Facilitating Synchronized Movement during Ice-Breaking Scenarios through a Real-World Reinforcement Learning Agent Using Non-Verbal Behaviors

Ge (Serena) Guo
Cornell University
Ithaca, New York
gg372@cornell.edu

Qi Yang
Cornell University
Ithaca, New York
qy244@cornell.edu

Rejoice Hu
Cornell University
Ithaca, New York
rh635@cornell.edu

Guy Hoffman
Cornell University
Ithaca, New York
hoffman@cornell.edu

2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI) | 979-8-3503-7893-1/25/$31.00 ©2025 IEEE | DOI: 10.1109/HRI61500.2025.10973921

***Abstract*—This study investigates the application of real-world Reinforcement Learning (RL) in facilitating synchronized movement and fostering social connections during initial interactions between strangers. We introduced a novel, projection-based system where an RL agent is projected onto a table to interact with participants through its movements. Our research involved the development and testing of a Deep Q-Network (DQN) integrated with a Long Short-Term Memory (LSTM) model that identifies synchronized movements and determines the reward for the DQN. The preliminary results showcase the effectiveness of the LSTM model, revealing that synchronized movement can be a potent reward mechanism, enabling RL agents to learn effective policies. The initial findings indicate the feasibility of the RL approach in enhancing social interactions, providing partial support for our designed actions.**

***Index Terms*—Social Robot; Human-human-robot Interaction (HHRI); Reinforcement Learning; Agent, Deep Q-Network (DQN); Non-verbal Behavior**

I. Introduction

The field of Human-human-robot interaction (HHRI) has seen increasing interest in the design of robots capable of moderating and enhancing social interactions between humans. Prior research has explored using robots to shape interpersonal behaviors and evaluation in terms of conflict regulation [9], turn-taking [14], and empathy evoking [8], etc. However, few studies have investigated how robot behaviors could help encourage social connection by creating synchronized movement during initial interactions between previously unacquainted humans.

Our study aims to explore whether a robot can detect the movements of humans and adapt its own to foster synchronized behavior between strangers to promote a sense of social affiliation and closeness during icebreaker scenarios. Physical synchronized movement (or interactional synchrony) is defined as the simultaneous coordination of physical actions between two or more people [3]. Research has highlighted the link between synchronized movement and various prosocial outcomes including cooperation, feelings of unity, and self-other overlap [18], and an evolutionary mechanism for enabling social human bonding among large groups [12].

This study investigates whether a reinforcement learning agent can learn policies to encourage synchronized movement between strangers during an ice-breaking scenario, to foster a sense of connection. The agent will be projected on a table and choose from various movement strategies to prompt two participants to coordinate their motions.

II. Related Work

A. Mediating interpersonal communication through robot's non-verbal behaviors

Research in Human-human-robot interaction (HHRI) explored diverse methodologies to mediate interpersonal communication, with a particular emphasis on the non-verbal behaviors exhibited by robots. For example, Rifinski et al. demonstrated that even minimal socially responsive gestures, such as gaze and leaning motions from a simple robot, can positively impact interpersonal evaluations between two humans engaged in a face-to-face debate conversation [17]. Similarly, Press et al. revealed that robotic objects performing humorous behaviors, characterized by a mix of unexpected slow and fast movements can reduce social awkwardness between two strangers [15]. Besides, Tennent et al. have used an everyday object, a microphone (Micbot), to perform back-channeling movements to signal encouragement for participants to speak during team conversations. Their findings suggest that those movements of a robot can effectively increase group engagement and performance[19]. These findings suggest that subtle robot behaviors can positively shape human-human social interactions. Building on this, our study seeks to extend this research by exploring whether dynamic movement behaviors could play a similarly constructive role.

