Engineering Social Learning Mechanisms for Minimalistic Multi-agent Robots

Owais Hamid[1]
[1]*Systems Design Engineering*
University of Waterloo
Waterloo, ON, Canada
omabdulh@uwaterloo.ca

Kerstin Dautenhahn[2,1]
[2]*Electrical and Computer Engineering*
University of Waterloo
Waterloo, ON, Canada
kerstin.dautenhahn@uwaterloo.ca

Chrystopher L. Nehaniv[1,2]
Systems Design Engineering
University of Waterloo
Waterloo, ON, Canada
cnehaniv@uwaterloo.ca

2020 3rd International Conference on Control and Robots (ICCR) | 978-1-7281-7562-1/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICCR51572.2020.9344158

***Abstract*—Social learning, which can include complex or simple social mechanisms, allow us to understand cooperation and communication in animals, giving them better chances to survive for longer and thrive as a society. In order to translate this understanding into socially rich behavior among multi-agent robots, this study utilizes social learning mechanisms that are simple yet quite effective. Small and simple swarm robots are utilized to understand how such social mechanisms might play a role in establishing rules for emergent group behavior and how social rules might be engineered to gain useful effects in a group of robots. The study investigates exploratory behavior without interaction (asocial) and with interaction (social) among a group of robots. The results from this exploratory study suggest that deterministic asocial exploration is best performed by Spiral exploration mechanisms. However, these asocial exploration strategies are eclipsed by certain types of social reward sharing strategies as long as sharing occurs for at least half the lifetime of the robots. Sharing locations of reward caches for all time is of course the most optimal, but comes at the cost of communicating longer and hence using more energy both on the sender and receiver's end. An analysis of a compromise strategy between completely asocial exploration and social reward location sharing is performed using strategies termed critical and conditional learning. It is found that the number of reward caches located through critical and conditional learning are intermediary to the two extremes, namely completely asocial and completely social foraging.**

***Index Terms*—Multi-Agent Robotics, Social Learning, Social Robotics, Local Enhancement**

I. Introduction

Social Learning has been shown to be a beneficial enrichment and a powerful source for acquiring new skills or new information in many animal species. The advantages of social learning for animals or humans are varied and applicable to different circumstances. Similarly, roboticists have argued that machines or robots capable of social learning (learning from humans or other robots) can take advantage of lessons learned earlier and hence will be much more robust to changes in environment and survival [1].

Social learning can be attributed to various mechanisms of survival, allowing animal societies to thrive and providing a medium of information and skill exchange. However, social learning in and of itself, is not always beneficial [2] and previously acquired skills or information through social learning may or may not be relevant in a particular situation. Furthermore, principles of social learning in biological species may or may not be applicable to multi-agent robotics. Thus, it is necessary to investigate the conditions under which social learning among robots proves to be helpful.

The remainder of this article is structured as follows. Section II discusses related literature on animal studies and relevant definitions. Related literature in robotics is discussed in Section III. Section IV states the research questions and hypotheses. The experimental methodology and procedure is discussed in Section V, followed by results (Section VI). The statistical analysis of the results are discussed in section VI-C. Section VII discusses the experimental outcomes with respect to the initial research questions. Conclusions, limitations of the study and future work follow in Section VIII.

II. Asocial and Social Exploration in Nature

There are both advantages and disadvantages to social learning. Many negative consequences of social learning have been highlighted by Giraldeau [2], one important instance of this being that once information is outdated due to a change in the environment, a socially propagated piece of information may become useless. It has been argued that asocial exploration is the only type that samples the environment and produces information. Thus, scroungers, those who copy[1] rather than produce information, only perform better than producers when fellow scroungers are rare [3].

Anthropologist Alan Rogers coined what is now known as Rogers' paradox [4], the observation that a social learner's fitness would be at most the same as an asocial learner's, conflicting with the common assumption that culture always enhances fitness, ecological success and population growth.

A. When to Learn

However, the situation is more complex. It was pointed out that a strict dichotomy of copying or learning asocially does not exist realistically, rather social learners can create information too [4]. In fact, an important distinction was made regarding *when* information is copied, distinguishing between critical and conditional social learning. Individuals that learn *critically* will first try social learning but, if the learned solution proves of little value or needs perfection, will switch

[1]We assume the word 'copy' in the context of this paper to mean any type of social learning, but refer to [3] for a discussion of different types of social learning

to asocial learning. Its opposite, *conditional* social learning refers to strategies where an individual learns asocially first, then moves and switches to social learning if faced with failure [3].

Critical social learning has been called an Evolutionarily Stable Strategy (ESS) [5]. This concept has high value when asocial exploration is too costly or a solution is too complex to be discovered by a single agent [3]. There are exceptions, either where the environment is highly variable, or when the cost of social learning is higher than asocial learning. Within a highly variable environment, conditional social learning might be beneficial because the environment is sampled first [6], which could lead to obtaining more up to date information.

B. Strategic Learning: Social Exploration

Social learning has been called strategic in nature [3]. A lot of research in the social learning literature has asked questions about why social learning happens; who is copied, when copying happens and where it happens. Copying can be unbiased and random, but more often has a reason, such as context or content. Context has been studied by far the most, and may be derived from a variety of factors [3].

- The animals might be in a state of uncertainty (thus, copy if uncertain), if individually acquired information is outdated or not satisfactory.
- An agent can copy a strategy (of foraging) that the majority conforms to, or if demonstrators are consistent. Thus, copying could be dependent on the frequency of the demonstrators using a specific strategy.
- Demonstrators would be preferred if they were familiar, similar in age, of a certain gender or kin, or if they performed actions that resulted in better payoff or if they were known to be successful in the society.

