Exploring Intervention in Co-Evolving Deliberative Neuro-Evolution with Reflective Governance for the Sustainable Foraging Problem

Aishwaryaprajna[1*] and Peter R. Lewis[1]

[1]Faculty of Business and Information Technology, Ontario Tech University, ON, Canada L1G 0C5
aishwaryaprajna@ontariotechu.ca*

Abstract

Cooperation has been widely studied in multi-agent foraging tasks. However, the impact of agent-environment interactions on the longer term and the achievement of sustainability have been largely unexplored in this context. This work contributes to the development of a testbed for exploring social dynamics between agents: the 'sustainable foraging problem'. This testbed explores the effect of agent behaviour and the agent's dilemma of choosing between individual reward and collective long-term goals for sustainable resource management. To incorporate varied levels of replenishment rates in this testbed, forest, pasture and desert environment types are formulated. A co-evolving deliberative loop with neuro-evolution that asks the agents to act with greedy or moderate behaviour is demonstrated. This deliberative layer is shown to be insufficient in situations of social dilemma where the agents learn to increase their individual rewards instead of collectively increasing these rewards through the sustainability of the environment. A simple reflective governor based on the notion of the agent's self-awareness is illustrated to allow the agents to occasionally reason about the long-term impacts of their immediate actions on future resource availability in the environment, which may eventually ensure sustainability.

Introduction

The decisions and actions of interactive social agents in collective systems affect the agents and the environment directly or indirectly. For the development of cooperative behaviour in multi-agent systems through learning, foraging tasks have been important (Zedadra et al., 2017). Foraging tasks involve terrain exploration, navigation and object identification, manipulation and transport of resources in multi-agent systems and are a canonical problem for investigation of agent-agent interactions (Winfield, 2009). These tasks have often been thought of as social games (Panait and Luke, 2005). There has been extensive research on the evolution of cooperation in game-theoretic contexts with social dilemmas (Axelrod and Hamilton, 1981; Kollock, 1998; Doebeli and Hauert, 2005; Boyd and Richerson, 2009). Developing cooperation may be beneficial in goal-oriented short-term tasks like efficient resource collection and management, but achieving sustainability requires consideration of the impact of the agent-environment interactions over a longer term. The analysis of sustainability of environment with moderate use of resources has been yet largely unexplored in the context of multi-agent foraging tasks.

We are interested in formulating a testbed of experiments that allows us to ask precise questions regarding development of different forms of cognitions of agents for sustainable resource management corresponding to varied levels of resource availability. The achievement of sustainable environment is highly important in real-world issues like climate change.

Recent literature (Dignum and Dignum, 2020; Barnes et al., 2020) has argued that rationality of agents regarding utility functions or goals is not enough to bring cooperation in socially situated multi-agent systems in the absence of social norms. Intentional cooperation and coordination in social systems may be achieved by social self-awareness with capabilities of perception and reasoning about others in the system and the environment (Bellman et al., 2017). Our work builds on this notion and targets the building of reflective cognitive agents that are capable of reasoning about their immediate actions and the corresponding long-term impact on their environment. We hypothesise that computational reflection in foraging agents may support the long-term sustainability of the environment and hence, also of themselves, due to the consideration of the higher-level of goals at each step of learning and decision-making. We illustrate the development of a simple reflective governor loop that is built on the notion of the agent's self-awareness to reason about the connections between the agent's immediate actions and the sustainability of the environment.

To test this hypothesis and evaluate reflective foraging agents, we need a dynamic social environment testbed that captures the features of a complex real-world scenario. This approach of *principled simplification* has been adopted in recent literature on agent-based modelling (Barnes et al., 2022) and is traditionally used in theoretical computer science research where simple combinatorial benchmarks like ONEMAX or KNAPSACK problems are often widely studied instead of harder and complex real-world systems. The basic idea behind this involves creating a simple yet sufficient

problem that retains the key features of the harder problem. In this way, the insights obtained from the simple problem may be generalized to more complex problems.

The 'Sustainable Foraging Problem'

A general framework of different environment types - *Forest*, *Pasture* and *Desert*, corresponding to varied levels of replenishment rates is proposed. The agents' goal is to forage and collect resources for their subsistence in these environments. However, depending on the amount of replenishment, foraging would impact the resource amount and the total population of the agents. We assume that each agent has the choice to adopt either a *greedy* or a *moderate* strategy while foraging. The moderate strategy assumes that the agent only collects resources when it is needed for subsistence. Whereas, the greedy strategy assumes that the agent collects resources at all times irrespective of its needs.

We assume that the total available resources in the environment at any time instance t is defined as r_t. We consider the replenishment rate of the resources to be α.

Let there be n_t agents at time t that are foraging in the environment and collecting resources. Let us assume that there are g_t greedy agents and m_t moderate agents at any time t, where $n_t = g_t + m_t$.

At time t, each agent has an energy of E_t. This is initialized at the first time instance as $E_1 = \eta$. At each iteration, each agent requires at least s amount of energy to survive, and perhaps more if it engages with other activities that also require energy. If $E_t \leq 0$, then the agent dies. Please note that the unit of the resource in the environment and the energy gained by the agent after foraging the resource is equivalent.