B. Computer-mediated synchronized movements

Research has explored different interventions that encourage synchronized movements in small groups. For instance, the *Two Riders* installation, which resembles two interconnected bicycles, encourages synchronized pedaling among pairs [11]. Similarly, the *21 Balancoires* installation invites people to

Authorized licensed use limited to: Cornell University Library. Downloaded on November 06,2026 at 18:37:23 UTC from IEEE Xplore. Restrictions apply.

swing together to change the background music, creating a playful and collaborative environment [4]. the *SocialStools* is a set of three interactive stools that leverages imageries and music to mediate strangers' proxemics interactions [6]. Though not investigated in a human-human interaction context, Bartkowski et al. showed that a robot matching a human's movements led to increased ratings of trust compared to non-matched movements [2]. Although this research focused on human-robot trust, it provides preliminary evidence that synchronization of motions may support positive assessments. However, there is a need for more research specifically addressing synchronized movements in collaborative human-robot contexts. Our study aims to conduct such an investigation to fill this gap in a human-human-robot interaction setting.

To mathematically model synchronization behaviors for human-robot joint action scenarios, prior research has employed dynamical systems techniques as well as phase and event estimation methods [13]. Our study similarly adopts a computational approach utilizing temporal sequence predictions and reinforcement learning algorithms to automatically derive adaptive robot behavior for initiating and sustaining synchronized movement between humans.

III. System

The system comprises two main components: (1) a physical setup enabling participants' real-time interaction with the simulated agent; (2) a Deep Q-Network (DQN) agent paired with a Long Short-Term Memory (LSTM) model, determining the agent's reward. Our system's components are illustrated in Figure 1.

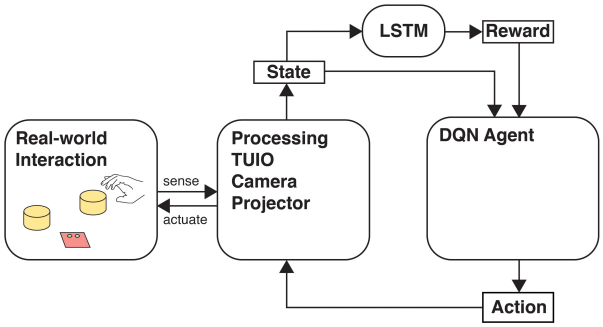


Fig. 1. System Diagram of Actuation, Sensing, LSTM-Based Reward Classification, and the DQN Agent.

To reduce the training time and to quickly test the feasibility of our model, we used hand-held objects as a proxy of the physical setup with stools. This relatively low-fidelity prototype will enable us to simulate real-world stool interactions.

The rest of the setup is similar to a previous study done by Guo et al. [6], [7]: Our physical system consists 2 wooden cylindrical objects with fiducial markers that represent participants' avatars, a 4K webcam mounted overhead to identify the stool positions, and a 1080p projector with 3000 Lumens brightness for projecting agent and player avatars onto the floor (Figure 2). We used reacTIVision, an open-source computer vision framework to acquire the position and orientation of the markers.

The visual projection and locomotion of the robots are coded and actuated using Processing [16] with TUIO protocol [10]. The robot is represented as a red square projection with two green eyes. The human is represented as a yellow circle with two red eyes. The position and orientation data from real-world interactions are acquired via the camera, processed in Processing, and then sent to an LSTM model to determine the robot's reward. The state is also sent to the DQN agent to determine the next action. The action is then sent back to Processing, and visualized in the physical world. We implemented LSTM and DQN in Python and used User Datagram Protocol (UDP) to enable communication between programs.

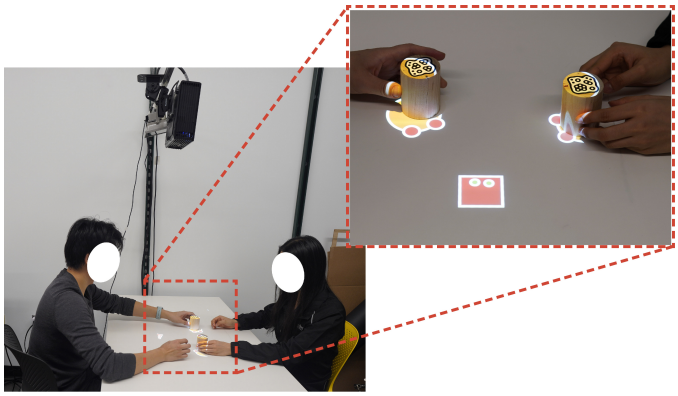


Fig. 2. Experimental Setting for the Online Session with Two Researchers Interacting with the DQN Agent.