C. Local Enhancement

Social learning often gets reduced to one very specific type that has been studied widely in humans, other animals and robots, i.e. imitation [7]. Imitation is the most complex form of social learning. There are, however, several other simpler types of social learning mechanisms that play a very important role in communicating strategic information that helps societies survive and thrive.

Local Enhancement is one such mechanism and has been identified as the social mechanism at play in a large number of studies on animal behavior [3] [8].

As defined by Hoppitt and Laland, *Local Enhancement* occurs when a demonstrator interacts with an object at a certain location, and the chances of an observer visiting the same location to interact with objects at that location increases[2] [3]. In a very basic manner, it has also been defined as the simple redirection of attention towards the location where a demonstrator is interacting with an object of interest [3]. An example of local enhancement are wild Trinidadian guppies [8] who are guiding their conspecifics to a particular location in a net that has a hole in it, in order to facilitate the escape of others from a net [8].

[2]For the purpose of this paper the definition of local enhancement by used by Hoppitt and Laland [3] is utilized.

III. Exploration in Robotics

Several other articles have previously explored very relevant questions regarding the behavior of a swarm of robots foraging under various conditions, and these will be reviewed to some extent here.

A. Asocial Exploration

A number of research studies investigated cooperation for foraging in static environments [9] [10]. Such work often focuses on multiple robots that do not interact or communicate with each other (we use the term 'asocial' in this context) with fixed foraging sites that involve robots searching and bringing back such rewards to a base, such as in [11].

Spiral search has proven to be a highly desirable optimal solution for a single agent foraging and searching [9]. An extension of such search algorithms to multiple agents was used in [10] to compare an ant foraging strategy with the spiral based deterministic search.

B. Co-operative Foraging

Foraging techniques based on animal societies have been the focus of a number of robotics applications. Ant inspired foraging techniques involving pheromones, its dispersion and evaporation have been studied through actual chemical trails tracked by actual chemical sensors [12], vision sensor based pheromones [13], or virtual pheromone trails stored on the software server side [14]. Bee inspired foraging techniques have also been explored, one of which being relevant to this work helps robots converge to an area of interest [15].

Local Enhancement is not by any means a new concept in swarm robotics research. Previous work such as [16] utilizes what is called *recruitment*, i.e. sharing information about the location of a foraging site through direct communication, and bringing another idle robot to complete the same task that the initial agent was assigned[3].

It is quite probable that social learning might be incorporated simply as a 'mode' of operation that only gets activated when the agent autonomously determines that its best chance at foraging or survival might be the use of social learning.

Some related previous work has been done on this subject with a parallel research question, i.e. how can a social 'mode' be programmed where newer agents can take advantage of the experience of older agents using simpler mechanisms of social learning. In this case, stimulus enhancement and emulation, which are two other simpler forms of social learning, were used [7].

Noble and Frank [7] performed a simulation study which purported to use simpler (than imitation) social learning mechanisms for an agent to learn tool use to exploit resources. The resource density was very high, with 80% of the search space filled with resources. An agent could take one of 12 actions in

[3]Again, we use 'social learning' for robots in this paper to mean the same, i.e. communication of reward cache locations with other robots also searching for reward caches

each time step and the actions were fed into a reinforcement learning algorithm that chose the next step with the maximum likelihood of a reward. Agents could travel to one of 8 block neighborhood locations. Agents died every 400 time steps and new ones were born.

Three types of social behaviors: Emulation, Contagious behavior or Following (defined as stimulus enhancement) were possible. Following means an observer can follow the demonstrator for the first 25% of the lifetime, contagious behavior meant a 10% probability the observer would do the same as a demonstrator, emulation meant that if a demonstrator obtained a non zero payoff, the observer would know there is payoff present, but does not know how to exploit it, and finally, imitation meant that the observer could take the demonstrator's perspective if the payoff was non zero.

As expected, social mechanisms led agents to perform better than random moves. The best payoff came from emulation in this environment. Mean payoff also increased with an observer following the demonstrator for the first 25% of its life. This result is important because it confirms certain observations made regarding critical social learning, namely that under relatively stable[4] conditions, critical social learning is preferred, as the cost of individual learning is higher in comparison [3]. The results from the study also showed that it is important to imitate or emulate the right models, otherwise payoff decreases.

IV. Research Questions

Our research questions were derived from the previously discussed literature. On a fundamental level, it is intriguing to ask what strategy of copying benefits a society and an individual the most.

From a higher level perspective, it is also interesting to try to understand what sort of social learning parameters can optimize foraging and other activities a society undertakes. These parameters include, but are not limited to, the number of social learners, when social learning happens and under what circumstances it is successful.

Given the understanding we have regarding social learning in general, and local enhancement in particular, the research questions posed are as such:

1) How does the performance of social exploration in robots compare to asocial exploration under time constraints?
2) How does the performance of group exploration in social and asocial exploration scenarios compare under conditions of varying amount of rewards?
3) What are the effects of learning socially first (critical learning) as compared to learning socially later (conditional learning)?

Firstly, it is hypothesized (based on social learning theory) that a simple social learning mechanism, when infused in a multi-agent robot society, performs better than simple asocial explorers.

[4]Stable refers to a stable environment, i.e. one where reward locations stay the same and are not perturbed

Another hypothesis is that critical social learning performs better than conditional social learning in a stable environment (where reward locations do not change often) when primary reward density is low.

We have decided to focus this study on *Local Enhancement* for a number of reasons. Firstly, there seems to be an overwhelming focus on imitation as a form of social learning among robotic agents. However, imitation is considered the most complex form of social learning [3] not least because of the high level of cognition (and hence computation) required, but also the need for higher level comprehension and interpretation of body parts and conversion of the demonstrators actions to the agent's own anatomy, what has been termed perceptual opacity by Heyes [17] and conceptualized as the Correspondence problem by Nehaniv, Dautenhahn and Alissandrakis [18]. The focus here is on minimal real time computation for a simple task such as foraging or finding resources for survival.

