Imitating Living Organism Level Altruistic Behaviour: Sharing Food

Soomin Shin

I. INTRODUCTION

If one aims to develop a robot capable of foraging for food and evading predators, the most effective approach might involve emulating animal behaviors. A proposed simple and robust method is as follows: the robot, equipped with a camera sensor, identifies an object as food if its characteristics do not exceed a predefined predator threshold; otherwise, it classifies the object as a predator [1].

The robot's actions, whether seizing food or avoiding predators, are fundamentally instinctive. Similar to the human fight-or-flight response [2], these behaviors do not emanate from complex cognitive processes.

The question then arises: is it feasible for artificial agents to replicate behaviors that exhibit psychological attributes? Differing from basic survival actions like foraging or predator evasion, altruistic behaviors are quintessential examples that encompass psychological dimensions. This study investigates the potential of designing computational algorithms that enable virtual agents to exhibit altruistic behaviors.

II. RELATED WORKS

A. The Methods Animals Acquire Behaviours

Animal behaviors can be broadly categorized into two overarching types: innate and learned. Innate behaviors are instinctual patterns exhibited by animals, often observable in newborns, such as in the case of baby sea turtles [3]. In contrast, learned behaviors develop as animals grow. An example is certain birds using songs as alarm calls to signal the presence of predators to their kin [4]. These alarm calls are not only learned but also serve a social function, benefiting their group. Paradoxically, while producing alarm calls benefits the flock, it may also pose a risk to the caller [5]. Birds thus learn to produce and appropriately utilize these calls, sacrificing their safety for the group.

B. Altruistic Behaviour: Sharing Food

Similar to the alarm calls of birds, other examples of altruism in animals, particularly reciprocal altruism, are prevalent [6]. This concept suggests that altruistic acts are motivated by long-term self-interest. Food sharing is a commonly observed form of altruism across various species, including bats [7], chimpanzees [8], and rats [9]. Schneeberger et al. (2012) [9] found that rats who had previously received altruistic acts, such as being given food, were more likely to reciprocate. This behavior contrasts with that of rats not having experienced such altruism, underscoring reciprocal altruism's principle of future reciprocal aid.

C. Programmable Behaviour Model

1) Approach to Apply Mechanism in Biology to Robots: Transforming a behavioral or perceptual model from living organisms to inanimate agents requires the model to be programmable. Murphy (2001) [10] proposed a three-levels computational theory framework [11]. The first level involves identifying an analogous function in a living organism. The second decomposes the organism's input and output information. The third level determines how to process the input data to generate the desired output. Notably, the mechanism in the artificial agent might display variances from its biological counterpart.

2) Schema Theory: Arbib (2002) [12] introduced *'Schema Theory'* to facilitate the transformation of biological behaviors into programmable constructs. Following the computational theory structure, this transformation entails (1) identifying the organism's specific actions or perceptions and (2) developing algorithms to process these. Arbib also divided behavioral schemas into two types: motor and perceptual. Each schema type processes its input data using a unique algorithm.

D. Granting Deliberative Intelligence to Robots

A robot operating solely on S-R (stimulus-response) schema lacks deliberative intelligence. To possess such intelligence, a robot must have a goal and a plan to achieve it. One method to develop this is through STRIPS [13], which includes goal and initial states, operators, and difference evaluators. The robot discerns the gap between its current and goal states and chooses operators to effectively bridge this gap.

The robot has information of goal state and its initial state, which means it can compute the difference between the goal and initial state. Based on the calculated difference, it evaluates the difference; depending on the calculated discrepancy, it can select proper operator which can reduce the discrepancy efficiently so that reach the goal state.

E. Reinforcement Learning

Reinforcement Learning (RL) is inspired by operant conditioning in behavioral psychology [14]. Sutton (1984) [15] presented a framework adapting operant conditioning to computational problems. RL involves learning optimal behaviors through interaction with the environment, similar to biological agents [16]. Agents receive rewards or punishments based on their actions. The greater the rewards, the more likely it is that the agent will repeat the behavior.

An agent's behavior is directed by its policy. As the agent interacts with its environment, it experiences rewards or penalties, which update its reward state. The policy aims to guide the agent toward actions that maximize its rewards.

III. Hypothesis

This research explores the possibility of virtual agents, controlled by computational algorithms, demonstrating altruistic behaviors similar to those in biological entities. The central hypothesis suggests that these virtual agents can mimic sharing behaviors using reinforcement learning (RL) methodologies. Additionally, consistent with reciprocal altruism theory, it is hypothesized that these agents will progressively adopt sharing behaviors as a means to augment their own advantages.

Building upon the principles of reciprocal altruism, it is theorized that these agents will demonstrate altruistic behavior predominantly in scenarios where mutual benefits are assured. Consequently, should the agents possess the capability to accumulate surplus food in their inventories for future consumption, it is expected that the frequency of altruistic behaviors will diminish. This reduction is predicated on the assumption that the immediate need for reciprocal altruistic acts is lessened when agents can independently satisfy their requirements.

