Social Reinforcement For Collective Decision-Making Over Time

Marco A. Montes de Oca*
Department of Mathematical Sciences
University of Delaware
Newark, DE 19716
mmontes@math.udel.edu

Abstract

Social interactions underpin collective decision making in all animal societies. However, the actual mechanisms that animals use to achieve a collective decision differ among species. For example, ants use pheromones to bias the decisions of other ants; birds observe and match the velocity of their neighbors; humans match the speed of other people while driving (even above speed limits). Despite the differences among these (and other) collective decision-making mechanisms, some basic principles underlying them exist, like the tendency to conform to the actions or opinions of others. In our examples, by following pheromone trails ants effectively follow on their nestmates' steps, birds in a flock are more likely to fly in the same direction, and we humans, while we do not always agree, we do not like to be in permanent conflict with others and eventually seek ways to compromise. Therefore, this principle's basic operating mechanism is that an individual who is exposed to the actions or opinions of others tends to perform the actions, or have the same opinions of the observed individuals.

In this communication, I describe two basic collective decision-making mechanisms based on the principle outlined above. These mechanisms are tested in a setting that simulates a robotics scenario in which a group of robots must collectively find the shorter of two alternative paths between two areas without measuring travel times or distances. First, I describe a mechanism that consists of robots forming teams of three robots (or a greater odd-number of robots) that decide which path to use by locally using the majority rule. Then, I describe a mechanism that consists in individual robots increasing the tendency to choose either path based on the path recently taken by another robot. In both cases, the group collectively chooses the shorter path with high probability. These mechanisms have been proposed as swarm intelligence mechanisms for optimal collective decision-making.

Keywords: Social Reinforcement, Collective Decision Making, Swarm Intelligence

*Homepage: http://www.math.udel.edu/~mmontes/

1 Introduction

There are many animal species whose members gather in large numbers and make important decisions as a group, that is, they make collective decisions [1]. We distinguish between two kinds of collective decisions: those that are characterized by the decentralization of the process through which individuals reach a decision, and those that involve a step of information centralization, like voting. In this abstract, we focus only on the first kind.

Examples of collective decisions exist across species of very different complexity levels, from insects to humans. A well-known example of a collective decision in insects is the selection of the shorter of two paths between two areas by ants [3]. In humans, an example of a decentralized collective decision is the agreement on the cruise speed on a highway (sometimes different than the established speed limit).

The mechanisms that make collective decision-making possible are very different, for instance, ants use pheromones, bees use waggle dances, humans may use language. Yet, despite the differences, there are some basic principles that underlie these mechanisms, such as the apparent tendency to do or to agree with what others do or think. There is even an adage suggesting that doing what others do is preferable to try and err: "When in Rome, do as the Romans do." In fact, it has been shown that copying others is beneficial in a wide range of environmental conditions [8]. Thus, we believe that it is worthwhile to explore the idea of exploiting the information contained in the actions and communications of others for collective decision-making.

Accordingly, we proposed two models of collective decision-making and tested them in a robotics application [6, 7]. Here, we describe these two models and discuss some of their features from a reinforcement learning point of view. In particular, we highlight the role that the group plays in reinforcing a particular behavior on individuals, what we refer to as *social reinforcement*, that gives the group the ability to make an optimal decision.

2 Models

Our first model is based on an opinion dynamics model that uses the majority rule to integrate the information that different individuals possess [5]. In this model, a population of agents each of which can assume one of two states (also called opinions) evolves as follows: First, a group of three randomly chosen agents is formed. Then, the state of the majority within the group is determined. Finally, all the members of the group adopt the state of he majority (thus, each time a group is formed, a maximum of one agent changes state) and the process is repeated. Regardless of the initial conditions, all the participating agents end up having the same state, that is, this model produces consensus on one of the two available states. The final state depends on the initial density of states: if more than 50% of the agents have one of the two states (say A, for example), then all the agents end up with state A. If exactly 50% of the agents are in state A, then the final state is either A or B with 0.5 probability.

We saw the potential of the dynamics of Krapivsky and Redner's model [5] as a collective decision-making mechanism and tried it in a robotics scenario [6]. However, a robotics application imposes a few constraints that have a big impact on the system's dynamics (as shown in Section 3). First, we interpret agents as robots, which gives them embodiment and situatedness, causing the environment to affect the system's dynamics. Second, states are interpreted as actions that the robots have to repeatedly execute while solving a task. Finally, we explicitly include the passage of time associated with action execution. In our model, the duration of an action is finite but stochastic, that is, two executions of the same action take different completion times as is usually the case in real life.