Secondly, foraging efficiently calls for sharing of *locations*. Since searching over large areas in an uncertain environment is difficult, and a task that would be well suited to a large number of robots with limited computational intelligence, quickly calling attention to an agent's location fits very well with the characteristics shown by local enhancement in animal species.

The goals and contributions of this study are to understand the interplay between asocial exploration and social learning, and to understand the advantages when strategies are switched from one mode to another. We further seek to establish baseline parameters of optimal reward foraging with and without sharing of information for further studies.

One of the other goals of the study was to try to accomplish autonomy for robotic agents using Local Enhancement. The autonomy would be in the context of agents finding reward caches at a location (either asocially or socially) that hold multiple rewards (hence the word cache) which they would consume one at a time as and how they were required to keep themselves powered up.

We envisage the applications of the concepts and results of this study to help search over large areas that are unmapped and unknown by multi-agent robotic systems. The multi-agent system may or may not act socially and would require a degree of autonomy to search for either constrained or low occurring resources, or specific targets in large areas.

V. Methodology and Experimental Procedure

Due to the simplistic nature of social learning being used in our study, simple robots with limited computation capacity were deemed most suitable. While originally Quanser based quadcopters were meant to be utilized for this research, under the current conditions of COVID19, due to inaccessibility of the robots, the online platform Robotarium, physically located at Georgia Tech in Atlanta was utilized as it allows experimentation with real robots with 'jobs' that are submitted online via its platform [19]. The jobs involve submitting python codes, that were tested in simulation on a local machine first, so that the code can run on real robots remotely. Simulation platforms

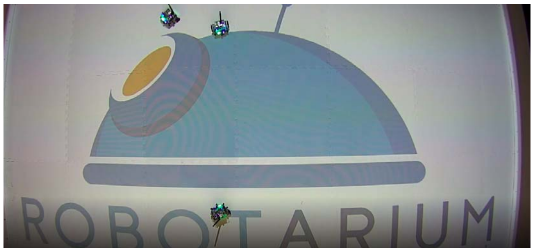


Fig. 1: A still image from an overhead camera of the Robotarium running an experiment with three GritsBots

such as Webots [20] were also considered, but working with real robots was given priority. However, one caveat that came with the Robotarium was its limited availability (only on Fridays), again due to the current COVID19 situation.

While the Robotarium does have a centralized server that controls all robots, every agent has separate controllers, and all features added on to a central program (in python) were executed as if they run on separate robots, because coding an individual robot's micro-controllers is not allowed.

A team of three robots were used for all experiments. The choice of three robots was made because we understand from the literature that animals often hunt in packs of three [21]. The Robotarium utilizes small robots called the GRITSbots [19]. The robot itself is quite small (measuring 130*90*180 mm in length, width and height). The robots run on ESP8266 microcontrollers operating at 160 MHz. The robots run in an enclosed arena, as shown in Figure 1 which is 3 * 1.8 m in dimension. This represents the total search space available to the robots.

Rewards are simulated on this platform, i.e. they are not physically tangible entities, rather they only exist in computer memory. This is due to the remote nature of the Robotarium which cannot be changed, e.g. we could not introduce objects in the environment.

Reward density refers to the number of total rewards that can be foraged. Six reward densities were studied: [5,10,20,30,40,50] which represent number of reward locations available in the search space.

The robots are capable of detecting a reward if they are within an *observational radius* of 0.03 m of the reward. This passes for sensing on the robot, since there are no actual sensors on the robot.

A. Methodology

For this exploratory study, a simple methodology involving a set of rewards to be foraged was set up, with the details described below.

The research questions posed in Section IV have been broken down into a work flow to systematically investigate the research questions.

The first experiment was designed to collect baseline statistics to determine an approximately optimal approach towards asocial exploration of the search space. Three types of search patterns based on previous research were chosen for the asocial exploration task. The workflow established was as follows:

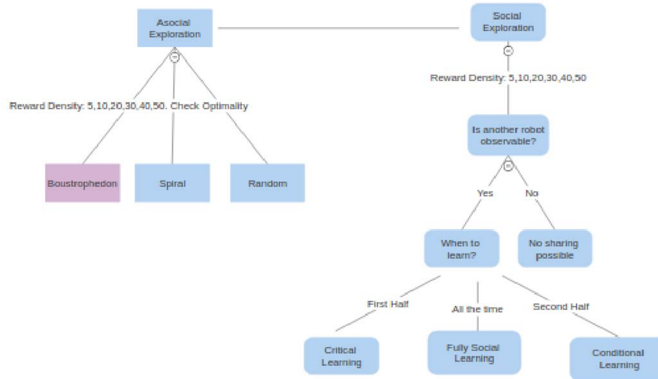


Fig. 2: A flowchart of the experimental methodology that was adopted in our multi-agent robot foraging study using the Robotarium

Establish parameters of success for asocial exploration

- Asocial exploration is deterministic and involves robots exploring the search space in a deterministic fashion with preset patterns.
- Based on previous research [10], the parallel deterministic spiral formation was adapted as one search pattern. A boustrophedon [22] search pattern, and randomized robot behavior (also tested in [22]) were the other two patterns compared for the baseline testing.
- The search space has different numbers of rewards, and a total of six scenarios are developed for each search pattern based on the six reward densities.

Establish rules of sociability in a group of multi-agent robots

- Program a proximity check for *observational radius* to introduce social learning
  - A proximity check refers to checking surroundings to locate other robots that might be in a certain radius. In real life, these represent sensor limits
  - A broadcast mechanism that simply communicates success in finding the reward to any other agent in the proximity
  - The location of a primary reward is stored in an agent's memory such that it can revisit the site when it needs more food (rewards) to survive
- Implementing local enhancement
  - Local Enhancement is the redirection of attention towards a potentially interesting location that may or may not lead to social learning
  - For local enhancement to happen, the engineered solution that we use is to create a matrix of proposed locations of interest. These locations are shared socially using the above mentioned broadcast mechanism.