A. DQN Agent Design

As our state space is immense and many similar states may share similar Q-values, evaluating the Q-table poses a challenge for the typical Q-learning agent. Furthermore, we have chosen synchronized movement as our agent's reward, a metric challenging to quantify using rule-based algorithms. Therefore, we have employed Deep Q-Networks (DQN) with Long Short-Term Memory (LSTM) to determine whether the movements of the three players are synchronized.

1) State: The dimension of the state space is 25 sequential frames x 9 features. As illustrated in Figure 3, the 9 features include (1) orientations of 3 players (2 humans, 1 agent), (2) distances between each pair, and (3) angles of the formed triangle. The Processing sends these 9 features to the DQN agent at 5 Hz. We include 25 sequential frames to capture the sequential movements of the three players and balance model complexity.

2) Action: We designed four actions that could potentially stimulate synchronized movement: *Initiate*, *Spin*, *Fidget*, and *Coordinate*. Each action should be considered a strategy comprising a sequence of low-level locomotion control and decision-making processes. In the *Initiate* action, the agent

Authorized licensed use limited to: Cornell University Library. Downloaded on November 06,2026 at 18:37:23 UTC from IEEE Xplore. Restrictions apply.

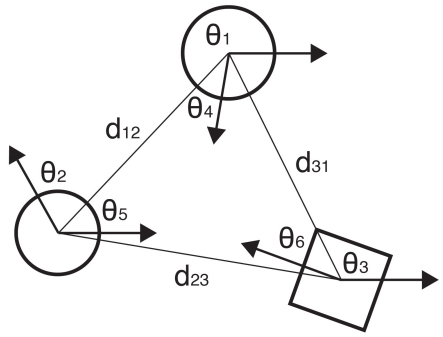


Fig. 3. States of the DQN Agent and Extracted Features for Classifying Synchronized Movement at One Time Step.

randomly selects and moves towards an individual to begin collaborative rotation. Should the individual remain stationary, the agent will rebound and attempt to approach up to three times before redirecting toward another person. During the *Spin* action, the agent rotates either clockwise or counterclockwise for five seconds in its current location. In the *Fidget* action, the agent moves slightly left and right in a repetitive manner. The *Coordinate* action involves the agent identifying the nearest vertex on the minor axis of an ellipse, whose major axis aligns with the line between two human positions, and consistently moving towards this vertex at a fixed speed. The four actions are shown in Figure 4.

Throughout all actions, the agent maintains a safe distance from humans and actively avoids collisions. It consistently orients itself towards the center of the triangle formed by the two humans and the agent.

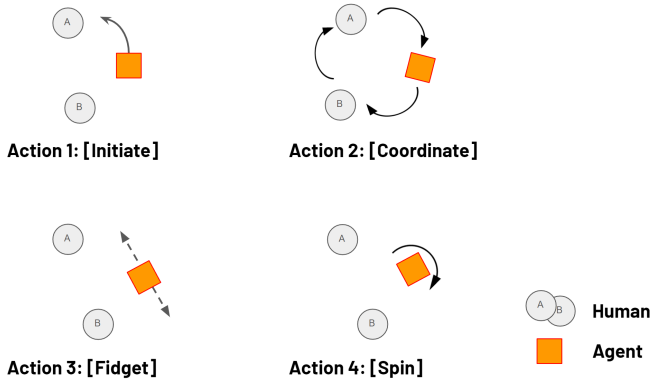


Fig. 4. Four Actions of the Agent: Initiate, Spin, Fidget, Coordinate

3) Reward: Successful classification of synchronized movement is critical for the DQN learning process. Rule-based algorithms (i.e. calculating the entropy of three players' velocity and setting up a threshold) might be simple and straightforward, but they cannot capture the complexity and diversity of synchronized movements. Therefore, we decided to take the machine learning approach.

Due to the sequential nature of our data, we trained an LSTM model for the binary classification task: classify synchronized movements from desynchronized movements given a 5-second sequence of movements. If the synchronized movement was identified, the agent received a 1.0 reward for the 5-second sequence. The agent received a -1.0 reward otherwise.