Expanding on reciprocal altruism principles, it is postulated that these agents will predominantly exhibit altruism in situations where mutual benefits are guaranteed. Therefore, the frequency of altruistic behavior is expected to vary based on the individual agents' ability to store surplus food and the specific conditions under which an altruistic act occurs.

Hypothesis 1: Virtual agents are capable of acquiring altruistic behaviors through reinforcement learning techniques.

Hypothesis 2: The range of an agent's inventory capacity is posited to have a significant influence on its propensity to engage in altruistic behavior.

Hypothesis 3: Depending on the sharing condition, agents' sharing aspects would be affected.

IV. Methods

A. Software

The implementation of embodied agents and their behaviors utilizes the Unity engine. Within this environment, reinforcement learning methodologies are employed to develop virtual agents, termed 'ml-agents' [17]. Unity is particularly conducive to the application of Proximal Policy Optimization (PPO), a technique within the reinforcement learning (RL) paradigm. The PPO algorithm in Unity trains Open Neural Network Exchange (ONNX) [18] in accordance with specific policies, aiming to optimize the expected reward [19].

The process for training and testing the customized agents includes the following steps:

1) Environment and Agent Coding: The development of the environment and agents involves writing new functions or adapting existing ones.

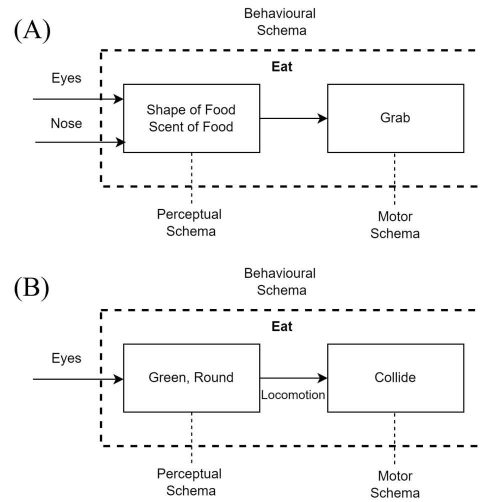

Fig. 1. The behavioural schema of eating food. (A) represents the behavioural schema of biological organism, a rat. (B) represents the behavioural schema of artificially implemented agents

2) Neural Network Training: The neural network is trained utilizing the PPO algorithm, operating within the revised policy framework.

3) Testing: The trained neural network is deployed across each agent and rigorously tested to verify their intended functionality.

V. System Design

A. Create Programmable Model From Rats

Since this study implement the higher level behaviour than intrinsic behaviour: foraging food, eating food and running away from predator, it first creates the basic model of eating food, after that it extends to the sharing behaviour using the basic model.

1) The Basic Model: Eating Behaviour: As suggested by [11] and [12], the first step is decomposing biological organism's behaviours and transforming into programmable objects. Before decomposing sharing behaviour, decompose the mechanism of rats finding foods and eat. The input sensory rats use for the eating behaviour is vision and olfactory information. With two types of sensory input, rats perceive the shape and the scent of food, which is categorized into perceptual schema, see Fig.1. Once they perceive the food, then using their locomotion ability, approach to the food and grab the food, i.e., motor schema.

Based on the decomposed programmable components, transform components into computationally executable functions. Due to the technical limitation, utilizing olfactory data has ignored. In this study only vision information is dealt for the input data (see Fig1). In this closed environment, food has been created as a green round shape. As living organism does, artificial agents move toward the food once they perceive the food. Instead of grabbing the food, in this virtual environment, agents collide with the food to perform eating behaviour.

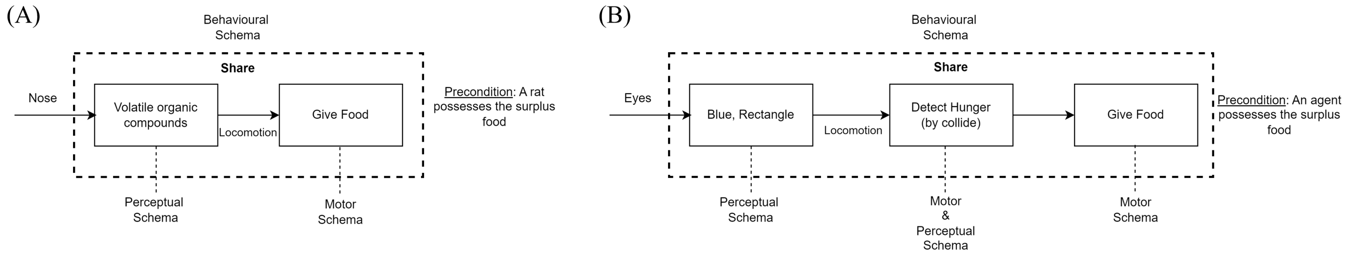

Fig. 2. The behavioural schema of sharing food, extended from the eating food model, see Fig 1. (A) represents the behavioural schema of a rat. (B) represents the behavioural schema of artificially implemented agents

2) The Extended Model: Sharing Behaviour: Sharing behaviour is implemented extended to the basic model. The biological organism, rats, sense counterpart's hunger by analyzing volatile organic compounds that the counterpart emits [20]. Once the odour information enters the rats' nose, they detect the chemical compounds and perceive if the fellow rat is hungry (perceptual schema), see Fig **??**. If they decide to share their quota, using their locomotion ability, they approach to the fellow rat and share their food. Another different part from the basic model is that for the extended model, precondition exists. They can only share if and only if they have surplus food.

