Motivating Physical Activity via Competitive Human-Robot Interaction

Anonymous Author(s)
Affiliation
Address
email

Abstract: This project aims to motivate research in competitive human-robot interaction by creating a robot competitor that can challenge human users in certain scenarios such as physical exercise and games. With this goal in mind, we introduce the Fencing Game, a human-robot competition used to evaluate both the capabilities of the robot competitor and user experience. We develop the robot competitor through iterative multi-agent reinforcement learning, and show that it can perform well against human competitors. Our user study additionally found that our system was able to continuously create challenging and enjoyable interactions for humans and the majority of human subjects considered the system to be entertaining and useful for improving the quality of their exercise.

Keywords: Competitive-HRI, Reinforcement Learning

1 Introduction

Competition is ubiquitous in the natural world [1, 2] and in human society [3, 4, 5]. Despite its universality, *competitive interaction* has rarely been investigated in the field of Human Robot Interaction, which has mainly focused on *cooperative interactions* such as collaborative manipulation, mobility assistance, feeding, and so on [6, 7, 8, 9, 10]. In some ways it is not surprising that competitive interaction has been overlooked: of course everyone wants a robot that can assist them; who would want a robot that thwarts their intentions? Yet, we also accept that human-human competition can be healthy and productive, for example in structured contexts such as sports. In this paper we explore the idea that human-robot competition can provide similar benefits.

We believe that physical exercise settings such as athletic practice, fitness training, and physical therapy are scenarios in which competitive HRI can benefit users. Physical exercise is essential to our physical and mental health. In addition, the quality of some physical rehabilitation is closely related to patients' commitment to exercising regularly. However, the major cause of poor adherence to physical exercise is due to a lack of motivation [11, 12, 13]. It has been shown that enjoyment, competition, and challenges are important motivational factors for the practice of physical exercise [14, 12]. In this project, our goal is to create a robot that can provide these motivations to human users via adversarial behaviors. The main contributions of this paper are listed below:

Motivating Competitive-HRI Research. We discuss how competitive interaction can influence people in a positive manner and the technical difficulties competitive-HRI tasks represent. This discussion motivates our proposed Fencing Game, a competitive physically interactive game, as an evaluation environment for competitive-HRI algorithms and user study.

System Design and Implementation. A highly generalizable robotic system is created to support our competitive-HRI research. A multi-agent reinforcement learning(RL) method is used to train a robot to play competitive games with multiple gameplay styles.

Two User Studies. Our first user study found that more than 80% of subjects found competitive interaction with our robot to be entertaining and desirable. Our system was able to provide challenging gameplay experiences that significantly increase human players' heart rates. Furthermore,

Submitted to the 5th Conference on Robot Learning (CoRL 2021). Do not distribute.

we observed that human subjects who constantly explore different strategies tend to achieve higher rewards in the long run. A second user study demonstrated that an RL-trained policy made it significantly more challenging for subjects to make quantitative improvement in a long sequence of games compared to a carefully designed heuristic baseline policy. The RL-trained agent also appeared to be more intelligent to the human subjects.

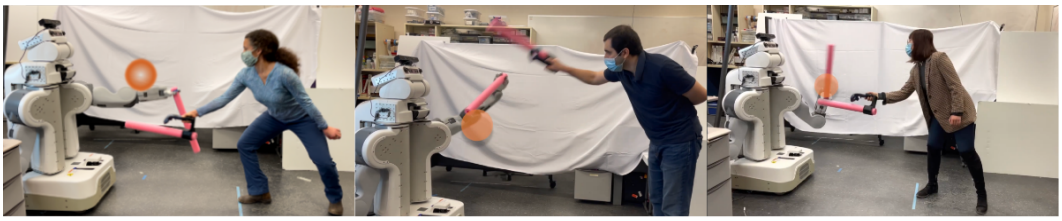

Figure 1: Competitive fencing games between a PR2 robot and human subjects. The detailed game rules are described in Sec. 2. Please refer to this link for anonymous example gameplay videos.

2 Competitive Interaction

In this section, we discuss the significance of competitive interaction and justify why we believe that it deserves increased attention in the HRI community. We will first discuss how competition can influence people in a positive manner from a psychological perspective. Afterward, we will discuss the technical challenges in competitive-HRI tasks and propose the use of the Fencing Game as the main task that this project will focus on.

Positive Influences of Competitive Interaction. Competition between players provides motivation and can foster improvement in performance at a given task. Plass et al. [15] compared competitive versus cooperative interactions in an educational mathematics video game. The results revealed that, compared to working individually, the subjects performed significantly better when working competitively than cooperatively. In particular, competitive players demonstrated higher effectiveness in problem solving compared to non-competitive players. This study also observed that subjects experienced a higher level of enjoyment via an increased tendency to engage in the game [16]. Furthermore, they also displayed higher situational interest, *i.e.* the subject paid more attention and had more interaction during the game [17]. Viru et al. [18] showed that competitive exercises can improve athletic performance in a treadmill running test. Subjects' average running duration was increased by 4.2%. During cycling and planking exercise, Feltz et al. [4] were able to inspire higher performance from subjects by placing them in competition with a manipulated virtual partner.

Inspired by these studies, we envision that a personal robot can become a competitive partner that provides enjoyment, increases motivation, and motivates improvement in activities such as physical exercise. For this reason, we initiate our competitive-HRI research by focusing on creating a physical exercise companion.

Technical Challenges. Creating an actual robot that can compete with a human physically is challenging. The robot needs to constantly reason about the human's intent via their actions, and strategically control its high degree-of-freedom body to counteract the opponent's adversarial behavior and maximize its own return. Therefore, a big part of our competitive-HRI research focuses on solving these technical challenges to create a robotic system with real-time decision making capability and body agility that is comparable to that of humans.

The Fencing Game. Based on the expected technical challenges, we designed a two-player zero-sum physically interactive game. The Fencing Game is an attack and defense game where the human player is the antagonist with the goal of maximizing their game score. The robot is the protagonist who aims to minimize the antagonist's score. Fig. 1 shows three images of human subjects playing the game, and the algorithm in Appendix A.1 summarizes the scoring mechanism for this game. The orange spherical area located between the two players denotes the target area of the game. The antagonist on the right earns 1 point for every 0.01 seconds that their bat is placed within the target area without contacting the opponent's bat. The antagonist will lose 10 points if the antagonist's bat is placed within the target area and makes contact with the protagonist's bat simultaneously.

Moreover, the antagonist will get 10 points of reward if the protagonist's bat is placed within the target area, waiting for the antagonist to attack, for more than 2 seconds. Each game will last for 20 seconds. The observation space for both agents includes the Cartesian pose and velocity of the two bats, as well as the game time in seconds. For the sake of simplicity, both agents in this project are non-mobile. Yet, mobility can be easily integrated into future iterations of this game.