The lifetime of a robot is defined as follows:

- For asocial exploration, an agent is provided with 6000 iterations (an iteration lasts 0.033s in real time in the

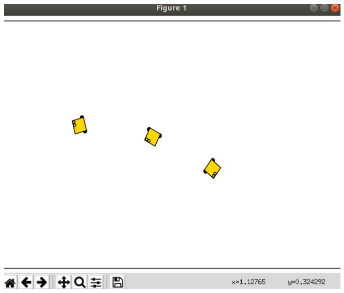


Fig. 3: Simulation environment of Robotarium in python

Robotarium). Any reward location can only be awarded once to the agent. This is to maximize the understanding of how many such caches an agent can detect and forage for itself.
- An agent may explore socially the entire time. This mode is called *fully social* and the agent spends the entire time foraging and scrounging.
- For the *critical social* exploration experimental runs, an agent is given 6000 iterations, the first half of it spent exploring socially. The second half is spent exploring asocially.
- For *conditional social learning*, the agent spends the first half of its lifetime, i.e. the first 3000 iterations exploring asocially, followed by social learning in the second half. All other conditions are the same as critical social learning.

A flowchart was created and it represents the experimental methododology used to investigate the research questions. This can be seen in Figure 2.

B. *Description of software*

Jobs on the Robotarium can be programmed in either Python or Matlab. Python was utilized for these experiments. The experiments need to be validated through simulation first.

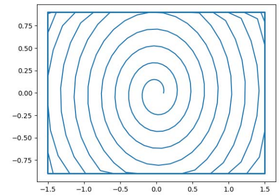
The simulations also provided an estimate of how long the experiment will take in the real Robotarium (runtime). Safety barrier certificates embedded in the code ensure that robots do not collide with each other during the experiment [19].

C. *Experimental Procedure*

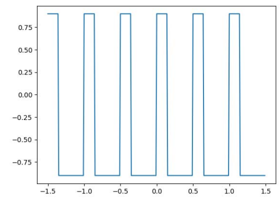
1) Asocial Exploration: In experiment 1 (per Table I), three search patterns were deployed, with random exploration serving as a control condition and both Boustrophedon and Spiral patterns used for comparison in order to decide on the optimal search pattern for the agents in this environment.

Goal points for Boustrophedon and Spiral formations given to the agent to follow are depicted in Figure 4. The random formation was implemented by simply giving an agent a randomized goal point, and, after reaching it, to provide another random goal point.

For asocial exploration, each search pattern was repeated ten times for each reward density at [5,10,20,30,40,50] for significance tests. As an example, for the spiral search pattern at reward density 10, ten repetitions were conducted to observe the behavior of the three robots and how many rewards they



(a) The archimedean spiral pattern used for exploration by all three robots



(b) Boustrophedon, or sawtooth patterns used for exploration

Fig. 4: Patterns of search utilized for the Experiments

TABLE I: Types of Experiments

	Type of Experiment	Conditions	Aim
1	Asocial Exploration	Boustrophedon Spiral Random	Compare which asocial strategy forages/discovers most reward caches
2.1	Social Exploration	Fully Social Critical Conditional	Understand how sharing of rewards improves reward cache location by individual agents
2.2		Local Enhancement	Understand if an agent can successfully return to previously identified foraging site in 400 iterations

This table shows the experiments that were designed, what conditions were present for the experiments, and what their aims are.

accumulate, thus giving us 10*3 (10 repetitions * 3 robots) = 30 data points.

2) Social Learning: **Two experiments** were designed using social learning, as outlined in Table I. In Experiment 2.1, a structure similar to the one used for asocial exploration was utilized while collecting experimental data for social learning with a few modifications.

Fully social tasks were performed with all robots detecting rewards by themselves using the spiral formation (the reason why spiral is used is explained in Section VI-C) along with sharing rewards with robots in the observational radius (set to 0.6 m) when they are discovered.

The same reward densities were used for both critical and conditional social learning. Social learning with an observational radius was allowed for the first 3000 iterations for the critical learning case, and the latter 3000 iterations were used for conditional learning.

Experiment 2.2 of the social learning experiment was to design a 'follow up' 400 iterations worth of added life (added to the 6000 iterations assigned at the start) if an agent successfully found a reward cache (which means it collected a reward), and would then have to go to the most convenient (i.e. closest) reward cache location that it either discovered by itself or was shared with it by another agent.

The identification of the most 'convenient' reward cache is a two stage process. The agent has in its memory the locations where reward caches can be found. These reward caches have been either discovered by the agent itself, or the agent has observed another agent find a reward cache here. These locations of interests are in essence **proposals** of where a robot *may* go back to if it so chooses in order to restock on food. In other words, they are the options the agent has to choose from to go back to re-stock on food so that it can continue foraging for more locations.

Therefore, the first stage is to create these proposals. The second stage is to identify the goal ultimately selected, and this is based on closest distance, i.e. which reward cache is conveniently closest.

A representation of proposals of locations of interest (for the agent), such that the robot could go back and forage from that location, and the goal ultimately selected for local enhancement is given in Figure 7 to provide a better understanding of local enhancement in this context.

VI. Results

A. Asocial Exploration: Spiral, Boustrophedon and Random search patterns

As stated in Section V-A, three patterns of search were employed for asocial exploration for Experiment 1. These are all deterministic in nature and provide a better understanding of which pattern provides an optimal search pattern. These patterns and the results are described in the form of means and standard deviations (error bars) of number of reward caches discovered (foraged) in Figure 5.