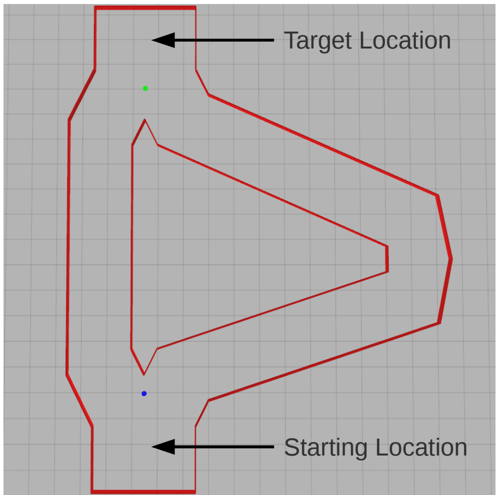
As a test scenario of the system, we used a setting similar to the one used by Goss *et al.* [3] to show that ants can find shortest paths (see Figure 1). The robots' task is to transport heavy objects from one room (starting location) to another (goal location) by forming teams, and going back and forth between these locations. Each time the robots are at the starting location, new teams are formed at random, which allows the mixing of the robots' states.

In our second model, no groups of agents are formed. Instead, individual agents observe the actions performed by other agents. After observation, each agent increases the tendency of performing the observed action. Let the binary variable $X_i \in \{0, 1\}$ to represent an agent i's state. This variable is in turn governed by an internal real-valued variable S_i and a threshold θ. The variable S_i can be thought of as the tendency of agent i to be in one of the two possible states (hereafter, we refer to S_i simply as agent i's tendency). The threshold θ is constant and common to all agents, while S_i is variable and private to each agent.

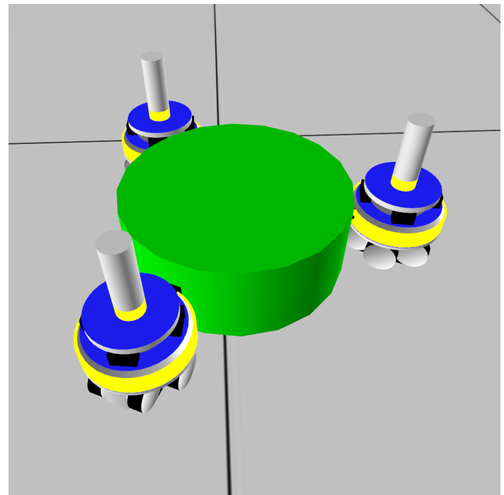
At each time step t of the system's evolution, an agent i might be able to observe the state of another random agent $j \neq i$. When an agent observes the state of another agent, the observing agent updates its tendency as follows:

$$S_i^{t+1} = (1 - \alpha)S_i^t + \alpha X_j^t \,, \tag{1}$$

where $\alpha \in [0, 1]$ determines how much importance is given to the agent's latest observation (X_j^t) as opposed to the agent's accumulated experience (S_i^t). If α is equal to zero, an agent does not change its tendency to imitate other agents; if α is



(a) Environment



(b) Robot Group

Figure 1: Test Scenario. The arena, shown in (a), is a double bridge whose branches differ in length. A group of robots attached to an object is shown in (b). The robots' mission is to transport objects (too heavy for individual robots to move) from the starting to the target location of the arena. The choice robots must make is to take either the longer or the shorter path.

equal to one, an agent copies whatever action another agent performs. After updating its tendency, an agent updates its state as follows:

$$X_i^{t+1} = \begin{cases} 1, & \text{if } S_i^{t+1} > \theta \\ 0, & \text{if } S_i^{t+1} < 1-\theta \\ X_i^t, & \text{if } 1-\theta \leq S_i^{t+1} \leq \theta\,, \end{cases} \quad (2)$$

where $\theta \geq \frac{1}{2}$ due to the symmetry of the actual threshold value that triggers the adoption of one or another state. Thus, an agent's state is a function of its tendency and the threshold θ.

Note how the rule that each individual uses to determine whether to perform a commonly observed action is similar to the basic exponential smoothing equation used for data filtering and time series forecasting [2, 4]. For this reason, we refer to this model as the exponential smoothing model.