We then created our training data. During the training, each of the three researchers manipulated the position of a wooden cylinder (the avatar) and rotated in circles with different speeds, distances, directions of rotation, and locations. In total, 132 5-second synchronized samples were collected. 96 5-second desynchronized movements were collected in a similar manner where three researchers randomly moved their avatars, or simply remained still.

Given the importance of the reward function in the DQN algorithm, we assessed several models. Rule-based algorithms are simple but fail to capture complex synchronized movements. KNN and SVM handle small datasets and high-dimensional spaces, respectively, but lack the ability to model temporal dependencies. Random Forest captures non-linear interactions but is not designed for sequential data. LSTM excels at modeling temporal patterns but requires extensive data and risks overfitting.

4) DQN Training: The DQN model architecture involves 3 Dense layers with 128, 64, and 4 units and ReLU activation functions for the first two layers and a linear activation function for the last. The MSE loss and Adam optimizer with a learning rate of 0.001 was used. The discount rate gamma was set to 0.95, the exploration rate epsilon to 1.0, the minimum epsilon to 0.01, and the epsilon decay rate to 0.995.

After obtaining a strong LSTM model, DQN training ensued in a simulated environment as a proof of concept. The robot randomly selects one of the 4 actions initially, and the human participants respond to its movements. During our pilot-test sessions, the researchers intentionally tried to coordinate with the robot to test the learning efficacy of the DQN agent.

IV. Results

A. LSTM Performance

Rule-based algorithms and KNN's were eliminated due to their limitations in capturing complex high-dimensional patterns. The best test accuracies obtained after optimal hyper-parameter tuning with our training data was 0.478 for SVM, 0.804 for Random Forest, and 0.870 for LSTM. Therefore, the LSTM model demonstrated superior test accuracy compared to these alternatives. This LSTM model's configuration consisted of 100 units, a dropout rate of 0.4, a learning rate of 0.001 using the Adam optimizer, a batch size of 16, and a training duration of 100 epochs. However, indications of overfitting were observed in the loss curve. To address this, we conducted further validations through online testing. During these tests, three researchers alternated between synchronized and desynchronized movements to assess the accuracy and responsiveness of the models. The results were notably promising as demonstrated in our video.

Authorized licensed use limited to: Cornell University Library. Downloaded on November 06,2026 at 18:37:23 UTC from IEEE Xplore. Restrictions apply.