3 Related Work

Reinforcement Learning in Competitive Games. Competitive games have been used as benchmarks to evaluate the ability of algorithms to train an agent to make rational and strategic decisions [19, 20, 21]. Multi-agent reinforcement learning methods allow agents to learn emergent and complex behavior by interacting with each other and co-evolving together [22, 23, 24, 25, 26, 27]. Many recent efforts have used multi-agent RL methods to learn continuous control polices that can achieve high complexity tasks. Bansal et al. [28] created control policies for simulated humanoid and quadrupedal robots to play competitive games such as soccer and wrestling. Lowe et al. [29] has extended DDPG [30] to multi-agent settings by employing a centralized action-value function.

Human-Robot Competition. There are a few studies focusing on human-robot interaction in competitive games. For instance, Kshirsagar et al. [31] studied the affect of "co-worker" robots on human performance when they are performing in the same environment and competing for a monetary prize. This study showed that humans were slightly discouraged when competing against a high performing robot. On the other hand, humans exhibited a positive attitude towards the low performing robot. Mutlu et al. [32] showed that male subjects were more engaged in competitive video games when they played with an ASIMO robot. However, the majority of the subjects preferred cooperative games when they played with this robot. Short et al. [33] analysed the "rock-paper-scissors" game and found that human subjects were more socially and mentally engaged when the robot cheated during the game.

Robots in Physical Training. In the context of using robots to assist humans in physical exercises, a robot developed by Fasola and Mataric [34] was able to provide real-time coaching and encouragement for a seated arm exercise. Süssenbach et al. [35] developed a motivational robot for indoor cycling exercise that employed a set of communication techniques in response to the subject's physical condition. The results showed that the robot sufficiently increased the users' workout efficiency and intensity. Sato et al. [36] created a system capable of imitating the motion and strategy of top volleyball blockers for assisting vollyball training.

These works are typically limited to a few competitive scenarios that require simple and repetitive motions. In this work, we leverage reinforcement learning to create a robotic system that can potentially play various physically competitive games against human players.

4 System Design and Implementation

Humans are highly efficient at recognizing patterns and learning skills from just a small number of examples [37]. Existing research also shows that human subjects can adapt to robots and improve their performance in just a few trials in physical HRI tasks [38, 39]. However, games against a robot competitor are less challenging if a human player can easily predict its behavior and quickly find an optimal counter-strategy. We hypothesized that human players can quickly learn to improve their performance against a given robot policy, but a change in the robot's gameplay style could interrupt this learning effect and keep the games challenging. A gameplay style is characterized by patterns in the agent's end-effector motion trajectories. For example, one antagonist may prefer using more stabbing movements, while another antagonist may prefer slashing movements. Therefore, there are two primary objectives that govern the generation of robot control policies. First, a policy should allow the robot to play the Fencing Game sufficiently well, such that the games are intense and engaging to the human users. Second, we propose to obtain three policies with unique gameplay styles for our user study. A multi-agent reinforcement method created the robot control policies that are used in the user study. We designed and implemented the physical system based on a PR2 general-purpose robot. The implementation of our physical system is detailed in Appendix A.2.

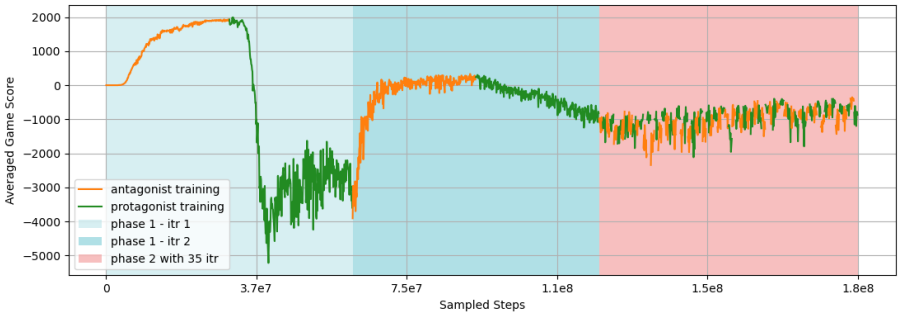

Figure 2: The antagonist's average game score during one complete round of phase one and two training. The two iterations of phase one training enabled two agents to interact according to the game rules. The phase two training performed 35 small updates resulting in random characteristics for both agents.

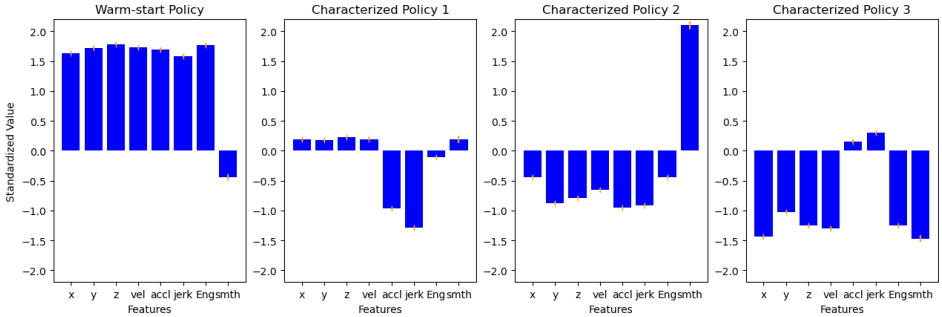

Figure 3: Visualization of quantified gameplay style of the four policies used in the user study. Error bars indicate the standard deviation of the feature values among the population.

4.1 Learning to Compete

To generate robot control policies that comply with the two aforementioned requirements, we formulate the Fencing Game as a multi-agent Markov game problem [40]. Both the antagonist agent and the protagonist agent are represented by a PR2 robot model in a Mujoco simulation environment. Both agents are trained in a co-evolving manner by playing games against each other. We break down the multi-agent proximal policy optimization method proposed by Bansal et al. [28] into two separate training processes, which reduces the computational processing needed to obtain multiple pairs of agents with acceptable performance. As a result, we were able to complete all sampling and training processes on a desktop computer (cpu: $1 \times i7$, gpu: $1 \times GTX970$). Our version of multi-agent PPO, the two-phase iterative training algorithm, is presented in Algo. 1.