3 Simulations

We performed a number of simulations in order to observe the models' dynamics. A summary of the setups and results is presented next.

Majority Rule Model. The duration of actions associated with states A and B are modeled as two normally distributed random variables with means μ_A and μ_B, and standard deviations σ_A and σ_B, respectively. Their ratio $r = \mu_B/\mu_A$ gives a measure of the difference between action execution times and is referred to as latency ratio. Figure 2 shows the dynamics of the majority rule model with a population of 900 agents and 200 teams. When $r \neq 1$, the system achieves consensus on the action of shorter duration even if initially only a minority has a state associated with it. This is seen by the lower critical initial fraction at which the probability of consensus on state A increases from practically zero to almost one.

Exponential Smoothing Model. As in the previous experiments, action durations are normally distributed. In Figure 3, we show the evolution of the average tendency in a population of 100 agents with $r = 2$. When the average duration of the actions associated with each of the two states is equal ($r = 1$) and $S_i^0 = 0.5$, that is, the population of agents will reach a consensus on any of the two states with equal probability (case not shown).

When $r = 2$, the duration of the actions is sufficiently different to induce a strong bias toward the state associated with the action with shorter average execution time even for $\alpha = 0.5$. However, a large value of α means that agents copy any observed state, which produces agents to switch too fast which leads the system to consensus on any of the two states with equal probability.

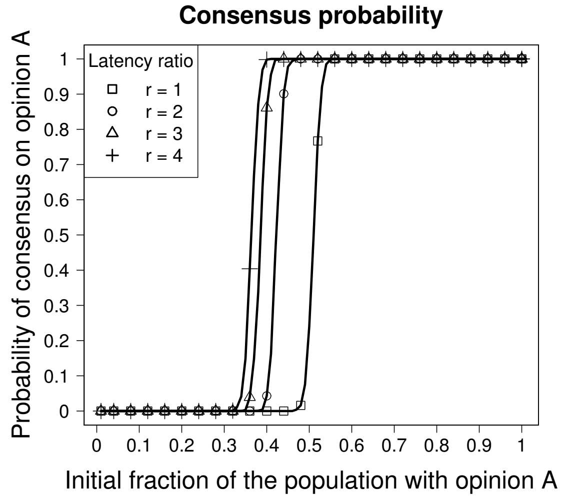


Figure 2: Dynamics of the majority-rule opinion formation model with normally distributed action durations with a population of 900 agents. Results obtained through 1,000 independent runs of a Monte Carlo simulation.

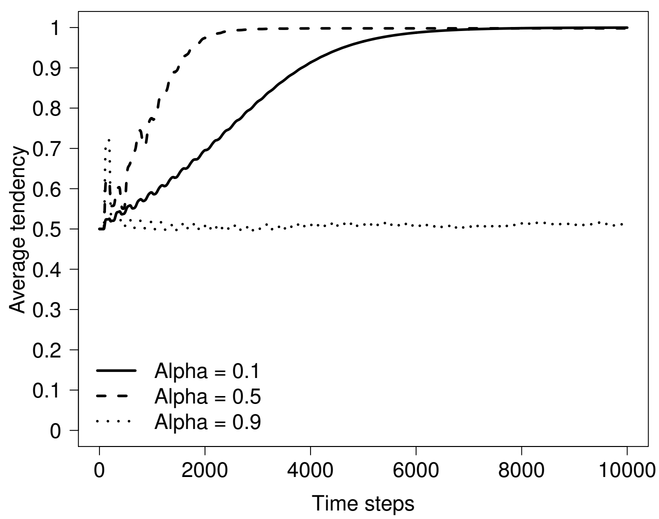


Figure 3: Average tendency over the whole population (100 agents). In these simulations, $\mu_A = 100$, $\sigma_A = \sigma_B = 10$ time steps. The acceptance threshold θ is equal to 0.6 in all cases. Averages obtained through 500 independent runs of a Monte Carlo Simulation.

4 Social Reinforcement

The rules that govern the behavior of the agents in the models presented in the previous section encode the idea of social reinforcement which we understand as the indirect transfer of environmental feedback to a focal individual through the behavior of others. In the following, we discuss how social reinforcement occurs in each of the presented models and its role in the collective decision process.