We assume that moderate agents only collect resources when their energy levels drop below a certain threshold, τ. Out of m_t moderate agents, m'_t agents need resource immediately at time instance t to escape death. Then $(g_t + m'_t)$ agents would collect maximum possible resource at any time instance t. Then we can say that $(n - g_t - m'_t)$ agents will not collect resource at time instance t, being moderate in nature.

Let us assume that an agent can collect resources as much as possible but not exceeding the amount it can 'carry', a. We assume that the maximum resource an agent can collect is twice the required energy to survive a time step, defined as, $2s = a$.

When the resource in the environment has depleted, and there is not enough resource for the agent(s), the maximum amount of resource an agent can collect can be determined by u_t, which is defined as follows:

$$u_t = \min\left(a, \frac{r_t}{(g_t + m'_t)}\right) \tag{1}$$

The resource collecting agent's energy update at each iteration, capturing the amount of energy foraged, u_t and the amount of energy expended, s can be given by:

$$E_{t+1} = E_t + (u_t - s) \tag{2}$$

The energy update equation for the moderate agent having energy above the threshold τ can be given as:

$$E_{t+1} = E_t - s \tag{3}$$

The reward (payoff) function of the agent at each time step is defined as the logarithmic function of the current energy level of the agent, defined as:

$$\mathcal{R}_t = k \log E_t \tag{4}$$

where, k is a scaling factor.

At each time, the resource depletion may be described as follows:

$$r_{t+1} = \alpha r_t - (g_t + m'_t) \times u_t \tag{5}$$

The presence of variances in initial resource availability r_t and replenishment rate α will give rise to different environment types, which are described as follows.

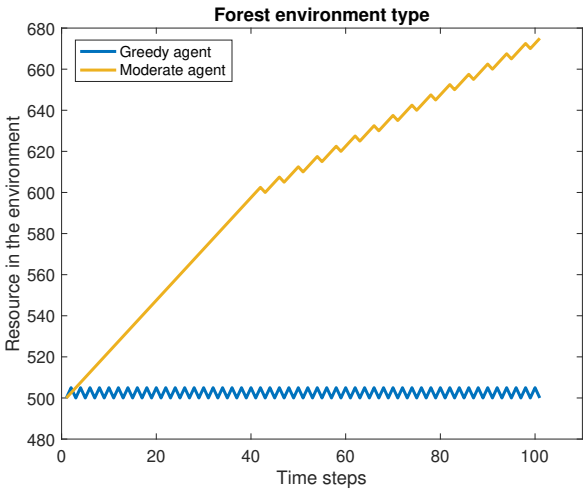

Figure 1: The resource depletion in forest environment for greedy and moderate agent behaviour

Forest:

In this environment type, the initial resource is abundant and the rate of replenishment of resources is very high. The maximum resource an agent can collect in this environment type would always be a irrespective of the agent's behaviour. The cost of exploitation of resources by the agents is low for achieving sustainability of the environment over long term.

When considered in a game-theoretic approach, the exploitation of resources by greedy agents can be incentivized, since this behaviour brings little harm to the environment or

future generations of the social agents. The payoffs/rewards of the moderate agents can be considered to be lesser than greedy agents since they acquire lesser amounts of resources in the long run. At this environment type, agents having the cognitive ability to reflect on their actions would likely offer no additional benefit here, due to an abundance of resources. This can be explained as follows for the resource collecting agents.

The resource depletion for the Forest environment type is described with a single agent in figure 1. Here, the chosen parameters are, $r_1 = 500, k = 100, \alpha = 1.01, E_1 = 100, \tau = 50$ where the following hold:

$$|r_{t+1} - r_t| >> m'_t \times u_t, \forall t \in [0, \infty) \quad (6)$$

$$|r_{t+1} - r_t| >> g_t \times u_t, \forall t \in [0, \infty) \quad (7)$$

This illustrates that the amount of replenishment at each time step is greater or at least equivalent to the maximum foraged amount of resources in this environment type. This shows that neither the greedy nor moderate behaviour of the agent can impact the sustainability of the resources.

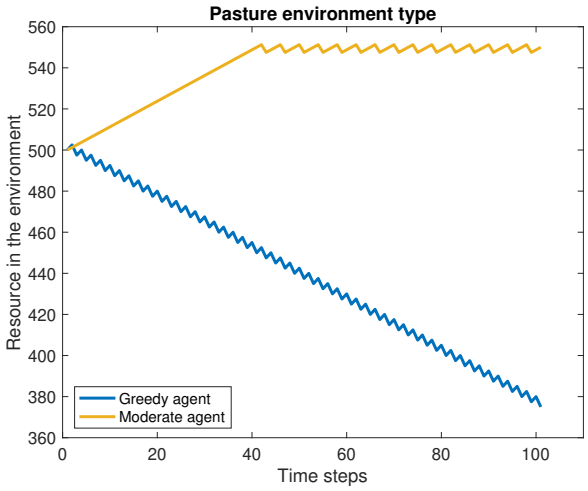

Figure 2: The resource depletion in pasture environment for greedy and moderate agent behaviour

Pasture:

This environment type models the situation when resource replenishes at a specific rate however, over-use of resources could lead to serious consequences to sustainability.