Learning to Move and Play. The first phase of the training can be seen as a pre-training process, which aims to allow both agents to quickly learn the motor skills required for joint control and the rules of the game. The agents are rewarded by both the weighted sum of a continuous reward and the game score at each timestep. The continuous reward encourages the agents' exploration in the task space. The policy of the antagonist μ with parameters θ_i^μ will first be trained by collecting trajectories that result from playing against the protagonist with its most recent policy. This process continues until timeout or μ has converged. The protagonist's policy ν with parameter θ_i^ν will then be trained against the antagonist with its most recent policy. We were able to acquire robot policies that play the Fencing Game sufficiently well by only running this training sequence twice ($N_{iter} = 2$). The pair of resulting policies from phase one will be called the warm-start policies.

Creating Characterized Policies. It has been shown that by using different random seeds to guide policy optimization toward different local optima, an agent can learn different behaviors and approaches to complete the same task [41, 26]. The highly variable nature of multi-agent systems enhances this random effect. The agents in multi-agent systems are more likely to learn drastically

different strategies and emergent behaviors when they continuously learn by competing with each other [28, 42, 43, 44]. The second phase of training generates a policy with random characteristics by exploiting this fact. Both agents are initialized with their warm-start policies from phase one and trained in the same iterative scheme shown in Algo. 1. Now the agents are solely rewarded by the game scores. When training each of the agents, instead of having an agent face against the opponent's latest policy, one of the previous versions of the opponent's policy in the history will be randomly selected in phase two. We created a policy library that contains six pairs of randomly characterized policies that resulted from six different rounds of phase two training. Fig. 2 demonstrates the change of game scores during the two-phase training process.

4.2 Selecting Agents With Distinct Gameplay Styles

In order to identify the most distinctive protagonist policies in the library described in the last subsection, each agent's gameplay style needs to be quantified and compared. We first generate trajectories for all protagonist policies in a tournament [45], where each agent plays 100 games with each of the six opponents. Eight end effector trajectory features are selected to quantify agent's gameplay styles: total displacement change on x, y, and z axis, average velocity, average acceleration, average jerk, total kinetic energy, and trajectory smoothness. These features are calculated for each of the games played by each of the protagonists. The quantified style of a protagonist agent is the averaged features across 600 games. The features of all protagonists are then compared via their three most significant principal components (less than 2% of information loss)[46], and the three most separable policies are selected for the user study. The protagonist's gameplay style for the warm-start policy and the three selected characterized policies are visualized in Fig. 3.

Algorithm 1: Iterative Two-Agent Learning

Input: Environment $\mathcal{E}$; Stochastic policies μ and ν; Instantaneous Reward Function *r(·)*
Initialize: Parameters θ_0^μ for μ and θ_0^ν for ν
for $i = 1, 2, ..N_{iter}$ **do**
  if *phase one training* **then**
    $\theta_i^\nu \leftarrow \theta_{i-1}^\nu$
  else
    $\theta_i^\nu \leftarrow \theta_{random_from_history}^\nu$
  end
  for $j = 1, 2, ..N_\mu$ **do**
    rollout $\leftarrow roll(\mathcal{E}, \mu_{\theta_i^\mu}, \nu_{\theta_{i-1}^\nu}, r(\cdot))$
    $\theta_i^\mu \leftarrow$ PPO_Update(rollout)
  end
  if *phase one training* **then**
    $\theta_i^\mu \leftarrow \theta_i^\mu$
  else
    $\theta_i^\mu \leftarrow \theta_{random_from_history}^\mu$
  end
  for $j = 1, 2, ..N_\nu$ **do**
    rollout $\leftarrow roll(\mathcal{E}, \mu_{\theta_i^\mu}, \nu_{\theta_i^\nu}, r(\cdot))$
    $\theta_i^\nu \leftarrow$ PPO_Update(rollout)
  end
end

4.3 Sim2Real

Due to the kinematic and dynamic mismatch between simulation and reality, policies that are trained solely on simulated data can perform poorly in reality. We use a combination of a Jacobian Transpose end-effector controller and a system identification (systemID) process to solve this problem. Instead of specifying the torque values for each joint, the policy outputs an offset from the current end-effector pose. The new desired pose is executed by the end-effector controller. We used the CMA-ES algorithm to optimize the following objective over the parameter space of both the controller and the robot model in the simulation.

$$(\theta_m^*, \theta_c^*) = \arg \min_{(\theta_m, \theta_c)} \sum_{t=0}^{T} (s_r^t - s_s^t)^2$$

Where θ_m represents the simulated robot model parameters: damping, armature, and friction loss. θ_c represents the proportional gains and derivative gains of the end-effector controller. T_r and T_s are trajectories sampled from the real robotic system and simulation respectively that result from the same control sequence. $s_r^t \in T_r$ and $s_s^t \in T_s$ are the robot's end-effector pose in reality and simulation at time t respectively. As a result, the difference in end-effector dynamics between the simulated robot and the real PR2 robot is reduced. The maximum controller output is bounded conservatively to prevent possible human injury.

Table 1: Subjective Ratings of the Modified TAM Questions

	1-Strongly Disagree	2-Disagree	3-Neutral	4-Agree	5-Strongly Agree
Perceived Usefulness	0%	6.25%	25%	37.5%	31.25%
Perceived Ease of Use	0%	18.75%	18.75%	62.5%	0%
Attitude	0%	6.25%	37.5%	25%	31.25%
Intention to Use	0%	18.74%	18.75%	25%	37.5%
Perceived Enjoyment	0%	6.25%	6.25%	31.25%	56.25%
Desirability	0%	6.25%	12.5%	31.25%	50%
Increase Engagement	6.25%	6.25%	31.25%	18.75%	37.5%

5 Experiments and Analysis

We performed two in-lab user studies in this work. The first user study performed a broad exploration on the idea of competitive-HRI under the fencing game setting. Sixteen human subjects were asked to play five games with each of the four RL trained policies that resulted from Sec. 4. Subjects' game scores, heart rates, arm movements, and their responses to a modified technology acceptance model (TAM)[47] were used to evaluate our system from the following three perspectives: 1. Is a competitive robot accepted by human users under the scenarios of physical games and exercise? 2. Can our system effectively create challenging and intense gameplay experience? 3. Can the robot interrupt the human learning effect by switching its gameplay style?

The second user study compared characterized policy 1 with a carefully designed heuristic-based policy. By placing its bat in between the target area and the point on the human's bat that is closest to the target area, the robot exploits embedded knowledge of the game's rules in order to execute a strong baseline heuristic policy. Action noise was added to the heuristic policy to create randomness in the robot's behavior. The design details of the heuristic-based policy are described in Appendix A.3. Ten human subjects were asked to play 10 games with each robot policy. This experiment compared the two policies via game scores, subjects' TAM responses, subjects' perception of difficulty, enjoyment, and robot intelligence. Details about the experiment procedures, subjective question design, and the demographic information for both user studies are discussed in Appendix A.3.