In case of artificial agents, olfactory information is not able to process. As such, for the artificial agents, they only use vision information to identify counterpart agents. Agents identify counterpart agents by vision information, and identify their hunger gauge when they collide each other. Right after they collide, the agent gives the food to the counterpart. As biological agents do, artificial agents share their quota if and only if they have surplus food.

B. Deliberative Intelligent Agents

This study incorporates STRIPS (Stanford Research Institute Problem Solver) techniques into the motor schema to endow agents with deliberative intelligence. The objective of this implementation is to enable agents to perform sharing behavior. In this context, the absence of sharing behavior experience represents the *initial state* in STRIPS. Consequently, the *goal state* is defined as the agent experiencing sharing behavior. The difference between the goal state and the initial state lies in the absence or presence of sharing behavior. To operationalize this difference, the study breaks down the transition from the initial state to the goal state into several stages, ultimately leading to the experience of sharing behavior.

When an agent's hunger level falls below 800 (the consumption threshold), it is considered hungry. The *Eat* operator is triggered only if the agent has surplus food in its inventory (see Table I). Upon colliding with a fellow agent, if the other agent's hunger level is not lower than its own—implying 'I am hungrier than you'—the agent disregards the encounter and continues to move. Finally, if the other agent's hunger level is higher, indicating greater hunger, the agent shares its food with this hungrier agent.

C. Learning Strategy

This study applied existing untrained Open Nueral Network Exchange (ONNX) [18], RL algorithm and basic food collecting algorithm to train the agents. On top of the existing algorithm and untrained network, this study modified the algorithm to imitate the biological organism's behaviour and trained from the scratch as follows.

1) Basic Algorithm of Experiment: The experimental environment will utilize a predefined setting provided by Unity [17] (refer to Fig.**??**). This environment features a fixed rectangular area where multiple agents and food items are generated. Agents are programmed to navigate within this rectangular field, interacting with other agents and food items. Upon contact with a food item, an agent consumes it, leading to the item's removal from the environment. Each consumption event rewards the agent.

For the project's extension beyond the basic environment setup, the following three packages from Unity will be employed: UnityEngine, UnityEngine.UI, and Unity.MLAgents

2) Modified Algorithm to Mimic Biological Agents: The pre-made environment primarily trains virtual agents to wander and consume food. For this study, several functions have been modified and created to the agents' functionalities using C# language. These include:
1) Inventory Capacity: Each agent is equipped with a following fixed inventory size (inventory size: 1,2,5) to store excess food.
2) Hunger Level: Agents possess a hunger level that starts at 1000 and gradually decreases over time to 0.
3) Consumption Threshold: When an agent's hunger level falls below the threshold (800), it consumes food if it forages the food or it has a food in its inventory.
4) Mortality Condition: An agent is programmed to receive punishment once its hunger level reaches 0.
5) Altruistic Sharing: Agents are designed to share food from their inventory with other agent only if that is in a state of greater hunger.
6) Reward for Sharing: To encourage sharing behavior, a reward system is implemented, awarding agents who share their food quota.

VI. Experimental Design

Regardless of experimental conditions, all hyperparameters for configuration are fixed and consistent over all conditions.

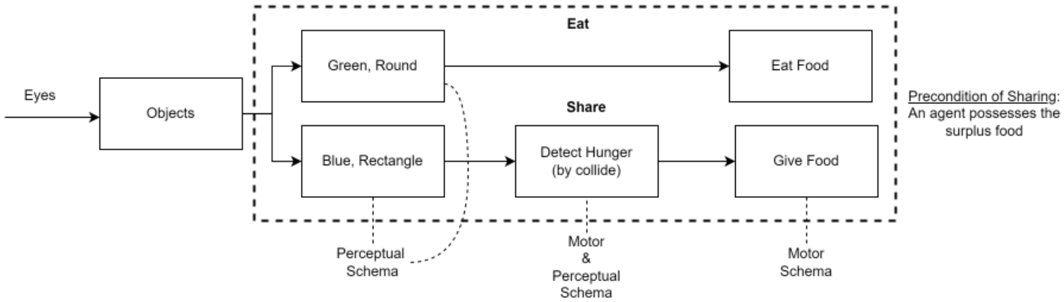

Fig. 3. Integrated model of eating food (Fig.1 (B)) and sharing food (Fig.2 (B)) behaviour models.