In this environment type, the moderate strategy of foraging would ensure the following:

$$|r_{t+1} - r_t| \geq m'_t \times u_t, \forall t \in [0, \infty) \quad (8)$$

However, the agent's greedy behaviour will lead to the following situation:

$$|r_{t+1} - r_t| < g_t \times u_t, \forall t \in [0, \infty) \quad (9)$$

This means that greedy agents will over-consume and there will be not enough resources for each resource-collecting agent in a multi-agent scenario. With uncontrolled collective exploitation, *Tragedy of Commons* would be imminent (Ostrom, 1999) where greedy strategies would impact the survival of agents. Before reaching this point, it is necessary for the agents to reason and reflect at the state of available resources to avoid imminent doom and potentially allow the resources to replenish with collective non-exploitative behaviour.

The chosen parameters for the Pasture environment type are $r_1 = 500, k = 100, \alpha = 1.005, E_1 = 100, \tau = 50$ illustrated in figure 2. This shows that moderate agents can avoid impacting sustainability in this environment type.

Desert:

When the resource is scarce, either of the considered agent strategies cannot bring sustainability of the environment. In this environment type as there will not be enough resource for any agent who is collecting due to the extremely minimal replenishment rate.

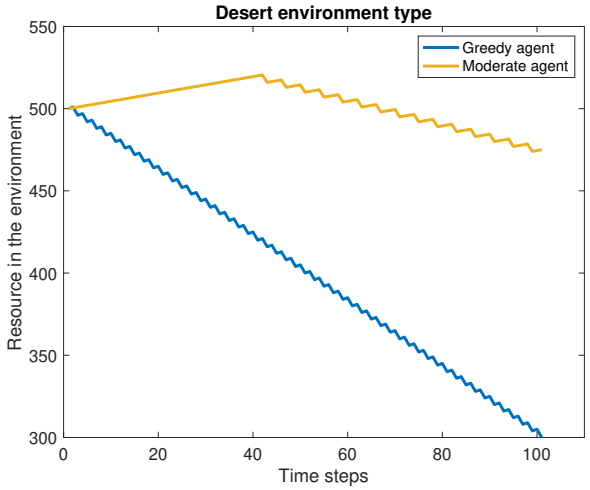

Figure 3: The resource depletion in pasture environment for greedy and moderate agent behaviour

At this stage, both moderate and greedy resource collection would bear a heavy cost to the existing resources which can be described as follows,

$$|r_{t+1} - r_t| < m'_t \times u_t, \forall t \in [0, \infty) \quad (10)$$

$$|r_{t+1} - r_t| < g_t \times u_t, \forall t \in [0, \infty) \quad (11)$$

Foraging in this environment type will give rise to a no-win situation. However, migration of agents to an environment with sufficient resources is a viable option. The impact of agent's behaviour in this environment type is demonstrated

in figure 3 with the parameters chosen as $r_1 = 500, k = 100, \alpha = 1.002, E_1 = 100, \tau = 50$.

We hypothesise that if the agents know which environment type they are in, then they may use meta-models to modify the payoffs in their interest. To ensure this, agents need to have cognitive models generated on the basis of gradients of resource depletion and the past actions of the agents.

Each environment type is considered to be well-mixed so that each agent has access to the 'common pool' of resources at all time instances. Despite the resources being common-pool in nature, the 'sustainable foraging problem' considers a much broader set of problem parameters that may support explorations in the dimensions of migration of agents between environment types and the agent-environment impact for the different agent cognition models varying with respect to the environment types.

Static Baseline Behaviour in Multi-agent Scenario

A population of 10 agents is considered where the initial energy of the agents E_1 is randomly initialised in the range $[95, 105]$. This randomization is implemented to avoid the situation where all moderate agents require to collect resources at the same time step.

In order to establish the static baselines in each environment type, each of the agents is considered to have just a deterministic behavioural layer. The reward accumulated by the agents while being alive through 500 time steps is illustrated in tables 1 and 2. T Since there are 10 agents, building on the calibration for the single agent scenario, the initial resource level in each of the environment types is chosen as $r_1 = 500 \times 10$. The remaining parameters are kept the same as in the single agent scenario.

Environment type	**Mean Reward (All Greedy)**	**Alive Agents (All Greedy)**
Forest	3.22×10^5	10
Pasture	1.68×10^5	0
Desert	1.35×10^5	0

Table 1: Static baselines for 10 greedy agents averaged over 30 independent runs for 500 time steps each

Environment type	**Mean Reward (All Moderate)**	**Alive Agents (All Moderate)**
Forest	1.91×10^5	10
Pasture	1.91×10^5	10
Desert	1.72×10^5	2.67

Table 2: Static baselines for 10 moderate agents averaged over 30 independent runs

In the forest environment type, neither moderate nor greedy behaviour impacts the survival of the agents or the sustainability of the environment. However, the best thing to do for all the agents is to become greedy to maximise their individual rewards over the short and long term. Therefore, we can say that there is no social dilemma existing in the forest environment type.

The agents in the desert environment type can survive a little longer if they are moderate. In some cases, this may give rise to forest environment type after some agents have died and the environment can replenish steadily.