5.1 User Study One Result: A Broad Exploration

Our first experiment demonstrates that participants highly accept the use of a competitive robot as an exercise partner. The majority of the subjects considered competitive games with our robot to be useful, entertaining, desirable, and motivating. Our system was able to provide a challenging and intensive interactive experience that significantly increased subjects' heart rates. While competing against our RL trained robot, most subjects struggled to significantly improve their performance over time. Yet, a subset of subjects who constantly explored different strategies achieved higher scores in the long run.

User Acceptance. Table 1 summarizes the subjects' responses to the technology acceptance model. The majority of the subjects (*i.e.* 68.75%) agreed that a competitive robot could improve the quality of their physical exercise. Moreover, 87.5% of subjects agreed that competitive human-robot interactive games are entertaining, and 81.25% of subjects agreed that a competitive robot exercise companion would be desirable in the future. Interestingly, the intention to use (62.5% agree + strongly agree) and increased engagement (56.25% agree + strongly agree) metrics are not as high as the vast agreement on perceived enjoyment and desirability. Some subjects who didn't have a strong intention to use our system explained their reasons in the open-ended question. They stated that it is not immediately clear how competitive robots can play a role in their routine exercises, such as jogging, and weight training. On the other hand, most subjects (*i.e.*, 71%) who exercise less than three hours per week agreed that a competitive robot partner will increase their engagement with physical exercise. Therefore, our competitive robot is more effective in motivating people who do relatively little exercise. Future research should explore how to effectively apply competitive-HRI to common exercises.

Increased Heart Rate. Most of the gameplay with our competitive robot increased the human subjects' heart rates significantly. Subjects' peak heart rates were higher than their resting heart rates in more than 99% of the games, and their peak heart rates in 92.6% of the games were higher than their walking baseline heart rates. Since the subjects were asked to keep their feet planted on the ground during a game, their body maneuverability was very limited. Despite this movement limitation, subjects' peak heart rates were significantly higher (*i.e.*, 20% to 58%) than their walking baseline in 29% of the games. We found that, other than physical effort, subjects' cognitive effort and reported emotions also corresponded to a rise in heart rate. Figure 4. b. shows that sections with higher average heart rate corresponded to people feeling cognitive demand, motivated, frustrated, and intimidated by the robot. In short, playing competitive games with our robot can be cognitively demanding and can trigger noticeable emotional reactions.

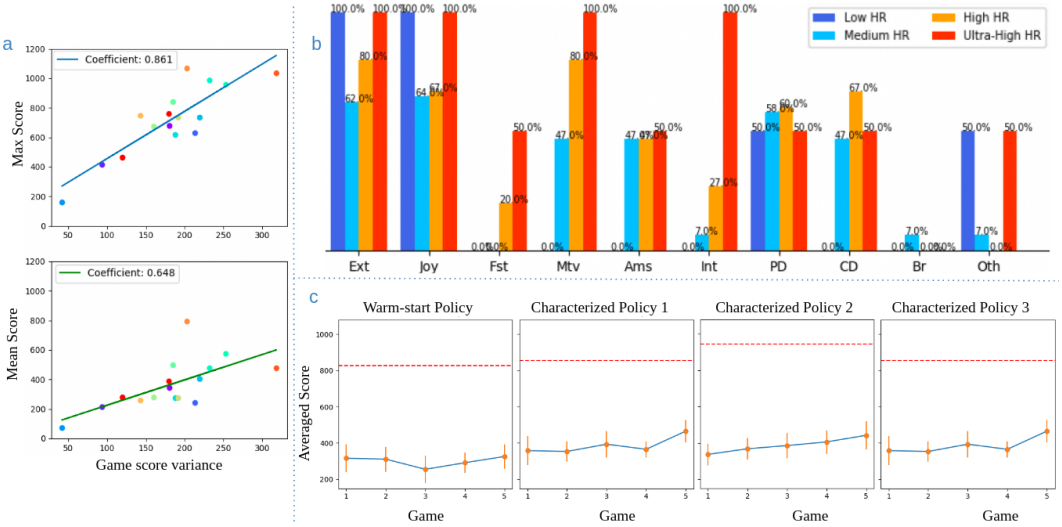

Figure 4: **a.** The 16 points in each subplot represent the 16 subjects. The variance of subjects' game scores is positively correlated to their achieved maximum and mean scores. **b.** Comparison of subjective descriptions between sections with low, medium, high, and ultra-high averaged heart rates. The definition of each group is detailed in Appendix A.5. The x-axis of each subplot shows the adjective describing each section of games (Exciting, Joyful, Frustrating, Motivating, Amusing, Intimidating, Physically Demanding, Cognitively Demanding, Boring, Others). The y-axis indicates the percentage of subjects in the corresponding group who selected the corresponding answer. **c.** Each subplot shows the average game scores of all subjects on the five games against each robot policy. The error bars indicate the standard errors over the samples. The red horizontal lines indicate the best mean score achieved by one of the subjects against the corresponding robot policy.

The Human Learning Effect. As mentioned in Sec. 4, we wanted to test if changing the robot's gameplay style interrupts the human learning process and keeps the interaction challenging throughout the whole experiment. This hypothesis was based on our assumption that most human players can make significant improvement within five games against a fixed robot policy. Surprisingly, this assumption was invalidated by our user study data. Fig. 4. c. shows the average game scores and standard deviation of all subjects in the five consecutive games against each of the four robot policies. An analysis of variance (ANOVA) suggests that subjects have no significant performance increase between the five games for a given policy (Warm-start Policy: $p = 0.96$, Characterized Policy 1: $p = 0.74$, Characterized Policy 2: $p = 0.85$, Characterized Policy 3: $p = 0.43$). The red horizontal dashed line in each subplot of Fig. 4. c. represents the best mean score achieved by a subject, which approximates the performance of an observed best human policy. The average performance of all subjects were lower than (*i.e.*, much lower in most cases) the performance of the best human policies, and most subjects still have much room for performance improvement. Our assumption on human learning was based on evidence from noncompetitive HRI experiments. In contrast, our competitive setting creates a much more dynamical environment and resulted in a more challenging learning environment for the human. Since no significant performance variance was observed across the four sections either (ANOVA $F = 1.51, p = 0.26$), our system was able to

remain challenging to users throughout the whole experiment. However, more studies are needed to analyze how changes in gameplay style interrupt human learning.