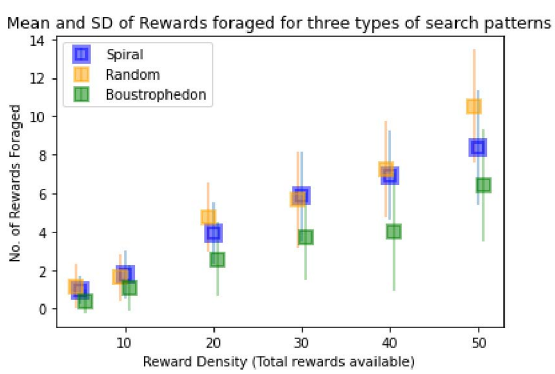


Fig. 5: Means and standard deviations for all robots aggregated together to get a single value (Mean± Std Dev), for different asocial search patterns at each reward density for Experiment 1. Data points were shifted slightly along the horizontal axis to make the standard deviation bars clearer to read.

B. Social Learning

As outlined in Section V-A and specifically in Table I, Experiment 2.1 on social exploration yielded foraging results that were summarized through means and standard deviations for the different reward densities in Figure 6. Since it was observed in the previous asocial exploration experiment that the spiral formation performed best for most reward densities, the performance of spiral exploration has also been included for reference.

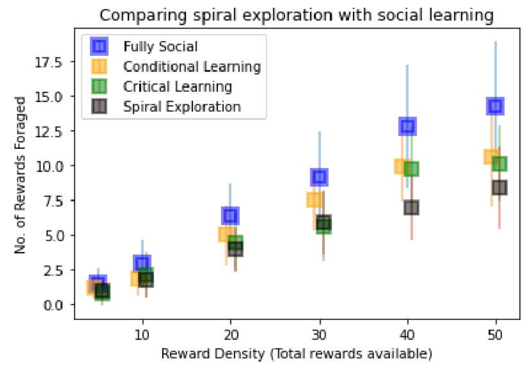


Fig. 6: Comparison of means and standard deviations for reward foraging from three different types of social learning for the first social exploration experiment, i.e. Experiment 2.1. One selected type of asocial exploration (spiral) is also present for comparison. Data points have been shifted slightly to make the standard deviation bars clearly visible.

For the second social exploration experiment, i.e. Experiment 2.2, Figure 8 represents the results of the second experiment for social learning, i.e. a representation of how many agents were able to successfully go back to a location of interest (that they had either discovered themselves or that had been shared socially with them).

If they successfully found or had their attention drawn to at least one reward cache location in the initial 6000 iteration life, the blue bar represents the number that actually reached the location of interest in the next 400 iterations (a time that was deemed sufficient to reach the closest point of interest) against red, i.e. those that did not make it either because they did not discover a location of interest in the initial 6000 iterations, or were unable to reach one in the following 400 iterations. A total of 3 agents over 10 repetitions gives 30 instances of robots attempting to reach the location of interest in this figure.

C. Data analysis

The data collected in the manner described above was subjected to statistical tests. The question of which statistical test to use was aided in part with recommendations given in [23], from which the Analysis of Variance (ANOVA) was determined to be the most suitable. The study was designed with two independent variables: the reward density and type of exploration. Thus, the two-way ANOVA was utilized along with the Tukey's Pair-wise HSD test for post-hoc analysis separately for observations from asocial exploration and exploration with social learning.

Tables II and III describe the Two Way ANOVA and post-hoc pair wise Tukey's HSD for asocial exploration respectively.

Tables IV and V again describe the Two Way ANOVA and post-hoc Tukey's test respectively for socially enabled exploration.

TABLE II: Two way ANOVA for Asocial Exploration

	sum_sq	df	F	PR(F)
C(strategy)	444.4481481	2	48.47849829	4.88E-20
C(rew_density)	3698.209259	5	161.3539235	1.87E-103
C(strategy):C(rew_density)	176.5740741	10	3.851988577	4.73E-05
Residual	2392.833333	522		

Two way ANOVA test results for asocial exploration with Reward Density and Social Exploration strategy (termed rew_density and strategy respectively) as the independent variables. The test results demonstrate that rewards that are foraged depend significantly on both the strategy and reward density, and that there is an interaction effect with $P < 0.05$

TABLE III: Post-hoc Analysis for Asocial Exploration

group1	group2	meandiff	p-adj	lower	upper	reject
Bous/20rew	random/20rew	2.1667	0.0121	0.2282	4.1052	TRUE
Bous/30rew	random/30rew	3.2667	0.001	1.3282	5.2052	TRUE
Bous/30rew	spiral/30rew	2.9333	0.001	0.9948	4.8718	TRUE
Bous/50rew	random/50rew	4.1333	0.001	2.1948	6.0718	TRUE
Bous/50rew	spiral/50rew	1.9667	0.0426	0.0282	3.9052	TRUE
random/50rew	spiral/50rew	-2.1667	0.0121	-4.1052	-0.2282	TRUE

The post-hoc Pair wise Tukey's HSD test was performed for asocial exploration with pairs of all possible combinations being compared against each other. Since there were over 150 pairs, only the most relevant data that differs significantly is shown in this table, i.e. that reject the Null Hypothesis (hence reject = TRUE). The results confirm the hypothesis that Boustrophedon performs worst of the three strategies at reward densities 20, 30, 40 and 50, but while we can observe the means of random and spiral being higher at 5 and 10 reward densities, these differences are not statistically significant (i.e. $P > 0.05$).