The pasture environment type gives rise to a social dilemma in all instances due to the structure of the payoffs. In the initial time steps the individual rewards at each time step are better for the greedy agents than the moderate ones. However, greedy behaviour quickly depletes the resource and eventually the agents die. The cumulative reward of the moderate agents is higher since they can survive longer. This n-player game makes moderate behaviour costly for the agents in the short term, however, provides a larger cumulative reward in the long run.

Co-evolving Deliberation through Neuroevolution

Each agent's deliberative layer is designed with neuroevolution inspired from (Robinson et al., 2007; Borg et al., 2011; Barnes et al., 2019). In general, neuroevolution can evolve topology or weights of the neural network, or both. In this case, we have only evolved the weights in line with the previous studies as mentioned above. The agent decides on the immediate action (greedy or moderate) with a two-layer fully-connected neural network: with the input layer having four neurons representing the state variables, the hidden layer with three hidden neurons and the output layer with one neuron representing the agent's actions. The input and the hidden layer have an additional neuron each corresponding to the bias. The remaining input neurons correspond to the following state parameters: the number of greedy agents, the number of resource collecting agents, the number of alive agents and the current resource level in the environment. The output values 0 and 1 correspond to moderate and greedy behaviour respectively.

Each agent has an independent neural network for which the weights are independently evolved on the basis of the state variables. The weights of the neural network are uniformly random in the range of $[-100, 100]$ at the start of an episode. Each episode with a particular weight vector is run for 500 time steps. The weight update occurs in a *(1+1)EA hill-climber* style. A mutated weight vector is generated with a mutation rate of the inverse of the number of input and hidden neurons and the mutation adds a centred Gaussian variate with a variance of 20 to the original weight. The variance is one-fifth of the maximum one-sided range of the initially chosen random weights. A mutated weight vector

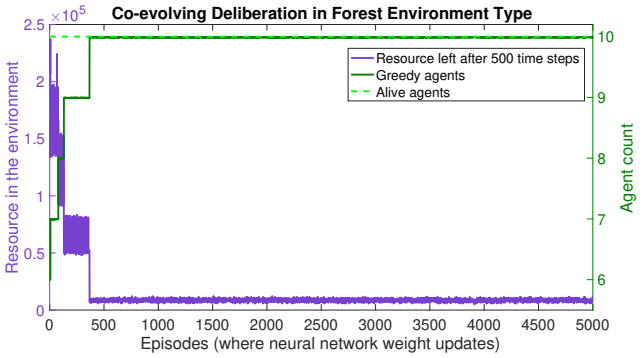

Figure 4: Forest: The state of resource and agent behaviour after 5000 episodes of weight updates with 500 time steps with 10 agents. The average number of greedy agents during an episode and the number of alive agents after the end of the episode has been shown.

is accepted only when the cumulative reward gained by the corresponding agent over the whole episode is better than the original cumulative reward.

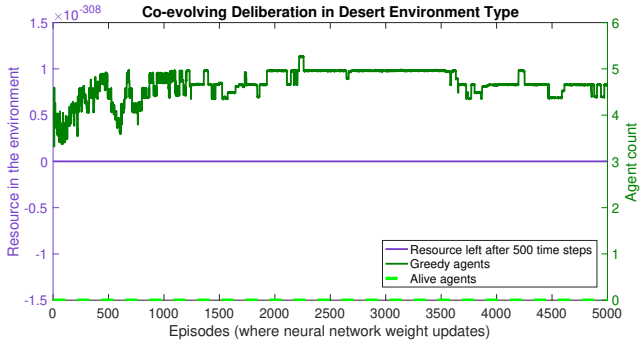

Figure 5: Desert: The state of resource and agent behaviour after 5000 episodes of weight updates with 500 time steps for 10 agents. The average number of greedy agents during an episode and the number of alive agents after the end of the episode have been shown.

With the evolution, we see that all the agents become greedy in the forest environment type. Even though the agents turn greedy and they are still able to survive as there is enough resource for all. They all become greedy as that increases their cumulative reward in their lifetime per episode. This is illustrated in figure 4. The deliberative layer works as intended and due to absence of a dilemma, the agents can survive in the forest environment type irrespective for the nature of their deliberation.

As illustrated in figures 5 and 6, in the desert environment type, at the end of each of the 5000 episodes of learning and updating the weight vector for the network, the agents

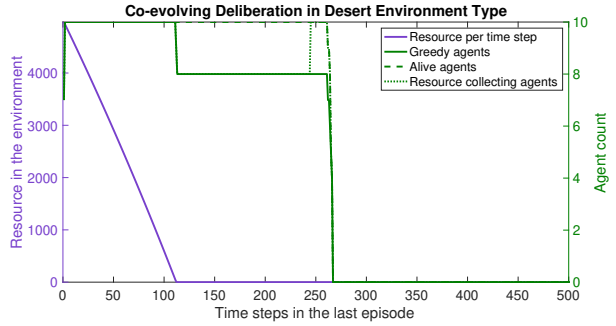

Figure 6: Desert: The state parameters at the last episode (5000th) of weight updates with 500 time steps for 10 agents