Human Performance. Despite the fact that no significant learning effect was found in the subject population, we observed that some subjects tried multiple strategies in the experiment, which could be interpreted as exploration within a learning framework. This observation motivated us to analyze the relationship between strategy exploration and performance, which we first quantified as variance in game scores, and then as the featurized gameplay styles from Sec. 4. As shown in Fig 4. a., a larger variance in score corresponds to better performance in terms of maximum and mean scores. We then compared the variance of the gameplay style between the five subjects with the highest maximum scores and the five subjects with the lowest maximum scores. For the sake of simplicity, we only compared the end-effector displacement change on x, y, and z axes and the averaged velocity. The variance of each selected feature for subjects with high performance were at least 29% higher than subjects with low performance. This suggested that subjects who have high variance in score tend to take risks by constantly exploring different strategies, resulting in better rewards in the long run. Future research could examine whether a robot can help human users achieve better performance by verbally or implicitly encouraging them to explore different strategies.

5.2 User Study Two Result: Baseline Comparison

This experiment compared an RL trained policy (characterized policy 1) to a strong heuristic baseline policy in a long gameplay sequence setting (10 games per policy). Compared to the baseline, the RL trained policy has a slightly better game score performance. In contrast to the first experiment, the human learning effect was observed in both policies. However, the RL trained policy was significantly better at suppressing human subjects from learning to make progress even without switching gameplay styles during the experiment. While the responses to the TAM model were very similar for both policies, the subjects considered the RL policy to be more intelligent because of its "defensive" and "diverse" behavior.

Game Scores: When playing against the baseline policy, the subject population's average game scores (**Baseline mean:** 383.5, **RL mean:** 349.1), the maximum game score (**Baseline max:** 929.0, **RL max:** 744.0), and the minimum game score (**Baseline min:** -291.0, **RL min:** -320.0) are higher than than those against the RL policy. Instead of switching policies every five games, we used a longer sequence of game-plays to evaluate each policy in this study. We were able to find a positive correlation between game scores and the amount of game-play experience against a specific policy. A linear regression over the **Baseline** method data (slope=30.0, coefficient=0.56, p value=9.3e-10) has a larger positive slope, stronger correlation coefficient, and a smaller statistical significant p value compared to that of the **RL policy** (slope=15.7, coefficient=0.34, p value=0.0005).

Subjective Responses: Subjects' responses to most of the modified TAM questions for the two policies are very similar as shown in Table A.5. However, 70% of the population considered the **RL policy** to be more intelligent. In the responses for short questions 2, 3, and 4 in Table 2, the **Baseline** policy was described as "fast" by 2 subjects, "follows my movement" by 4 subjects, and has "repetitive/predictable" behavior by 4 subjects. Meanwhile, the **RL** policy was considered to be "defensive" by 5 subjects, "strategic" by 2 subjects, and to have "diverse behavior" by 2 subjects.

6 Conclusion

This work motivated research in competitive-HRI by discussing how competition can be beneficial to people and the technical difficulties competitive-HRI tasks represent. The Fencing Game, a physically interactive zero-sum game is proposed to evaluate system capability in certain competitive-HRI scenarios. We created a competitive robot using an iterative multi-agent RL algorithm, which is able to challenge human users in various competitive scenarios, including the Fencing Game. Our first user study found that human subjects are very accepting of a competitive robot exercise companion. Our competitive robot provides entertaining, challenging, and intense gameplay experiences that significantly increases the subject's heart rate. In our second user study, one of the policies resulting from the proposed RL method was compared to a strong heuristic baseline policy. The RL-trained policy outperformed the baseline policy in terms of game scores and suppression of the human learning affect, and appeared to be more intelligent to 70% of the population.

References

[1] F. Christiansen and V. Loeschcke. Evolution and competition. In *Population biology*, pages 367–394. Springer, 1990.

[2] I. Ispolatov, E. Alekseeva, and M. Doebeli. Competition-driven evolution of organismal complexity. *PLoS computational biology*, 15(10):e1007388, 2019.

[3] J. H. Dyer and H. Singh. The relational view: Cooperative strategy and sources of interorganizational competitive advantage. *Academy of management review*, 23(4):660–679, 1998.

[4] D. L. Feltz, S. T. Forlenza, B. Winn, and N. L. Kerr. Cyber buddy is better than no buddy: A test of the köhler motivation effect in exergames. *GAMES FOR HEALTH: Research, Development, and Clinical Applications*, 3(2):98–105, 2014.

[5] I. Heazlewood, J. Walsh, M. Climstein, S. Burke, J. Kettunen, K. Adams, and M. DeBeliso. Sport psychological constructs related to participation in the 2009 world masters games. *World Academy of Science, Engineering and Technology*, 7:2027–2032, 2011.

[6] T. Bhattacharjee, E. K. Gordon, R. Scalise, M. E. Cabrera, A. Caspi, M. Cakmak, and S. S. Srinivasa. Is more autonomy always better? exploring preferences of users with mobility impairments in robot-assisted feeding. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pages 181–190, 2020.

[7] A. Bobu, D. R. Scobee, J. F. Fisac, S. S. Sastry, and A. D. Dragan. Less is more: Rethinking probabilistic models of human behavior. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pages 429–437, 2020.

[8] M. Chen, S. Nikolaidis, H. Soh, D. Hsu, and S. Srinivasa. Trust-aware decision making for human-robot collaboration: Model learning and planning. *ACM Transactions on Human-Robot Interaction (THRI)*, 9(2):1–23, 2020.

[9] A. Edsinger and C. C. Kemp. Human-robot interaction for cooperative manipulation: Handing objects to one another. In *RO-MAN 2007-The 16th IEEE International Symposium on Robot and Human Interactive Communication*, pages 1167–1172. IEEE, 2007.

[10] R. Luo, R. Hayne, and D. Berenson. Unsupervised early prediction of human reaching for human–robot collaboration in shared workspaces. *Autonomous Robots*, 42(3):631–648, 2018.

[11] M. J. Matarić, J. Eriksson, D. J. Feil-Seifer, and C. J. Winstein. Socially assistive robotics for post-stroke rehabilitation. *Journal of NeuroEngineering and Rehabilitation*, 4(1):1–9, 2007.

[12] I. Portela-Pino, A. López-Castedo, M. J. Martínez-Patiño, T. Valverde-Esteve, and J. Domínguez-Alonso. Gender differences in motivation and barriers for the practice of physical exercise in adolescence. *International journal of environmental research and public health*, 17(1):168, 2020.

[13] J. H. Van der Lee, R. C. Wagenaar, G. J. Lankhorst, T. W. Vogelaar, W. L. Devillé, and L. M. Bouter. Forced use of the upper extremity in chronic stroke patients: results from a single-blind randomized clinical trial. *Stroke*, 30(11):2369–2375, 1999.