TABLE IV: Two way ANOVA for Social Learning

	sum_sq	df	F	PR(F)
C(strategy)	979.4819444	3	51.18205889	6.70E-30
C(rew_density)	9445.090278	5	296.1274597	1.32E-169
C(strategy):C(rew_density)	453.8597222	15	4.743216587	8.71E-09
Residual	4439.833333	696		

Two way ANOVA test results with Reward Density and Social Exploration strategy (termed rew_density and strategy respectively) as the independent variables. The test results demonstrate that rewards that are foraged depend significantly on both the strategy and reward density, and that there is an interaction effect with $P < 0.05$

TABLE V: Post-hoc Analysis for Social Learning

group1	group2	meandiff	p-adj	lower	upper	reject
conditional/40rew	social/40rew	2.9	0.0024	0.5183	5.2817	TRUE
conditional/40rew	spiral/40rew	-2.9333	0.0019	-5.3151	-0.5516	TRUE
conditional/50rew	social/50rew	3.6667	0.001	1.2849	6.0484	TRUE
critical/30rew	social/30rew	3.5667	0.001	1.1849	5.9484	TRUE
critical/40rew	social/40rew	3.1	0.001	0.7183	5.4817	TRUE
critical/40rew	spiral/40rew	-2.7333	0.007	-5.1151	-0.3516	TRUE
critical/50rew	social/50rew	4.1	0.001	1.7183	6.4817	TRUE
social/20rew	spiral/20rew	-2.4	0.0457	-4.7817	-0.0183	TRUE
social/30rew	spiral/30rew	-3.3	0.001	-5.6817	-0.9183	TRUE
social/40rew	spiral/40rew	-5.8333	0.001	-8.2151	-3.4516	TRUE
social/50rew	spiral/50rew	-5.8667	0.001	-8.2484	-3.4849	TRUE

The post-hoc Pair wise Tukey's HSD test was performed for social exploration with pairs of all possible combinations being compared against each other. Since there were over 280 pairs, only the most relevant data that differs significantly is shown in this table, that reject the Null Hypothesis (hence reject = TRUE). The results confirm the hypothesis that fully social strategy performs the best of the three social strategies (significantly better than spiral, which is asocial but used for comparison, at 20,30,40 and 50, conditional learning at 40 and 50, and critical learning at 30, 40 and 50). There is no conclusive evidence that fully social performs significantly better at lower reward densities such as 5 and 10.

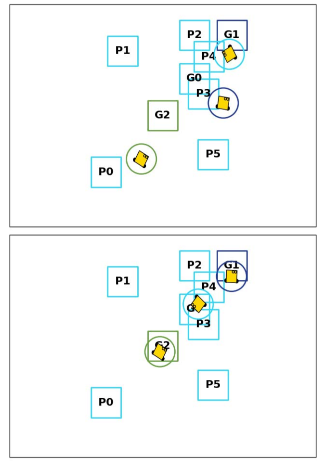


Fig. 7: A demonstration for the second part of the social learning experiment, i.e. Experiment 2.2. The top figure shows Proposals (named P0-P4, in light blue squares) that are available to Robot 1 (with the light blue circle around it). The Robot heads to G0 (its goal) because it has found the locally enhanced region to be the most 'interesting' (since it is closest).
The bottom figure shows the agent successfully reaching its goal, and hence successfully realizing local enhancement by investigating a proposal that it found 'interesting'. Note that the other proposals were not investigated since they were not found 'interesting'.

For asocial exploration at higher reward densities, i.e. at 20, 30 and 50 rewards in the environment, it was found that spiral and random patterns outperform Boustrophedon significantly. Similarly, at higher reward densities, fully social exploration outperforms all other search strategies. However, there are no significantly different results for lower reward densities and in terms of statistical significance, the results are inconclusive at those densities.

Likewise, for Experiment 2.1, at higher densities, fully social reward foraging trumps any other strategy. For most higher reward densities, social learning of any kind (critical or conditional) also trumps the best of asocial exploration strategies (i.e. spiral). At lower densities, there is no significant difference between these strategies, although looking at the Mean values, it can be said that social strategies out perform asocial ones and fully social learning is still the most optimal.

VII. Discussion of Research Questions and Outcomes

It is important to put the results derived in Section VI in perspective with regard to the original research questions (Section IV).

One of our primary concerns were that if 'fully social' learning performs best, as is expected of it, why stop learning socially at any time? Since both critical and conditional learning switch either into social learning or out of it, why stop learning socially at all?

One crucial reason why switching between social and asocial exploration can be important is because in our scenario,

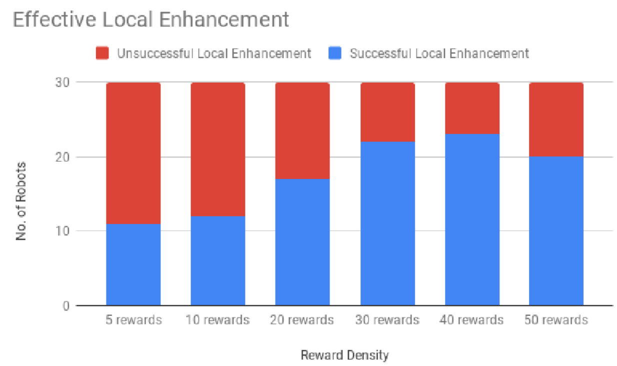


Fig. 8: The bar chart shows how many robots across the varying reward densities were able to successfully locate the closest reward cache location and visit it so as to gather food/energy to forage for the next 400 iterations. This represents the results of the second social exploration experiment (Experiment 2.2) where robots are supposed to return to a reward cache that they have in their memory once they have exceeded the initial 6000 iteration steps (lifetime)

the robotic agent is exploring an environment with other robots but does not know what the reward density of the environment is. Should the reward density be low, it would become advantageous to switch from asocial exploration to social learning.