TABLE I

Difference	Operator
my_hunger_level <= 800 && food in inventory >=1	Eat
my_hunger_level >counterpart_hunger_level	Ignore
my_hunger_level <counterpart_hunger_level && food in inventory >=1	Share

All conditions utilized the same algorithm, Proximal Policy Optimization (PPO). The entire training steps were 2,000,000. Hunger level decreased by 2 at each second from 1000.

A. *Independent Variables*

Sharing behaviour condition, capacity of inventory and the number of food differed depending on the conditions.

1) Sharing behaviour condition: divided into three conditions: baseline, unconditional sharing, conditional sharing.

- Baseline: Not equipped with sharing function. Except the sharing behaviour, modified eating algorithm has applied.
- Unconditional Sharing: Agents share their surplus food whenever they encounter fellow agents.
- Conditional Sharing: Agents share their surplus food *if and only if* the fellow agent is more hungry.

2) Inventory Capacity: The inventory capacity in this study varies among three sizes: 1, 2, and 5. Consequently, each sharing behavior condition encompasses all three inventory size variations, culminating in a total of nine experimental conditions.

3) Differentiating Sharing from Stealing: In addition to the aforementioned conditions and modifications, an algorithm that distinguishes stealing behavior from sharing has been applied across all conditions (excluding the baseline). To engage in sharing behavior, agents are required to collide with other agents for less than 1 second. Furthermore, if an agent assesses a fellow agent's hunger level within a brief time window, it is not necessary for it to repeatedly collide with the same agent. Consequently, this study has developed an algorithm specifically to differentiate between stealing and sharing, as detailed in Algorithm 1.

Algorithm 1

$C_N \leftarrow$ "" $\triangleright$ Initialize the collided agent's named
$CT_C \leftarrow curentTime$
if $C_N \neq$ "" **then**
  if $CT_C \geq CT_P + 1$ **then**
    $Value \leftarrow +1$
  else
    $C_N \leftarrow FellowAgnet'sName$
    $CT_P \leftarrow curentTime$
  end if
else
  $C_N \leftarrow FellowAgnet'sName$
  $CT_P \leftarrow curentTime$
end if

VII. Results

As the data did not meet the criteria for normality, the Kruskal-Wallis Test was used for analysis, and the Mann-Whitney U Test was employed for post-hoc analysis.

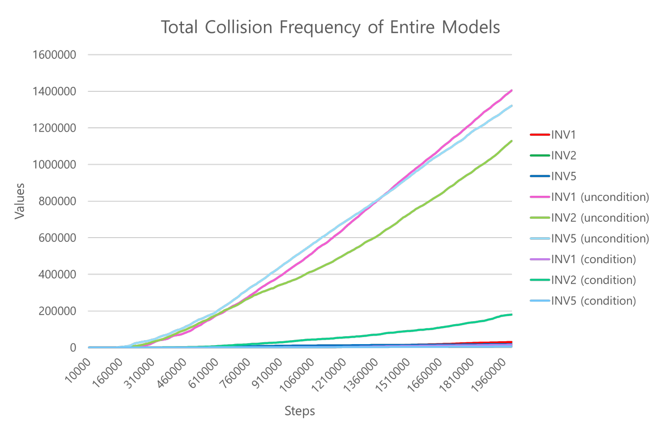

Fig. 4. Comparison the frequency of collision over entire models.

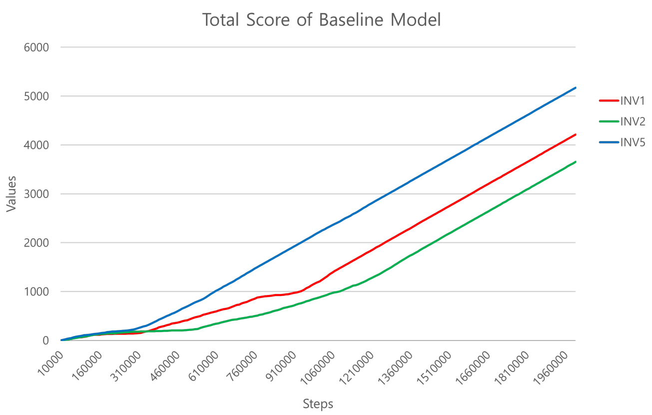

Fig. 5. INV1 corresponds to Inventory size 1, INV2 to Inventory size 2, and INV5 to Inventory size 5. The X-axis represents the number of training steps (*Max*=2,000,000).

