in the project ideation, problem statement formulation, approach, coding and report writing.

Acknowledgments

The authors would like to thank Prof. Mykel Kochenderfer and the Teaching Assistants of AA228/CS238 - Decision Making under Uncertainty, with special mention to Chelsea Sidrane and Robert Moss for their valuable suggestions and feedback from time to time.

References

[1] Rapoport, A., "Spread of information through a population with socio-structural bias: I. Assumption of transitivity," *The bulletin of mathematical biophysics*, Vol. 15, No. 4, 1953, pp. 523–533.

[2] Bujalski, J., Dwyer, G., Kapitula, T., Le, Q.-N., Malvai, H., Rosenthal-Kay, J., and Ruiter, J., "Consensus and clustering in opinion formation on networks," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, Vol. 376, No. 2117, 2018, p. 20170186.

[3] Olfati-Saber, R., Fax, J. A., and Murray, R. M., "Consensus and Cooperation in Networked Multi-Agent Systems," *Proceedings of the IEEE*, Vol. 95, No. 1, 2007, pp. 215–233. https://doi.org/10.1109/JPROC.2006.887293.

[4] Garofalo, F., Iudice, F. L., and Napoletano, E., "Herding as a consensus problem," *Nonlinear Dynamics*, Vol. 92, No. 1, 2018, pp. 25–32.

[5] Eekhoff, K., "Opinion formation dynamics with contrarians and zealots," *SIAM J. Undergraduate Research Online*, Vol. 12, 2019.

[6] Goffman, W., and Newill, V., "Generalization of epidemic theory," *Nature*, Vol. 204, No. 4955, 1964, pp. 225–228.

[7] Masuda, N., "Voter models with contrarian agents," *Physical Review E*, Vol. 88, No. 5, 2013, p. 052803.

[8] Mobilia, M., Petersen, A., and Redner, S., "On the role of zealotry in the voter model," *Journal of Statistical Mechanics: Theory and Experiment*, Vol. 2007, No. 08, 2007, p. P08029.

[9] Golub, B., and Jackson, M. O., "Naive learning in social networks and the wisdom of crowds," *American Economic Journal: Microeconomics*, Vol. 2, No. 1, 2010, pp. 112–49.

[10] Galam, S., and Moscovici, S., "Towards a theory of collective phenomena: Consensus and attitude changes in groups," *European Journal of Social Psychology*, Vol. 21, No. 1, 1991, pp. 49–74.

[11] Ho, C., Kochenderfer, M. J., Mehta, V., and Caceres, R. S., "Control of epidemics on graphs," *2015 54th IEEE Conference on Decision and Control (CDC)*, 2015, pp. 4202–4207. https://doi.org/10.1109/CDC.2015.7402874.

[12] Boutilier, C., "Planning, learning and coordination in multiagent decision processes," *TARK*, Vol. 96, Citeseer, 1996, pp. 195–210.

[13] Zhang, K., Yang, Z., Liu, H., Zhang, T., and Basar, T., "Fully decentralized multi-agent reinforcement learning with networked agents," *International Conference on Machine Learning*, PMLR, 2018, pp. 5872–5881.

[14] "POMDP.jl," GitHub Repository, 2021. URL https://github.com/JuliaPOMDP/POMDPs.jl.

[15] "Consensus of Multi-Agent Systems," GitHub Repository, 2021. URL https://github.com/gnitish18/Consensus_of_Multi-Agent_Systems.