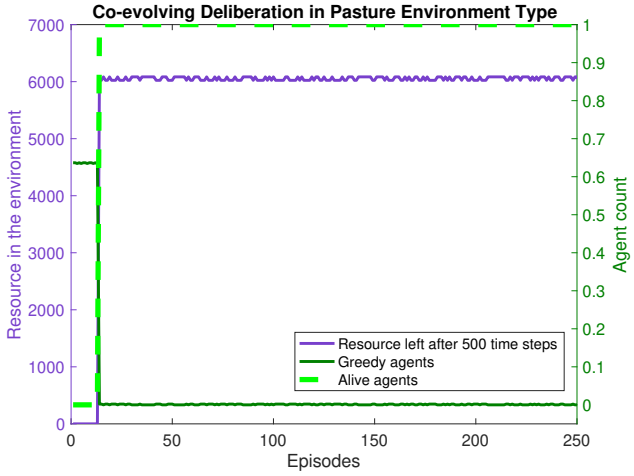

Figure 7: Pasture: The state of resource and agent behaviour after 250 episodes of weight updates with 500 time steps for a single agent. The average number of greedy agents during an episode and the number of alive agents after the end of the episode have been shown.

die. With evolution, they all become greedy to maximize their reward, as shown in figure 5 in terms of the average number of greedy agents in the total lifetime of the agents in a episode. Since they die quickly the average goes down. The behaviour of the agents in a particular episode (5000) is demonstrated in figure 6. This shows that the population gets dominated by greedy agents which brings about the depletion of resources quickly. The short term cost of being moderate impacts the long term goal of staying alive and allows the resources to replenish for longer. In this environment type, it is a no-win situation for the agents, where neither being moderate nor being greedy may allow to have a sustainable environment as illustrated in tables 1 and 2. This implies that the deliberative layer could do better in making the agents survive a bit longer, but would not be able to bring sustainability of the environment. Migration from this envi-

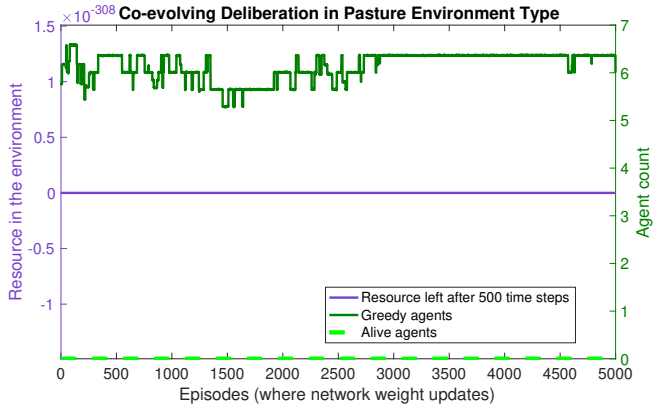

Figure 8: Pasture: The state of resource and agent behaviour after 5000 episodes of weight updates with 500 time steps for 10 agents. The average number of greedy agents during an episode and the number of alive agents after the end of the episode has been shown.

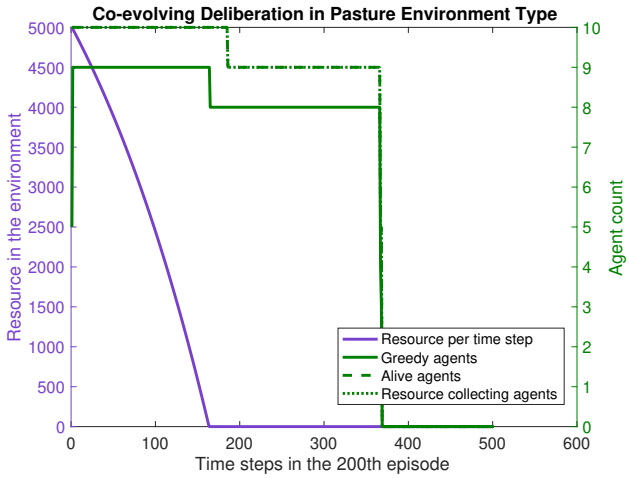

Figure 9: Pasture: The state parameters at the 200th episode of weight updates with 500 time steps for 10 agents.

ronment type to another with enough resources may a viable option for the agents though.

The forest and the desert environment types being the two extremes of the sustainable foraging testbed, the pasture environment type is the most interesting due to the arise of the co-evolutionary dynamics in agent-agent interactions (though implicit) and the agent-environment interactions. In the single agent scenario, the deliberation layer is able to learn the consequent impacts of of greedy behaviour in the long term as illustrated in the figure 7. The deliberative layer is enough to deliberate moderate behaviour to promote sustainability of the environment and eventually the agent's own life within a few episodes of learning and evolution.

However, for the multi-agent scenario, the n-player game makes individual goal more attractive that the collective long term goal, and that impacts sustainability. In the initial episodes, the agents begin to evolve in the same way as the forest environment type, because there is temporarily sufficient resource. But this quickly leads to the resource being depleted due to the nature of insufficient replenishment to carry 10 greedy agents. At this point, the greedy agents have built up substantial personal resources which allows to survive a little longer without foraging. The moderate agents on the other hand despite having contributed to the longevity of the environment, have little personal reserve and thus quickly die out. Once their personal reserves run out the greedy agents will also die out and the pasture will turn out like the desert environment type. These situations are illustrated with figures 8, 9, 10 and 11.