It is further envisaged that for applications of robots in a large remote environment searching for rewards (which could be resources of some kind, such as minerals), both communication and computational intelligence requirements need to be minimal to conserve battery life. An instinct switching system that decides when switching between social and asocial exploration is optimal could help make these groups of robots be more energy efficient and robust to challenges related to the environment in terms of finding resources.

A comparison between social strategies of reward foraging and asocial types of exploration was conducted with results presented in Section VI-C along with the relevant post-hoc analysis. It confirms that foraging using fully social learning discovers rewards in the most optimal manner. Switching to asocial learning part way through does make an agent discover fewer rewards than learning socially all the time, however it has the potential to save computation that may extend the agent's life time.

An intriguing area of investigation is when the environment has low to medium reward density. In this situation it was observed that a lot of robots are unable to find any rewards that led to them not being able to find a reward cache location which they can revisit should they find themselves running low on rewards. As a matter of fact, we define *frequency* as the number of times an agent was unable to detect a reward cache. Not detecting a reward cache is considered quite debilitating because not only does a robot not identify a food source, it is then unable to continue beyond 6000 iterations to continue foraging in the environment in accordance to our goal of achieving autonomy in the context we presented earlier in Section IV.

For a total of 30 agents, a total of 10 instances were recorded with spiral asocial exploration where no reward was detected. This frequency falls to 6 during fully social learning, but fluctuates wildly for critical and conditional learning at 14 and 4 respectively. It would appear that learning socially first seems detrimental. While animal studies in primatology (which has inspired our research), show clear benefits of social learning, these are cases where social learning is usually not only about location but also about learning skills (using more complex forms of social learning, e.g. imitation or emulation learning). Such complexity is not reflected in our study with simple robots, since they are highly constrained in their skill set.

Utilizing the Mean observed, the number of reward caches located during critical and conditional learning is higher than those gained with spiral exploration which makes those social learning strategies advantageous for lower and medium density environments.

The above discussion is meant to address the first and second research question, i.e. how does asocial and social exploration compare and where do we see the benefits.

Local Enhancement, as defined in [3] involves a brief re-direction of attention that may or may not lead to the visitation of that particular site or locale in the future, depending on how curious and/or motivated an agent is about exploring further resources. The third research question has been addressed with the second social experiment with results presented in Figure 8 and a reference to how it works in Figure 7.

In both social learning experiments, a demonstrator agent provides the location of a potentially important piece of information, the location of a reward cache, to the attention of an observer agent foraging close by, which the observer may or may not utilize in the future. In the second social experiment, a 'location of interest' is defined as the closest reward cache location that the agent can utilize to get a reward. Should the reward be attained, the robot then has a chance to continue foraging until either all 'interesting' reward caches are depleted or it reaches a certain goal defined for it. This is meant to simulate a situation where robot agents find their own resources ('food') that sustains them for a limited period of time beyond their battery life. Foraging while traveling is a very old concept among humans and animals.

Figure 8 shows an interesting trend, especially for lower reward densities (5 and 10) where 11 and 12 respectively out of a total of 30 agents (3 agents per experiment repeated 10 times) were able to reach a location of interest. This indicates a 33-35% success rate, where local enhancement was established according to our definition. This percentage of agents may be expected to continue foraging activities until they exhaust their resources or achieve their objectives of finding resources.

The percentage increases rather linearly from 10 to 30 reward densities, where it remains constant, indicating a saturation effect.

VIII. Conclusions, Limitations and Future Work

A. Conclusion

An exploratory study regarding individual robot exploration as a team, and sharing of rewards (social behavior) among the team was conducted with some significant results.
The study lays down foundations to enable switching behavior between asocial exploration and social learning or vice versa, should a team of foraging robots find themselves in a changing or low reward density environment where social learning might prove crucial. A very crude switching system was introduced to study the effects of social learning employed from the start of a run as opposed to later on. While we found no significant difference between the two, it would appear, based on mean rewards foraged, that it is beneficial to share reward locations later, i.e. to use conditional learning especially for low reward density environments. This seems to contrast the findings in animal societies where social behavior is copied first, and then modified later to suit an individuals' goals [3]. As we speculated before, this might be due to the huge differences between what critical learning and conditional learning means for animal species, who have to survive in their natural ecological environments that they have adapted to in both evolutionary and individual developmental timescales, and how it could be defined and be applied to multi agent robot systems, as used in our study.

B. Limitations

This study has several limitations.

Firstly, it was of interest to study how reward sharing mechanisms translate into survival capacity while there is sharing (or no sharing, or sharing under different conditions) within a group of three agents. Clearly, this concept also needs to be extended to larger numbers of agents to observe how they perform in conditions with different reward densities.

Secondly, the Robotarium has a single server that controls all robots. This precludes research on truly individual autonomously running robots, although in our experiments we simulated individual autonomy of the robots.

Lastly, due to the low availability of runtime in the Robotarium, it was not possible to run more than 10 repetitions for each condition, which results in the current (relatively) low numbers of observations.

C. Future Work

This work was meant to lay the foundation for a research program into engineering social behavior among groups of robots working together. Providing robots with Computational intelligence, while making rapid progress, is still quite limited if robots are to be produced cheaply and en masse, which supports our research direction to investigate simple mechanisms of social learning that could run on relatively simple agents. For future work, we hope to use fully autonomously running robots, as soon as our robotics laboratories re-open.

There are plenty of avenues of further investigations. One pressing work is to devise a much more robust switching mechanism. The methodology used for switching in this study is based on time, i.e. occurring after 3000 iterations. This strategy could be further optimized to be event based instead; the time at which switching takes place itself might be optimized further. The challenge for any agent is to try and estimate the reward density of an environment it is placed in. A distinct switching system such as this may take the form of a Machine Learning (ML) algorithms such as a Naive Bayes classifier or a classification based feedforward neural network. An analysis of how such a switching system helps optimize foraging would be important future work.