[14] D. L. Gill, L. Williams, D. A. Dowd, C. M. Beaudoin, and J. J. Martin. Competitive orientations and motives of adult sport and exercise participants. 1996.

[15] J. L. Plass, P. A. O'Keefe, B. D. Homer, J. Case, E. O. Hayward, M. Stein, and K. Perlin. The impact of individual, competitive, and collaborative mathematics game play on learning, performance, and motivation. *Journal of educational psychology*, 105(4):1050, 2013.

[16] K. Salen, K. S. Tekinbaş, and E. Zimmerman. *Rules of play: Game design fundamentals*. MIT press, 2004.

[17] S. Hidi and K. A. Renninger. The four-phase model of interest development. *Educational psychologist*, 41(2):111–127, 2006.

[18] M. Viru, A. Hackney, K. Karelson, T. Janson, M. Kuus, and A. Viru. Competition effects on physiological responses to exercise: performance, cardiorespiratory and hormonal factors. *Acta Physiologica Hungarica*, 97(1):22–30, 2010.

[19] J. Schaeffer, N. Burch, Y. Björnsson, A. Kishimoto, M. Müller, R. Lake, P. Lu, and S. Sutphen. Checkers is solved. *science*, 317(5844):1518–1522, 2007.

[20] M. Campbell, A. J. Hoane Jr, and F.-h. Hsu. Deep blue. *Artificial intelligence*, 134(1-2):57–83, 2002.

[21] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, et al. Mastering the game of go without human knowledge. *nature*, 550(7676):354–359, 2017.

[22] W. D. Hillis. Co-evolving parasites improve simulated evolution as an optimization procedure. *Physica D: Nonlinear Phenomena*, 42(1-3):228–234, 1990.

[23] K. O. Stanley and R. Miikkulainen. Competitive coevolution through evolutionary complexification. *Journal of artificial intelligence research*, 21:63–100, 2004.

[24] H. He, J. Boyd-Graber, K. Kwok, and H. Daumé III. Opponent modeling in deep reinforcement learning. In *International conference on machine learning*, pages 1804–1813, 2016.

[25] A. Tampuu, T. Matiisen, D. Kodelja, I. Kuzovkin, K. Korjus, J. Aru, J. Aru, and R. Vicente. Multiagent cooperation and competition with deep reinforcement learning. *PloS one*, 12(4): e0172395, 2017.

[26] D. Silver, T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre, D. Kumaran, T. Graepel, et al. A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. *Science*, 362(6419):1140–1144, 2018.

[27] J. N. Foerster, R. Y. Chen, M. Al-Shedivat, S. Whiteson, P. Abbeel, and I. Mordatch. Learning with opponent-learning awareness. *arXiv preprint arXiv:1709.04326*, 2017.

[28] T. Bansal, J. Pachocki, S. Sidor, I. Sutskever, and I. Mordatch. Emergent complexity via multi-agent competition. *arXiv preprint arXiv:1710.03748*, 2017.

[29] R. Lowe, Y. I. Wu, A. Tamar, J. Harb, O. P. Abbeel, and I. Mordatch. Multi-agent actor-critic for mixed cooperative-competitive environments. In *Advances in neural information processing systems*, pages 6379–6390, 2017.

[30] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra. Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*, 2015.

[31] A. Kshirsagar, B. Dreyfuss, G. Ishai, O. Heffetz, and G. Hoffman. Monetary-incentive competition between humans and robots: experimental results. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 95–103. IEEE, 2019.

[32] B. Mutlu, S. Osman, J. Forlizzi, J. Hodgins, and S. Kiesler. Perceptions of asimo: an exploration on co-operation and competition with humans and humanoid robots. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pages 351–352, 2006.

[33] E. Short, J. Hart, M. Vu, and B. Scassellati. No fair!! an interaction with a cheating robot. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 219–226. IEEE, 2010.

[34] J. Fasola and M. J. Mataric. Robot exercise instructor: A socially assistive robot system to monitor and encourage physical exercise for the elderly. In *19th International Symposium in Robot and Human Interactive Communication*, pages 416–421. IEEE, 2010.

[35] L. Süssenbach, N. Riether, S. Schneider, I. Berger, F. Kummert, I. Lütkebohle, and K. Pitsch. A robot as fitness companion: towards an interactive action-based motivation model. In *The 23rd IEEE international symposium on robot and human interactive communication*, pages 286–293. IEEE, 2014.

[36] K. Sato, K. Watanabe, S. Mizuno, M. Manabe, H. Yano, and H. Iwata. Development of a block machine for volleyball attack training. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1036–1041. IEEE, 2017.

[37] B. Landau, L. B. Smith, and S. S. Jones. The importance of shape in early lexical learning. *Cognitive development*, 3(3):299–321, 1988.

[38] L. Ke, A. Kamat, J. Wang, T. Bhattacharjee, C. Mavrogiannis, and S. S. Srinivasa. Telemanipulation with chopsticks: Analyzing human factors in user demonstrations. *arXiv preprint arXiv:2008.00101*, 2020.

[39] S. Nikolaidis, S. Nath, A. D. Procaccia, and S. Srinivasa. Game-theoretic modeling of human adaptation in human-robot collaboration. In *Proceedings of the 2017 ACM/IEEE international conference on human-robot interaction*, pages 323–331, 2017.

[40] M. L. Littman. Markov games as a framework for multi-agent reinforcement learning. In *Machine learning proceedings 1994*, pages 157–163. Elsevier, 1994.

[41] N. Heess, D. TB, S. Sriram, J. Lemmon, J. Merel, G. Wayne, Y. Tassa, T. Erez, Z. Wang, S. Eslami, et al. Emergence of locomotion behaviours in rich environments. *arXiv preprint arXiv:1707.02286*, 2017.

[42] B. Baker, I. Kanitscheider, T. Markov, Y. Wu, G. Powell, B. McGrew, and I. Mordatch. Emergent tool use from multi-agent autocurricula. *arXiv preprint arXiv:1909.07528*, 2019.

[43] S. Liu, G. Lever, J. Merel, S. Tunyasuvunakool, N. Heess, and T. Graepel. Emergent coordination through competition. *arXiv preprint arXiv:1902.07151*, 2019.

[44] F. Martinez-Gil, M. Lozano, and F. Fernandez. Emergent behaviors and scalability for multi-agent reinforcement learning-based pedestrian models. *Simulation Modelling Practice and Theory*, 74:117–133, 2017.

[45] K. Tuyls, J. Perolat, M. Lanctot, J. Z. Leibo, and T. Graepel. A generalised method for empirical game theoretic analysis. *arXiv preprint arXiv:1803.06376*, 2018.