The figure 8 shows the presence of the social dilemma where except brief 'lucky' periods in which the multiple agents randomly explore moderate behaviour, this always collapses back to a population dominated by greedy agents. The deliberative layer falls short of learning the implications of the actions on environmental state and acting upon that to ensure sustainability. The next section discusses the possibility of an intervening reflective governance layer that is overriding the deliberative layer when connections between past actions actions and the current environmental state are identified.

Building Reflective Agents to Enable Sustainability of Environment

In situations like the sustainable foraging problem where social dilemmas impact collective actions, reflective architectures have been explored to ensure collective self-governance as well as sustainability (Dryzek and Pickering, 2017; Pitt et al., 2020; Scott and Pitt, 2023).

In the original definitions of computational reflection (Maes, 1988), reflection was considered to be reasoning performed by a system about itself. In recent years, the research considering reflection as part of the broader notion of self-awareness (i.e. becoming the object of the agent's own attention) (Morin, 2006) extended this to include not just reasoning about the agent itself (in its context) but also learning about itself (in its context) (Lewis et al., 2011; Kounev et al., 2017).

We consider a progression of levels of agent self-awareness in a similar way to (Lewis and Sarkadi, 2023), which can be described as follows:

0. With no self-awareness, the agents are only capable to collect resources by providing no attention to the availability of resources or the relationship between resource collection and environmental impact.

1. At the first stage of self-awareness, the agent is capa-

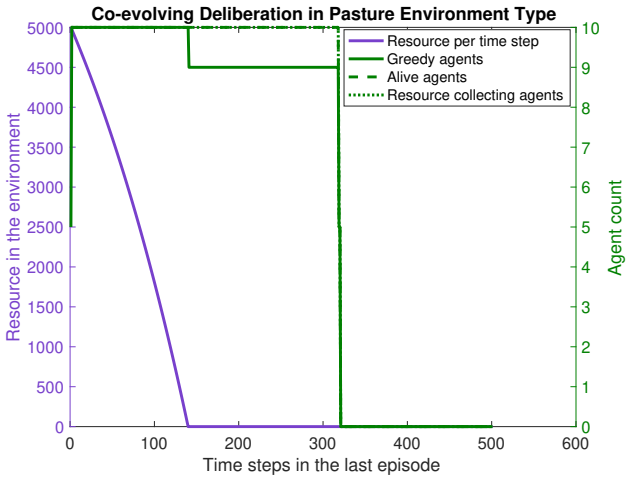

Figure 10: Pasture: The state parameters at the last, 5000th episode of weight updates with 500 time steps for 10 agents.

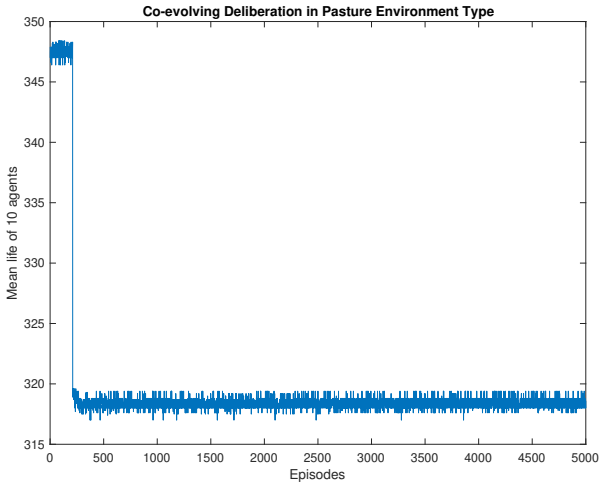

Figure 11: Pasture: The average life of 10 agents over 5000 episodes.

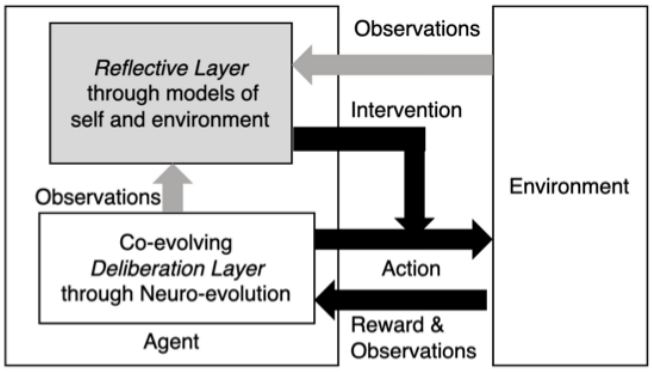

Figure 12: Proposed Reflective Agent Architecture for the Sustainable Foraging Problem

ble to collect resources when they are available. At this stage, the agent acquires knowledge regarding the availability of resources. However, the relationship between environmental impact is not explored. This level of self-awareness may be implemented with *if-else* logical statements.

2. At the second stage, the agent is capable of considering short-term goals of maximising payoffs by collecting resources at each scenario. This awareness is ecological in nature and may be achieved with deliberation layers with neuroevolution (as discussed in this work) or maybe RL-based frameworks.