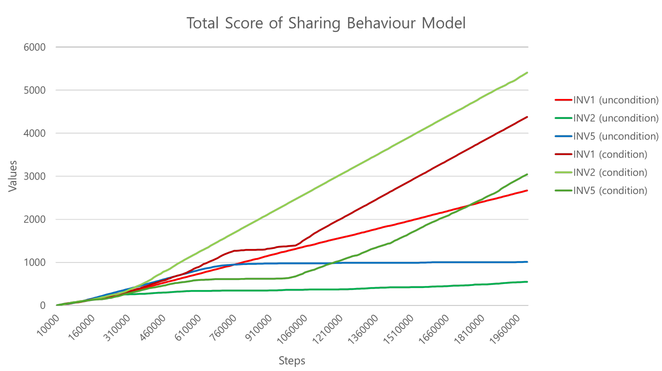

Fig. 6. The total score of the unconditional/conditional sharing model. INV1 corresponds to Inventory size 1, INV2 to Inventory size 2, and INV5 to Inventory size 5.

A. Baseline

In the baseline model, all experimental conditions show an incremental trend in values (INV_1, $\mathbf{R}^2$ = 0.958; INV_2, $\mathbf{R}^2$ = 0.920; INV_5, $\mathbf{R}^2$ = 0.992) (see Fig.5).

The three experimental conditions are significantly different (*H(2)* = 40.027, *p* = .001). The most notable performance is observed in the condition with the largest inventory capacity; INV_5 outperforms the smaller inventory sizes (*z* = -4.013, *p* = .000), and INV_5 also shows better performance than INV_2 (*z* = -6.155, *p* = .001). This is followed by INV_2 and INV_1 (*z* = -2.455, *p* = .014).

B. Unconditional Sharing Behavior Model

1) Total Score: The total score includes values from the food eating score (+1.0), sharing score (+0.5), and punishment score (-1.0). Although all conditions show an increasing trend, except for INV_1, INV_2 and INV_5 exhibit weak performance (INV_1, $\mathbf{R}^2$ = 0.999; INV_2, $\mathbf{R}^2$ = 0.857; INV_5, $\mathbf{R}^2$ = 0.678) (see Fig.6).

The three conditions significantly differ (*H(2)* = 207.799, *p* = .000). In this scenario, INV_1 demonstrates the most pronounced performance compared to INV_2 (*z* = -11.652, *p* = .000) and INV_5 (*z* = -12.398, *p* = .000). Additionally, INV_5 scores higher than INV_2 (*z* = -6.518, *p* = .000).

2) Sharing Score: Significant differences are observed in the frequency of sharing behavior across the three conditions (*H(2)* = 206.969, *p* = .000), as seen in Fig.7 (A). Since the total score includes sharing behavior scores, the order of performance in sharing behavior follows that of the total score, with INV_2 performing better than INV_5, which in turn outperforms INV_1 (*z* = -13.178, *p* . 000; *z* = -5.417, *p* . 000). Therefore, *Hypothesis 2* is supported in the unconditional sharing behavior scenario.

3) Time in Physical Proximity: Fig.7 (B) illustrates the duration of collisions involving more than two agents. To accurately measure the duration of these collisions, the algorithm (see Algorithm. 1) differentiates whether the collision is for sharing or stealing purposes.

As shown in Fig.7 (C), which presents the difference between (A) and (B), agents spend a considerable amount of time clustered together relative to the frequency of sharing behavior. The order of clustering duration is $INV_1 > INV_5 > INV_2$ (*z* = -4.496, *p* = .000; *z* = -2.697, *p* = .007). This indicates that agents tend to cluster together even when sharing behavior is not occurring.

C. Conditional Sharing Behavior Model

1) Total Score: Even in this experiment, where food sharing occurred only if the other agent was hungrier, the total score increased significantly compared to the unconditional sharing scenario.

All conditions demonstrate increasing values over time (Inventory size 1, $\mathbf{R}^2$ = 0.974; Inventory size 2, $\mathbf{R}^2$ = 0.996; Inventory size 5, $\mathbf{R}^2$ = 0.905) (see Fig.6).

The three experimental conditions are significantly different (*H(2)* = 70.785, *p* = .000). INV_1 outperforms INV_2 (*z* = -3.849, *p* = .000) and INV_5 (*z* = -5.322, *p* = .000). Moreover, INV_2 scores higher than INV_5 (*z* = -7.965, *p* = .000).

2) Sharing Score: Significant differences are noted in the sharing behavior frequency among the three conditions (*H(2)* = 147.867, *p* = .000), as depicted in Fig.8 (A). The performance in sharing behavior varies depending on the inventory capacity, akin to the conditional sharing scenario. The performance order is as follows: $INV_2 > INV_5 > INV_1$ (*z* = -10.677, *p* = .000; *z* = -3.422, *p* = .001). Given these findings, *Hypothesis 2* is supported in the conditional sharing behavior model.

3) Time in Physical Proximity: Similar to the unconditional sharing behavior model, the conditional model shows agents clustering over time. To specifically examine this, the time spent in clusters was subtracted from the total number of sharing behaviors (see Fig.8 (C)). The graph for INV_5 remains close to zero. Although INV_1 shows a slightly lower graph than INV_5 (*z* = -4.541, *p* = .000), it also stays near zero. In contrast, the graph for INV_2 exhibits a noticeably lower parabolic shape (*z* = -9.409, *p* = .000).