[46] S. Wold, K. Esbensen, and P. Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987.

[47] T. L. Chen, T. Bhattacharjee, J. M. Beer, L. H. Ting, M. E. Hackney, W. A. Rogers, and C. C. Kemp. Older adults' acceptance of a robot for partner dance-based exercise. *PloS one*, 12(10): e0182736, 2017.

[48] P. Mertikopoulos, C. Papadimitriou, and G. Piliouras. Cycles in adversarial regularized learning. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2703–2717. SIAM, 2018.

A Appendix

COVID-19 Effects. We were fortunate to be able to perform an in-person experiment on a real-world robot and human subjects, yet the subject recruiting process was particularly challenging during the current COVID pandemic. With limited access to potential experiment participants, we were only able to run a pilot test with two people prior to the actual experiment reported here. Solely from the pilot study data, we were not able to disqualify our eventual experimental assumption that there is a significant learning effect among participants even within only five games. Instead, as described in Sec. 5.1, it was not until obtaining our full experimental results that we were able to discard this assumption due to observing a large variance in the degree of human learning across participants. Future research that aims to understand the human learning effect in competitive-HRI tasks should collect more data for each individual and facilitate the participant's ability to learn. In particular, increased learning can be achieved by reducing environmental complexity, such as decreasing the robot's joint velocity, reducing the manipulator's reachability in task space, using a robot with a lower degree of freedom (DoF) and so on.

A.1 The Fencing Game

Algo. 2 summarizes the scoring mechanism for the Fencing Game described in Sec. 2.

Algorithm 2: The Fencing Game Scoring Mechanism

```
Initialize: Game score s = 0; Timestep = 0.01 Sec; Game horizon = 20 Sec
bat_a → Antagonist's bat
bat_p → Protagonist's bat
target → Target Area
for every timestep in this game do
    if bat_a in target then
        if bat_a contacts bat_p then
            s -= 10
        else
            s += 1
        end
    if bat_p in target for more than 200 consecutive timesteps then
        s += 10
end
```

A.2 System Implementation Details

The PR2 robot has two 7 degree of freedom (DoF) arms and its overall form-factor is similar to a human adult. It is comparable to a human player in a competitive game in terms of body size and arm flexibility. The human player's bat is attached to an infrared photo-diode array tracker, and its position and orientation is perceived by the robot via two tracking base stations. An audible scoring feedback system is created to report the scoring situation of each game in real-time. The participants will hear a higher frequency (440 Hz) signal sound when scoring, and a lower frequency (300 Hz) signal sound when receiving penalty from the robot. The system implementation pipeline is shown in Fig. 5.

Algorithm Details. This paragraph provides extra technical details on the two-phase iterative training algorithm described in Sec. 4. There are two major differences between phase one and two training in Algo. 1. **(1)** Phase one training uses a continuous reward to facilitate agents' development of basic motor skills. In phase two training, the agents are solely rewarded by the game scores. **(2)** In phase one training, each agent is always learning against the latest version of the opponent. But in phase two training, an agent would constantly and randomly load a previous version of opponent from the history, after a short period of training. The use of continuous reward encourages the robot to quickly explore the task space, which makes the phase one training good for quickly initializing the policies for the antagonist and protagonist. However, continued used of the iteration strategy in phase one training would likely trap both agents in a low quality local equilibrium and/or chasing each other in circles in parameter space in a long sequence of training [48]. In contrast, the reward and iteration mechanisms of phase two training create a high variance learning environment

which effectively mitigates the circling problem. In addition, because of the high variance nature of the phase two training, agents are more likely to learn emergent behaviors and converge to more sophisticated policies.

Training Details. As shown in Fig. 2, at the beginning of the training (i.e., phase 1 - itr 1) the game scores tend to be largely biased toward whichever agent is currently learning. This is because both agents' policies are simple/naive at the beginning phase and the opponent can easily find a counter strategy to dominate the game during learning. After the warm-start training (i.e., phase 1 - itr 2), the two agents converged to an area in their policy space where both agents are challenging each other without completely dominating the game.

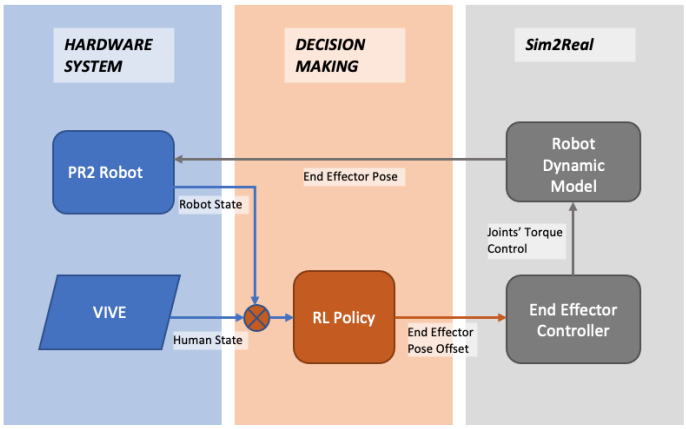

Figure 5: A block diagram demonstrating the pipeline of the proposed robotic system.

A.3 User Study

Demographics There were 16 human subjects (10 male, 6 female, of age $M = 28.8$ years, $SD = 5.56$) recruited for the first human-subject studies. Nine out of 16 subjects reported doing more than three hours of physical exercise per week, and seven subjects reported less than 3 hours of weekly exercise. Jogging, walking, cardio, and weight training were the most common exercises chosen among all subjects. Ten subjects from this population participated the second user study.

Before the Experiment. Before the experiment, each subject was asked to sit for 3 minutes and then walk for 1 minute to record two average heart rate baseline values. In the experiment, both the robot and a subject held a polystyrene bat to play the games. A safety line was drawn on the ground, and every subject stood behind it to prevent a potential collision. Since the robot was not mobile, each subject also kept his/her feet planted on the ground during a game. The target area was not directly visible; an audible scoring feedback system notified a subject when his/her bat was placed within the target area. A high-frequency signal sound indicated that the human player was scoring, and a low-frequency signal sound indicated that the human player was getting penalization. Before the experiment, a subject had 5 minutes to explore this target area with their bat so that the target area's location and the scoring mechanism were clear to the subject. Afterward, the subject played two warm-up games with the robot to get further familiarized with the system. In these warm-up games, the robot's actions were slowed down and no data was collected.