3. The highest order of self-awareness considered here requires understanding the long term effects of different actions at each scenario/game model. With episodic learning approaches, then the reward accumulation may be assessed to create a model of environmental impact depending upon each course of action. However, this may also be achieved if the agent interacts with the environment and tries to reflect on the particular game model its the environment belongs. This would enable the agent to change the rules of the game and potentially escape from situations where the game is set up against the interests of the agent. We are interested to build the latter category of cognitive capability for agents in the considered sustainable foraging task.

We build on the previous work on reflective agent architecture (Lewis and Sarkadi, 2023) which marries architectures from self-aware computing (Kounev et al., 2017) and classic learning agent architectures (Russell and Norvig, 1995). One of the key ideas here is that of a reflective governor, which can intervene on the actions arising from deliberation, and is triggered by reflective reasoning. The agent architecture that has been implemented in this work is demonstrated in figure 12.

The reflective governor chosen here implements a simple *if* statement as follows:

if
$(r_{t-2} > r_{t-1})$ && $(E_{t-2} < E_{t-1})$ && $(E_{t-2} > \tau)$
then
| action = ¬(deliberation)
end

This reflective governor in each agent checks if there has been resource depletion in the previous time step and if the agent's energy has increased in that time step, then it does not act as the deliberative layer has asked it to. This means that the deliberative layer has influenced an action that has reduced the environmental resource and is directly related

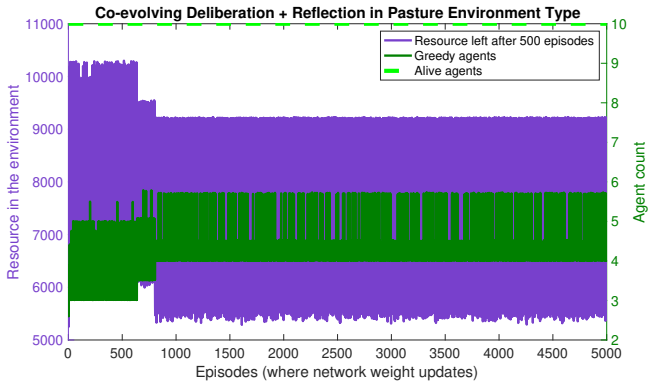

Figure 13: Pasture: The state of resource and agent behaviour after 5000 episodes with 500 time steps for 10 agents. The agents have a deliberative layer and a reflective governor intervening with the deliberative layer. The average number of greedy agents during an episode and the number of alive agents after the end of the episode have been shown.

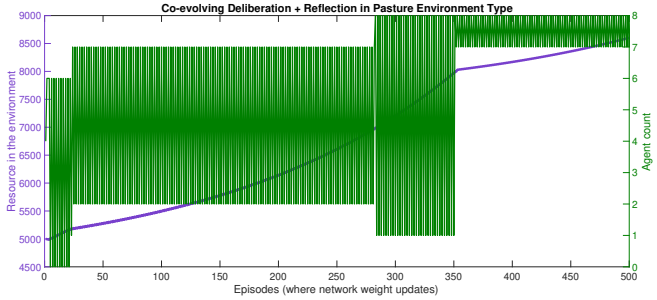

Figure 14: Pasture: The state parameters at the 450th episode with 500 time steps for 10 agents. The reflective governor intervenes with decisions made by the deliberative layer.

to the agent's own action. In such a situation, the reflective governor forces the agent to do the opposite of the deliberation it received. However, it needs to be checked that the energy level of the agent is not below the threshold τ beyond which both greedy and moderate agents collect resources for survival.

The implications of the presence of this reflective governor in the meta-layer has been illustrated in figures 13 and 14. The reflective governor has been set up in a way such that it overrides the deliberation by the neural network in alternate time steps. The reasoning is built when the agent realises the agent-environment interaction during the active resource collection in the previous time step.

As illustrated in figure 14, in the initial time steps of the 450th episode, the number of greedy agents alternate between none and six until the moderate agents have reached the threshold τ. This kind of flip-flopping behaviour of the greedy agents is caused by the reflective governor.

Until this time point of the energy of moderate agents being close to τ, since there exist alternate time steps where none of the 10 agents is collecting resources, the replenishment rate in the pasture environment type actually enables the resource to replenish. In the following time steps, the deliberative network learns to ask agents to be more greedy than before since there exist more resources than in the initial stage. At this stage, the moderate agents, require foraging in alternate time steps being close to the energy threshold.

It is interesting to note that despite having a tight bound of resources and the corresponding resource limit which can accommodate 10 agents when they are all moderate throughout the run, the reflective governor finds a solution to allow the resources to replenish and still keep the agents alive.

Conclusions and Future Steps

A sustainable foraging testbed which captures the features of a social dynamic environment has been developed for multiple agents. This testbed incorporates forest, pasture and desert environment types correlated to varying resource replenishment rates. The agents' behaviour of foraging impacts the sustainability of the environment depending upon the environment type in which the agents are foraging. The pasture environment type allows sustainability of resources only when the agents do not act with greedy behaviour of collecting resources at all instances to increase immediate rewards. A co-evolving deliberation with neuro-evolution has been implemented to learn the state of the environment and act accordingly. The deliberative layer falls short of taking into account the long term impacts of the immediate actions. A reflective governor based on the agent's self-awareness when added as a meta-layer intervening the deliberations when the resource shows signs of depletion, enables the agents to live longer and allows the resources to replenish.