B. DQN performance

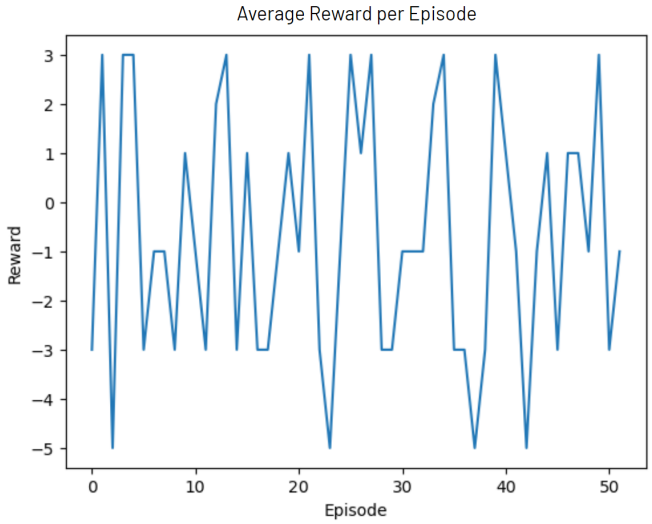


Fig. 5. Average Reward per Episode in the Online Interaction Session

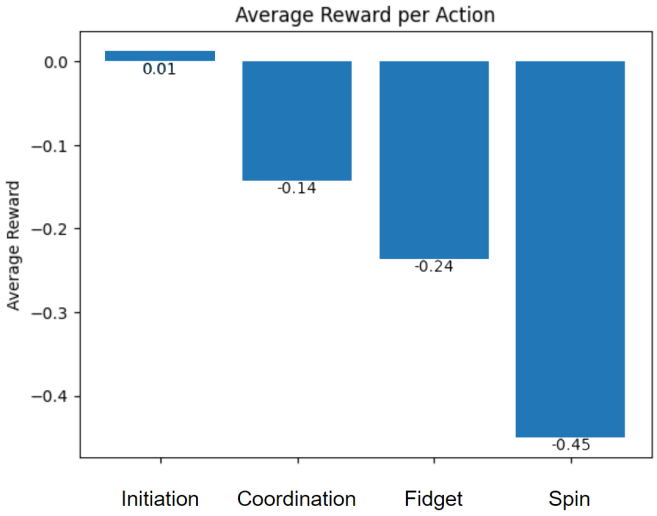


Fig. 6. Average Rewards for Four Different Actions During the Online Interaction Session

We interacted with the DQN for 50 episodes which took around 30 minutes. We reported the average reward per episode (Figure 5) and the average reward per action (Figure 6). In Figure 5, we did not observe an increase in received reward throughout time and the DQN clearly has not converged yet due to the limited number of episodes. However, Figure 6 demonstrated preliminary inspiring results. The Initiation and Coordination actions in which researchers intentionally coordinated with the agent clearly received higher rewards than the Fidget and Spin actions. The results of the current Initiation and Coordination actions can be successfully identified as synchronized movements when researchers intentionally coordinate with the agent.

The results highlighted key insights beyond extending the DQN agent's training. Negative rewards across all actions suggest the LSTM model occasionally misclassified synchronized movements, likely due to unaccounted human reaction time. Humans needed a few seconds to respond, often missing rewards as episodes ended prematurely. To improve, we propose lowering the reward threshold, making agent actions clearer, increasing classification frequency, and using majority voting to reduce noise.

V. Discussion

Our approach and experiment pose several interesting directions for future researchers.

Unlike state machine strategies with pre-defined if-then rules, our study used Reinforcement Learning (RL) to dynamically map state spaces to actions. While rule-based behaviors can perform well, they limit exploration of the policy space, leading to suboptimal performance in unforeseen scenarios and reducing the likelihood of natural, human-like behaviors. In contrast, RL allows agents to develop policies through interaction, enabling discovery of high-performing strategies with well-defined parameters. Our findings show that non-verbal behaviors, such as synchronized movement, can serve as effective rewards, guiding RL agents to moderate social interactions more effectively.

Determining a reward system for an RL agent in ice-breaking scenarios requires insights from social psychology. Social interactions typically progress through phases like initial attention [5], self-disclosure [1], and synchronized movements [18], and rewarding the agent for navigating these phases is beneficial. In our experiment, the agent sometimes missed rewards despite signs of human engagement, such as fidgeting, which indicates a willingness to interact. Rewarding actions that elicit such initiations could improve outcomes. Future work should develop a hierarchical reward system with a curriculum learning strategy aligned with the natural progression of human interactions to enhance the agent's effectiveness in fostering meaningful social engagement. Data augmentation, optimized DQN complexity, and pre-trained models based on proxy human behavior can also balance performance and data availability in real-world scenarios.

VI. Conclusion

We developed a DQN agent that learns to moderate social interaction through real-world interactions with humans using synchronized movement as the reward. The preliminary testing proves the feasibility of our system and the efficacy of the reward classification function. The result suggests more salient agent intention in actions, more robust reward evaluation, and longer training episodes will improve the DQN agent performance.

References

[1] Altman, I. & Taylor, D. Social penetration: The development of interpersonal relationships. (Holt, Rinehart & Winston,1973)

Authorized licensed use limited to: Cornell University Library. Downloaded on November 06,2026 at 18:37:23 UTC from IEEE Xplore. Restrictions apply.