4.1 Many-to-One Reinforcement

In our first model, robots form groups of three members before executing an action. The action chosen by the group corresponds to the one advocated by the local majority. A robot with a state different from that of the majority is forced to discard it and adopt the state of its peers. For this agent, the state of the majority represents the social reinforcement because the fact that two other agents share the same state represents information beyond the actual state value. In our setting, since the agents that choose the shorter path return to the starting location more often than the agents that choose the longer path, it is more likely to form groups whose majority is associated with the shorter path. Therefore, the length of the path is encoded in the probability of forming groups of robots with two or more robots advocating for the shorter path. In other words, the state of the environment is contained in the group and therefore, environmental feedback occurs only socially because agents do not measure time or lengths.

4.2 One-to-One Reinforcement

In this model, the signal used to update an agent's tendency is the observed behavior of another agent. If the frequency with which a certain behavior is increased, the the observing agent will increase its tendency to perform that behavior. This is in line with traditional Hebbian learning where repeated stimulation increases the strength of the association between action and stimulus. As with the majority model, if the frequency of observation is linked with environmental interactions, then social reinforcement becomes an indirect feedback channel between an agent and the environment.

5 Conclusions

Animals that form groups to address problems that are too difficult, or even impossible, for individuals to solve have to agree with their peers which actions to take. Upon agreement, we can say that a collective decision has been made. These decisions range from travel directions, to task allocation, to shelter selection. It is important to note that when collective decisions are made, the group is the problem solver, not the isolated individuals. Some researchers call this phenomenon *swarm intelligence*. The models presented in this communication are loosely inspired by natural phenomena, but do not model any real collective decision making mechanism used by animals (at least to best of our knowledge). This not to say that these models do not provide any insights into the dynamics of collective decision making. On the contrary, by reproducing the collective decision making ability of ants without simulating pheromones, our results show that there is a clear distinction between mechanisms and principles and that different mechanisms using the same principles can lead to the same collective-level behavior.

The models presented here illustrate the idea of social reinforcement, that is, the indirect transfer of environmental feedback to a focal individual through the behavior of others. In the majority rule model, social reinforcement occurs through the adoption of the state held by the local majority of the groups formed by robots. In the exponential smoothing model, social reinforcement occurs when a focal agent observes the action of another agent and updates its tendency accordingly. In both cases, when the agents' states are associated with actions with similar outcomes but that take time to perform, the population reaches a consensus on the action that takes less time to perform. In the test scenario used in our experiments, this result is translated into a group selecting the shorter of two paths between two locations without directly measuring time or distance. In these experiments, social reinforcement filters out suboptimal actions.

References

[1] L. Conradt and T. J. Roper. Consensus decision making in animals. *Trends in Ecology & Evolution*, 20(8):449–456, 2005.

[2] E. S. Gardner Jr. Exponential smoothing: The state of the art–Part II. *International Journal of Forecasting*, 22(4):637–666, 2006.

[3] S. Goss, S. Aron, J.-L. Deneubourg, and J.-M. Pasteels. Self-organized Shortcuts in the Argentine Ant. *Naturwissenschaften*, 76(12):579–581, 1989.

[4] R. Hyndman, A. Koehler, K. Ord, and R. Snyder. *Forecasting with Exponential Smoothing. The State Space Approach*. Springer, Berlin, Germany, 2008.

[5] P. L. Krapivsky and S. Redner. Dynamics of Majority Rule in Two-State Interacting Spin Systems. *Physical Review Letters*, 90(23):238701, 4 pages, 2003.

[6] M. A. Montes de Oca, E. Ferrante, A. Scheidler, C. Pinciroli, M. Birattari, and M. Dorigo. Majority-Rule Opinion Dynamics with Differential Latency: A Mechanism for Self-Organized Collective Decision-Making. *Swarm Intelligence*, 5(3-4):305–327, 2011.

[7] M. A. Montes de Oca, E. Ferrante, A. Scheidler, and L. F. Rossi. Binary Consensus via Exponential Smoothing. In R. Colbaugh et al., editors, *Proceedings of the Second International Conference on Complex Sciences: Theory and Applications (COMPLEX 2012)*. 2012. To appear.

[8] L. Rendell, R. Boyd, D. Cownden, M. Enquist, K. Eriksson, M. W. Feldman, L. Fogarty, S. Ghirlanda, T. Lillicrap, and K. N. Laland. Why Copy Others? Insights from the Social Learning Strategies Tournament. *Science*, 328(5975):208 – 213, 2010.