There are further concepts from animal behavior that can be incorporated as social behavior in a group of robots. Positive and negative rewards can be used to induce location enhancement and avoidance. The current study focuses only on positive rewards. Avoidance can be induced by introducing history (previous experience) along with the option of sharing such information for negative rewards (punishment), which might create a culture of avoidance in the robot society. This would require ML techniques to take history into account, such as regression trees or Recurrent Neural Networks.

Acknowledgment

This research was undertaken, in part, thanks to funding from the Canada 150 Research Chairs Program. The authors would like to acknowledge support from The Robotarium at Georgia Tech University. We would also like to acknowledge Dr. Moojan Ghafurian for her help with the statistical analysis presented in this paper.

Reproduction of Data

A minimalist example of the python code used to collect data for the asocial exploration part of this paper is open sourced at Owais Hamid's Github Repository.

References

[1] W. De Mulder, S. Bethard, and M.-F. Moens, "A survey on the application of recurrent neural networks to statistical language modeling," *Computer Speech & Language*, vol. 30, no. 1, pp. 61–98, 2015.

[2] L.-A. Giraldeau, T. J. Valone, and J. J. Templeton, "Potential disadvantages of using socially acquired information," *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, vol. 357, no. 1427, pp. 1559–1566, 2002.

[3] W. Hoppitt and K. N. Laland, *Social learning: an introduction to mechanisms, methods, and models*. Princeton University Press, 2013.

[4] M. Enquist, K. Eriksson, and S. Ghirlanda, "Critical social learning: a solution to rogers's paradox of nonadaptive culture," *American Anthropologist*, vol. 109, no. 4, pp. 727–734, 2007.

[5] K. N. Laland, *Darwin's unfinished symphony: How culture made the human mind*. Princeton University Press, 2018.

[6] L. Rendell, L. Fogarty, and K. N. Laland, "Rogers'paradox recast and resolved: Population structure and the evolution of social learning strategies," *Evolution: International Journal of Organic Evolution*, vol. 64, no. 2, pp. 534–548, 2010.

[7] J. Noble and D. W. Franks, "Social learning in a multi-agent system," *Computing and Informatics*, vol. 22, no. 6, pp. 561–574, 2012.

[8] S. M. Reader, J. R. Kendal, and K. N. Laland, "Social learning of foraging sites and escape routes in wild trinidadian guppies," *Animal Behaviour*, vol. 66, no. 4, pp. 729–739, 2003.

[9] H. M. A. Gabal and M. A. A. El-Hadidy, "Optimal searching for a randomly located target in a bounded known region," *International Journal of Computing Science and Mathematics*, vol. 6, no. 4, pp. 392–403, 2015.

[10] G. M. Fricke, J. P. Hecker, A. D. Griego, L. T. Tran, and M. E. Moses, "A distributed deterministic spiral search algorithm for swarms," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 4430–4436.

[11] P. Ulam and T. Balch, “Using optimal foraging models to evaluate learned robotic foraging behavior,” *Adaptive Behavior*, vol. 12, no. 3-4, pp. 213–222, 2004.

[12] R. Fujisawa, S. Dobata, K. Sugawara, and F. Matsuno, “Designing pheromone communication in swarm robotics: Group foraging behavior mediated by chemical substance,” *Swarm Intelligence*, vol. 8, no. 3, pp. 227–246, 2014.

[13] S. Garnier, F. Tache, M. Combe, A. Grimal, and G. Theraulaz, “Alice in pheromone land: An experimental setup for the study of ant-like robots,” in *2007 IEEE Swarm Intelligence Symposium*. IEEE, 2007, pp. 37–44.

[14] N. Hoff, R. Wood, and R. Nagpal, “Distributed colony-level algorithm switching for robot swarm foraging,” in *Distributed Autonomous Robotic Systems*. Springer, 2013, pp. 417–430.

[15] T. Schmickl and H. Hamann, “Beeclust: A swarm algorithm derived from honeybees,” *Bio-inspired computing and communication networks*, pp. 95–137, 2011.

[16] L. Pitonakova, R. Crowder, and S. Bullock, “Information flow principles for plasticity in foraging robot swarms,” *Swarm Intelligence*, vol. 10, no. 1, pp. 33–63, 2016.

[17] C. M. Heyes and E. D. Ray, “What is the significance of imitation in animals?” in *Advances in the Study of Behavior*. Elsevier, 2000, vol. 29, pp. 215–245.

[18] C. L. Nehaniv and K. Dautenhahn, “Like me?-measures of correspondence and imitation,” *Cybernetics & Systems*, vol. 32, no. 1-2, pp. 11–51, 2001.

[19] D. Pickem, P. Glotfelter, L. Wang, M. Mote, A. Ames, E. Feron, and M. Egerstedt, “The robotarium: A remotely accessible swarm robotics research testbed,” in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 1699–1706.

[20] O. Michel, “Webots: Professional mobile robot simulation,” *Journal of Advanced Robotics Systems*, vol. 1, no. 1, pp. 39–42, 2004. [Online]. Available: http://www.ars-journal.com/International-Journal-of-Advanced-Robotic-Systems/Volume-1/39-42.pdf

[21] I. Bailey, J. P. Myatt, and A. M. Wilson, “Group hunting within the carnivora: physiological, cognitive and environmental influences on strategy and cooperation,” *Behavioral Ecology and Sociobiology*, vol. 67, no. 1, pp. 1–17, 2013.

[22] C. Das, A. Becker, and T. Bretl, “Probably approximately correct coverage for robots with uncertainty,” in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2011, pp. 1160–1166.

[23] J. H. McDonald, *Handbook of biological statistics*. sparky house publishing Baltimore, MD, 2009, vol. 2.