This suggests that agents with a larger inventory size tend to cluster together, even if they have nothing to share. Conversely,

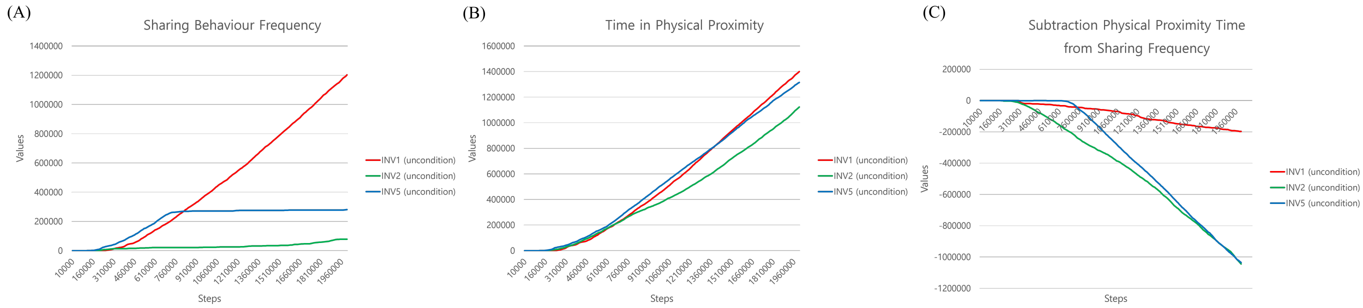

Fig. 7. Results of the unconditional sharing behavior model. The red line represents cases where agents have a single unit of inventory capacity, green for two units, and blue for five units. (A) indicates the cumulative frequency of sharing behavior per step. (B) shows the time spent after agents have formed a cluster. For clarity in differentiating between the number of sharing behaviors and clustering time, (C) depicts the results obtained by subtracting (B) from (A).

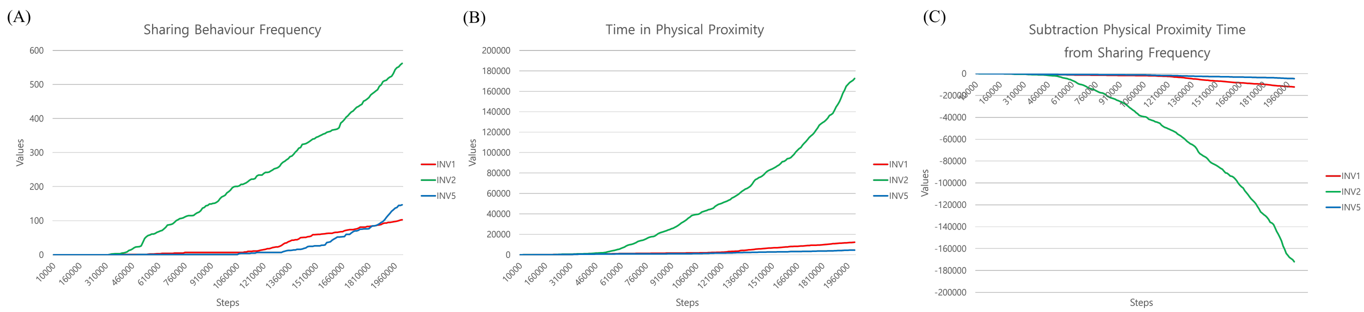

Fig. 8. Results of the conditional sharing behavior model.

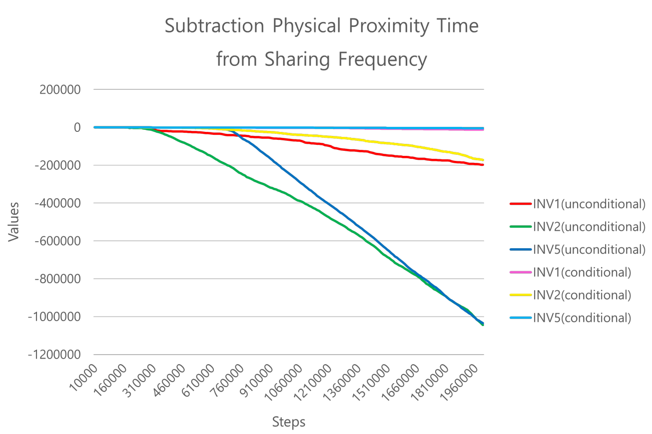

Fig. 9. In the unconditional sharing behavior model, significant differences are observed in the frequency of sharing behavior and clustering time for inventory sizes of two and five. In contrast, the conditional sharing behavior model with inventory sizes of one and five shows minimal differences between these variables.

agents with a single inventory unit collide primarily to share their food.