The episodic and offline nature of the deliberation considered in this work means that the agents need to die to learn the impact of their actions. The future direction of this research lies in exploring the impact of an online deliberative layer that finishes the learning episode during the agent's lifetime. An evolution of the topology of the deliberative network may be explored in this context. Due to the complex nature of the social dynamics in the sustainable foraging problem, reflective reasoning with online deliberation gives much fewer opportunities for active experimentation. This experimentation may eventually allow the agent's self-awareness to be more robust to act in accordance with different environment types and ensure sustainability. It would be interesting to consider a larger population of agents to explore the agent-environment interactions, as well.

References

Axelrod, R. and Hamilton, W. D. (1981). The evolution of cooperation. *science*, 211(4489):1390–1396.

Barnes, C. M., Ekárt, A., and Lewis, P. R. (2019). Social action in socially situated agents. In *2019 IEEE 13th International Conference on Self-Adaptive and Self-Organizing Systems (SASO)*, pages 97–106. IEEE.

Barnes, C. M., Ekárt, A., and Lewis, P. R. (2020). Beyond goal-rationality: Traditional action can reduce volatility in socially situated agents. *Future Generation Computer Systems*, 113:579–596.

Barnes, C. M., Ghouri, A., and Lewis, P. R. (2022). Explaining evolutionary agent-based models via principled simplification. *Artificial Life*, 27(3–4):143–163.

Bellman, K., Botev, J., Hildmann, H., Lewis, P. R., Marsh, S., Pitt, J., Scholtes, I., and Tomforde, S. (2017). Socially-sensitive systems design: Exploring social potential. *IEEE Technology and Society Magazine*, 36(3):72–80.

Borg, J. M., Channon, A., Day, C., et al. (2011). Discovering and maintaining behaviours inaccessible to incremental genetic evolution through transcription errors and cultural transmission. In *Advances in Artificial Life, ECAL 2011: Proceedings of the Eleventh European Conference on the Synthesis and Simulation of Living Systems*, pages 101–108. MIT Press.

Boyd, R. and Richerson, P. J. (2009). Culture and the evolution of human cooperation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1533):3281–3288.

Dignum, V. and Dignum, F. (2020). Agents are dead. long live agents! In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*, pages 1701–1705.

Doebeli, M. and Hauert, C. (2005). Models of cooperation based on the prisoner's dilemma and the snowdrift game. *Ecology letters*, 8(7):748–766.

Dryzek, J. S. and Pickering, J. (2017). Deliberation as a catalyst for reflexive environmental governance. *Ecological Economics*, 131:353–360.

Kollock, P. (1998). Social dilemmas: The anatomy of cooperation. *Annual review of sociology*, pages 183–214.

Kounev, S., Lewis, P. R., Bellman, K. L., Bencomo, N., Camara, J., Diaconescu, A., Esterle, L., Geihs, K., Giese, H., Götz, S., et al. (2017). The notion of self-aware computing. In *Self-Aware Computing Systems*, pages 3–16. Springer.

Lewis, P. R., Chandra, A., Parsons, S., Robinson, E., Glette, K., Bahsoon, R., Torresen, J., and Yao, X. (2011). A survey of self-awareness and its application in computing systems. In *2011 Fifth IEEE Conference on Self-Adaptive and Self-Organizing Systems Workshops*, pages 102–107. IEEE.

Lewis, P. R. and Sarkadi, S. (2023). Reflective artificial intelligence. *arXiv preprint arXiv:2301.10823*.

Maes, P. (1988). Computational reflection. *The Knowledge Engineering Review*, 3(1):1–19.

Morin, A. (2006). Levels of consciousness and self-awareness: A comparison and integration of various neurocognitive views. *Consciousness and cognition*, 15(2):358–371.

Ostrom, E. (1999). Coping with tragedies of the commons. *Annual review of political science*, 2(1):493–535.

Panait, L. and Luke, S. (2005). Cooperative multi-agent learning: The state of the art. *Autonomous agents and multi-agent systems*, 11(3):387–434.

Pitt, J., Dryzek, J., and Ober, J. (2020). Algorithmic reflexive governance for socio-techno-ecological systems. *IEEE Technology and Society Magazine*, 39(2):52–59.

Robinson, E., Ellis, T., and Channon, A. (2007). Neuroevolution of agents capable of reactive and deliberative behaviours in novel and dynamic environments. In *Advances in Artificial Life: 9th European Conference, ECAL 2007, Lisbon, Portugal, September 10-14, 2007. Proceedings 9*, pages 345–354. Springer.

Russell, S. and Norvig, P. (1995). *Prentice Hall series in artificial intelligence*. Prentice Hall Englewood Cliffs, NJ:.

Scott, M. and Pitt, J. (2023). Interdependent self-organizing mechanisms for cooperative survival. *Artificial Life*, pages 1–37.

Winfield, A. F. (2009). Towards an engineering science of robot foraging. In *Distributed autonomous robotic systems 8*, pages 185–192. Springer.

Zedadra, O., Jouandeau, N., Seridi, H., and Fortino, G. (2017). Multi-agent foraging: state-of-the-art and research challenges. *Complex Adaptive Systems Modeling*, 5(1):1–24.