[2] Bartkowski, W., Nowak, A., Czajkowski, F., Schmidt, A. & Müller, F. In Sync: Exploring Synchronization to Increase Trust Between Humans and Non-humanoid Robots. *Proceedings Of The 2023 CHI Conference On Human Factors In Computing Systems*. pp. 1-14 (2023)

[3] Bernieri, F. & Rosenthal, R. Interpersonal coordination: Behavior matching and interactional synchrony. (Editions de la Maison des Sciences de l'Homme,1991)

[4] Bloom, J. Musical Swings: An exercise in collaboration. Impacts of a temporary installation on public spaces in three US cities — A comparative analysis.. (2016)

[5] Byrne, D. The Attraction Paradigm. (Academic Press,1971)

[6] Guo, G., Leshed, G. & Green, K. "I normally wouldn't talk with strangers": Introducing a Socio-Spatial Interface for Fostering Togetherness Between Strangers. *Proceedings Of The 2023 CHI Conference On Human Factors In Computing Systems*. (2023,4), https://dl.acm.org/doi/abs/10.1145/3544548.3581325

[7] Guo, G., Leshed, G., Pinch, T. & Evan Green, K. SocialStools: A Playful, Socio-Spatial Interface for Fostering Togetherness Across Strangers. *Extended Abstracts Of The 2022 CHI Conference On Human Factors In Computing Systems*. (2022), https://doi.org/10.1145/3491101.3519877

[8] Hoffman, G., Zuckerman, O., Hirschberger, G., Luria, M. & Shani Sherman, T. Design and Evaluation of a Peripheral Robotic Conversation Companion. *Proceedings Of The Tenth Annual ACM/IEEE International Conference On Human-Robot Interaction*. pp. 3-10 (2015), https://doi.org/10.1145/2696454.2696495

[9] Jung, M., Martelaro, N. & Hinds, P. Using Robots to Moderate Team Conflict: The Case of Repairing Violations. *Proceedings Of The Tenth Annual ACM/IEEE International Conference On Human-Robot Interaction*. pp. 229-236 (2015), https://doi.org/10.1145/2696454.2696460

[10] Kaltenbrunner, M., "TUIO: A Protocol for Tabletop Tangible User Interfaces," Version 1.1, 2005–. [Online]. http://www.tuio.org/

[11] Rinott, M. & Tractinsky, N. Designing for interpersonal motor synchronization. *Human–Computer Interaction*. **37**, 69-116 (2022)

[12] Launay, J., Tarr, B. & Dunbar, R. Synchrony as an adaptive mechanism for large-scale human social bonding. *Ethology*. **122**, 779-789 (2016)

[13] Mörtl, A., Lorenz, T. & Hirche, S. Rhythm patterns interaction-synchronization behavior for human-robot joint action. *PloS One*. **9**, e95195 (2014)

[14] Mutlu, B., Shiwa, T., Kanda, T., Ishiguro, H. & Hagita, N. Footing in human-robot conversations: How robots might shape participant roles using gaze cues. *2009 4th ACM/IEEE International Conference On Human-Robot Interaction (HRI)*. pp. 61-68 (2009)

[15] Press, V. & Erel, H. Humorous Robotic Behavior as a New Approach to Mitigating Social Awkwardness. *Proceedings Of The 2023 CHI Conference On Human Factors In Computing Systems*., 1-16 (2023,4), https://doi.org/10.1145/3544548.3580821

[16] Processing Foundation, "Processing: A Flexible Software Sketchbook and a Language for Learning How to Code within the Context of the Visual Arts," Version 4.x, 2001–. [Online]. https://processing.org

[17] Rifinski, D., Erel, H., Feiner, A., Hoffman, G. & Zuckerman, O. Human-human-robot interaction: robotic object's responsive gestures improve interpersonal evaluation in human interaction. *Human–Computer Interaction*. 36, 333-359 (2021,7), https://doi.org/10.1080/07370024.2020.1719839

[18] Sebanz, N. & Knoblich, G. Prediction in Joint Action: What, When, and Where. *Topics In Cognitive Science*. **1**, 353-367 (2009)

[19] Tennent, H., Shen, S. & Jung, M. Micbot: A Peripheral Robotic Object to Shape Conversational Dynamics and Team Performance. *2019 14th ACM/IEEE International Conference On Human-Robot Interaction (HRI)*. pp. 133-142 (2019)

Authorized licensed use limited to: Cornell University Library. Downloaded on November 06,2026 at 18:37:23 UTC from IEEE Xplore. Restrictions apply.