D. Comparison of Unconditional and Conditional Sharing Models

Considering the sharing behavior frequency, agents exhibited the most remarkable performance with a single inventory unit (INV_1) in the unconditional sharing model ($\mathbf{R}^2$ = 0.972), outperforming the INV_2 condition in the conditional model ($\mathbf{R}^2$ = 0.968) (Z = -21.169, p = .000).

When clustering duration is included, the results reverse. Although the two aforementioned conditions showed high performance, if clustering time is subtracted, it appears their performance heavily relies on clustering (see Fig.7 (C) and 8 (C)).

Conversely, based on graph (C) of Fig.8, INV_5 in the conditional sharing behavior model indicates that sharing frequency and clustering time are similar, supporting *Hypothesis 3*.

INV_5 in the conditional model not only demonstrates an increase in sharing frequency, but also shows that this frequency is not due to clustering behavior aimed at stealing food. Therefore, we can conclude that *Hypothesis 1* is supported in the INV_5 condition of the conditional sharing behavior model.

VIII. Discussion

In this study, we discovered that by administering rewards and punishments, artificial agents can mimic altruistic behavior, such as sharing food. However, the agents' primary innate mechanism is selfishness, a trait that closely resembles the cognitive layer in humans.

Axelrod and Hamilton (1981) [21] used the prisoner's dilemma to elucidate human altruism. The essential motivation behind human altruism is the enhancement of self-interest. This theory can explain the behavior of the agents in this study.

TABLE II
SUMMARY OF RESULTS PER EMPLOYED MODEL

	Baseline	Unconditional Sharing Model	Conditional Sharing Model
Collision Frequency	Rarely occurs	The highest collision frequency	Close to Baseline
Clustering Time	No cluster formation	The highest clustering time	No cluster except the Inventory size 2
Sharing Frequency	None	Lower than conditional model	The highest sharing frequency

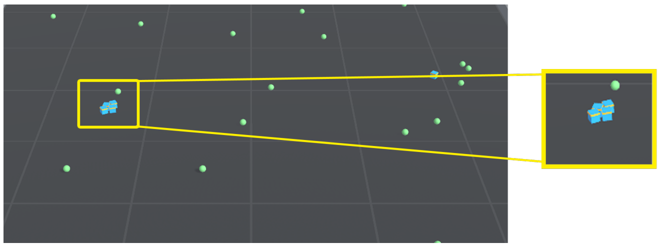

Fig. 10. This screenshot shows agents' performance in the unconditional sharing behavior model. Four out of five agents formed a cluster, even though none had surplus food in their inventory.

My findings indicate that agents' behavior can be altered by modifying the reward system conditions. In the baseline model, the frequency of collisions is significantly lower compared to other models (see Fig. 4). This suggests that an agent avoids colliding with its peers if there is no personal gain from such collisions. Therefore, the collision count in the baseline model represents the minimal number of collisions agents engage in while moving to feed themselves.

The addition of a reward system for sharing behavior changed the aspects of the agents' behavior. When the system rewarded sharing behavior, the collision count increased significantly. However, the resulting sharing behavior exhibited two different characteristics depending on the conditions of the sharing behavior.

A. Provoking Stealing Behavior

The study observed that when agents are programmed to share food with peers upon encounter, the number of collisions increased significantly, independent of inventory capacity. This behavior model also revealed that agents often collide and form clusters, remaining close to each other (see Fig. 10).

They maintain this formation even when none of the clustered agents possesses surplus food. This occurs because they understand that they can steal food whenever any agent locates it. Therefore, agents prefer to remain close to other agents to steal food rather than roam the environment for foraging. Under these conditions, they have learned that stealing is more advantageous than foraging in terms of increasing their self-interest.

B. Provoking Sharing Behavior

Reflecting human behavior, humans share when they have ample surplus food and when another individual is in need. Based on this premise, agents were programmed to share their quota only if the encountered agents were hungrier.

The study demonstrated that under the conditional sharing behavior model, agents could exhibit altruistic behavior akin to humans. Contrary to the previous model, agents neither formed clusters nor wandered to forage. However, this interpretation is limited when agents have an inventory size of two. The models with inventory sizes of one and five align with this interpretation.

C. Potential Application Domains

According to Chen et al. (2022) [22], a robot's altruistic behavior can impact the level of trust humans place in the robot. Thus, this study's findings could be applied to research aimed at increasing human trust in robots by enabling them to autonomously mimic altruistic behavior.

IX. LIMITATIONS

A. Arbitrarily Modified Variables

Five variables were arbitrarily determined: 1) hunger level, 2) when agents consume food, 3) when agents share food, 4) the capacity size of inventory, 5) the quantity of food.

While these variables are critical and have a significant impact, they were not established based on robust theory.

B. Black Box Process

The study noted variations in agent behavior depending on variable and algorithm modifications. As the outcomes resulted from reinforcement learning, it is not possible to elucidate the rationale behind the agents' choice of optimal behavior.

X. CONCLUSION

This study identified conditions under which virtual agents can imitate altruistic behavior. One scenario demonstrated optimal performance for sharing behavior: agents with a five-unit inventory capacity sharing food when they encounter agents hungrier than themselves.

Under this scenario, the frequency of sharing behavior steadily increased over time. This was not due to cluster formation. In other scenarios, physical proximity time increased exponentially, suggesting that sharing occurred mainly due to consistent collisions and proximity, akin to food being stolen.

Conversely, the frequency of sharing behavior and physical proximity time were similar under the optimal condition. This suggests that agents collided mainly to share food, which more closely resembles animal sharing behavior.

REFERENCES

[1] M. A. Arbib and J.-S. Liaw, "Sensorimotor transformations in the worlds of frogs and robots," *Artificial Intelligence*, vol. 72, no. 1-2, pp. 53–79, 1995.

[2] A. S. Jansen, X. Van Nguyen, V. Karpitskiy, T. C. Mettenleiter, and A. D. Loewy, "Central command neurons of the sympathetic nervous system: basis of the fight-or-flight response," *Science*, vol. 270, no. 5236, pp. 644–646, 1995.

[3] F. Scapini, "Heredity and learning in animal orientation," *Monitore Zoologico Italiano-Italian Journal of Zoology*, vol. 22, no. 2, pp. 203–234, 1988.

[4] A. J. Leavesley and R. D. Magrath, "Communicating about danger: urgency alarm calling in a bird," *Animal behaviour*, vol. 70, no. 2, pp. 365–373, 2005.

[5] P. Darlington Jr, "Altruism: Its characteristics and evolution," *Proceedings of the National Academy of Sciences*, vol. 75, no. 1, pp. 385–389, 1978.

[6] E. Fehr and U. Fischbacher, "The nature of human altruism," *Nature*, vol. 425, no. 6960, pp. 785–791, 2003.

[7] G. S. Wilkinson, "Reciprocal food sharing in the vampire bat," *Nature*, vol. 308, no. 5955, pp. 181–184, 1984.

[8] J. B. Silk, S. F. Brosnan, J. Henrich, S. P. Lambeth, and S. Shapiro, "Chimpanzees share food for many reasons: the role of kinship, reciprocity, social bonds and harassment on food transfers," *Animal behaviour*, vol. 85, no. 5, pp. 941–947, 2013.

[9] K. Schneeberger, M. Dietz, and M. Taborsky, "Reciprocal cooperation between unrelated rats depends on cost to donor and benefit to recipient," *BMC evolutionary biology*, vol. 12, pp. 1–7, 2012.

[10] R. R. Murphy, *Introduction to AI robotics*. MIT press, 2019.

[11] D. Man and A. Vision, "A computational investigation into the human representation and processing of visual information," *WH San Francisco: Freeman and Company, San Francisco*, vol. 1, p. 1, 1982.

[12] M. A. Arbib, *The handbook of brain theory and neural networks*. MIT press, 2002.

[13] R. E. Fikes and N. J. Nilsson, "Strips: A new approach to the application of theorem proving to problem solving," *Artificial intelligence*, vol. 2, no. 3-4, pp. 189–208, 1971.

[14] B. F. Skinner, "Operant behavior.," *American psychologist*, vol. 18, no. 8, p. 503, 1963.

[15] R. S. Sutton, *Temporal credit assignment in reinforcement learning*. University of Massachusetts Amherst, 1984.

[16] V. François-Lavet, P. Henderson, R. Islam, M. G. Bellemare, J. Pineau, *et al.*, "An introduction to deep reinforcement learning," *Foundations and Trends® in Machine Learning*, vol. 11, no. 3-4, pp. 219–354, 2018.

[17] A. Juliani, V.-P. Berges, E. Teng, A. Cohen, J. Harper, C. Elion, C. Goy, Y. Gao, H. Henry, M. Mattar, and D. Lange, "Unity: A general platform for intelligent agents," *arXiv preprint arXiv:1809.02627*, 2020.

[18] ONNX, "Open neural network exchange," 2024.

[19] A. Cohen, E. Teng, V.-P. Berges, R.-P. Dong, H. Henry, M. Mattar, A. Zook, and S. Ganguly, "On the use and misuse of absorbing states in multi-agent reinforcement learning," *RL in Games Workshop AAAI 2022*, 2022.

[20] K. Schneeberger, G. Röder, and M. Taborsky, "The smell of hunger: Norway rats provision social partners based on odour cues of need," *PLoS biology*, vol. 18, no. 3, p. e3000628, 2020.

[21] R. Axelrod and W. D. Hamilton, "The evolution of cooperation," *science*, vol. 211, no. 4489, pp. 1390–1396, 1981.

[22] N. Chen, Y. Zhai, and X. Liu, "The effects of robots' altruistic behaviours and reciprocity on human-robot trust," *International Journal of Social Robotics*, vol. 14, no. 8, pp. 1913–1931, 2022.