Experiment Procedure for User Study One. The experiment contains four sections. In each section, a subject will play five consecutive games with the robot (with a fixed robot policy) and rest for approximately 30 seconds between games. Subjects' heart rates will be recorded during each 20 second game. Due to the short duration of the games, our discussions in Sec. 5.1 use peak heart rate as a summarizing statistic of this recorded data. In the first section of the experiment, the robot will use the warm-start policy resulting from the phase one training. For the rest of the sections, the robot will load one of the three selected characterized policies in each section by following a random order. This ensures that all subjects will first learn to play with the robot at a regular speed, and then play with the robot with a new game style in each of the sections. After each section, a subject will

be asked to describe the interaction in the last 5 games by selecting one or more of the following adjectives: 'Exciting', 'Joyful', 'Frustrating', 'Motivating', 'Amusing', 'Intimidating', 'Physically Demanding', 'Cognitively Demanding', 'Boring', 'Others (please describe:)'. Finally, when a subject finishes all four sections of games, they will complete the last part of the questionnaire that assesses their acceptance of the competitive robot, and their subjective feelings towards the games. We modified the technology acceptance model (TAM) [47] and created the questions in Table. 2. We introduced two extra questions (*i.e.*, DE and IE) to understand the human subjects' acceptance and desirability of a competitive robot companion in the future, and if a competitive robot can motivate them to engage in physical exercise more frequently. Other than the open-ended question, all TAM questions were measured on a 5-point scale where 1 = "Strongly Disagree," 3 = "Neutral," 5 = "Strongly Agree". At the end of each gameplay section, a subject is also asked to consider and compare both the enjoyability and the difficulty of the games between the finished sections. One section can be equally, less or more enjoyable and difficult than another section. This will allow the participant to have two rankings of the four sections based on their perceived enjoyment and difficulty by the end of the experiment.

Experiment Procedure for User Study Two. This experiment contains two sections. In each section, a subject is asked to play 10 consecutive games with the same robot policy and rest for approximately 15 seconds between games. The order in which each subject plays against the baseline policy and characterized RL policy is randomized. After each section, a subject is asked to answer the modified TAM questions. After the final section, a subject is also asked to answer short questions 2, 3, and 4 in Table. 2.

Baseline Heuristic Policy. We aimed to design a strong baseline heuristic policy to create an intense human robot gameplay experience. Given an observation of the world, the robot orients its bat perpendicular to the human's bat with random angular offsets drawn uniformly from -25 to 25 degrees on the x, y, and z axes. In order to ensure that the robot is always executing a competitive defense, the policy commands the robot to position the center of its bat in between the target area and the point on the human's bat that is closest to the target area:

$$\bar{b_p} = \bar{tar} + (\bar{h_{close}} - \bar{tar}) \cdot uniform(0.5, 1)$$
$$\bar{h_{close}} = \bar{h_{low}} + ht \cdot (\bar{h_{up}} - \bar{h_{low}})$$
$$ht = \max\left(0, \min\left(1, (\bar{tar} - \bar{h_{low}}) \cdot (\bar{h_{up}} - \bar{h_{low}})/L_{sword}\right)\right)$$

Where $\bar{b_p}$, $\bar{tar}$, $\bar{h_{up}}$ and $\bar{h_{low}}$ represent the position of the robot's bat frame, the center of the target area, the upper end of human's bat, and the lower end of human's bat respectively. $\bar{h_{close}}$ indicates the point on the human's bat that is closest to the center of the target area, and L_{sword} indicates the length of a bat. The function $uniform(0.5, 1)$ randomly determines how far apart the robot's bat should be from the human's bat. In addition, there is a 50% chance for the robot to execute the desired bat position calculated from the last time step instead of the latest desired pose. The added uncertainties introduce randomness to the robot's behavior. This heuristic allows the robot to dominate the fencing game when it can move faster or as fast as the antagonist. However, human subjects are able to move slightly faster than our PR2 robot, which leaves room for human subjects to discover counter strategies.

A.4 Heart Rate

Fig. 4. b. compares the subjective descriptions between four groups of gameplay sections with different levels of average human heart rates. For each section in the user study, we first calculate the average peak heart rate over the corresponding five games in the section. A section's heart rate level l is calculated by dividing the section's average peak heart rate by the corresponding user's walking baseline heart rate, which results in a percentage value describing how much more or less the average section heart rate is compared to the baseline. The low, medium, high, and ultra-high heart rate groups contain the sections that $l \leq 100\%$, $100\% < l \leq 120\%$, $120\% < l \leq 140\%$, and $140\% < l$ respectively.

	Questions
Perceived Usefulness (PU)	Having a competitive robot companion would improve the quality of my physical exercise.
Perceived Ease of Use (PEOU)	Learning to earn higher score (make progress) in the games with a competitive robot would be easy for me.
Attitude (ATT)	Using a competitive robot exercise partner to improve my exercise quality is a good idea.
Intention to Use (ITU)	Assuming I had access to a competitive robot for exercise, I would intend to use it.
Perceived Enjoyment (PENJ)	I would find competitive human-robot gameplays are entertaining.
Desirability (DE)	Based on your experience today, future physical exercises and games with a competitive robot will be desirable.
Increased Engagement (IE)	Having a competitive robot companion would make me more likely to engage in physical exercise.
Short Question 1 (used in study one)	Are there anything you would like to change to improve the interaction experience?
Short Question 2 (used in study two)	Which robot (part 1, part 2, or equally) do you think is more challenging/difficult to play against? and why?
Short Question 3 (used in study two)	Which robot (part 1, part 2, or equally) do you think is more enjoyable/fun to play against? and why?
Short Question 4 (used in study two)	Which robot (part 1, part 2, or equally) do you think is more intelligent? and why?

Table 2: Modified Technology Acceptance Model and Open-ended Questions

	Baseline Median	RL Median	t-statistic	p-value
PU	3.6	3.6	0.0	1.0
PEOU	3.9	3.6	0.85	0.41
ATT	4.1	3.9	0.41	0.69
ITU	3.7	3.6	0.2	0.84
PENJ	3.9	3.9	0.0	1.0
DE	3.7	3.9	-0.45	0.65
IE	3.0	3.5	-0.9	0.37

Table 3: TAM Response Comparison Between Baseline Policy and RL Policy in Pair-T Tests

A.5 Experimental Results

Perceived Ease of Use. Interestingly, although we did not observe significant performance improvement within each section nor between sections, 62.5% of subjects perceived that it was easy for them to make progress when playing against the robot. Furthermore, some participants expressed a desire to beat the best score of previous participants. It is possible that some subjects focused only on attaining their perception of a "high" score for a small number of games rather than maintaining good performance in all 20 games. On the other hand, in a competitive game setting, we are still not sure exactly what it means for people to feel that getting a better score would be easy. In our experiment, it probably suggested that most participants considered the robot opponent to be surmountable because only a very small number of participants described their gameplay experience as "Frustrating", "Intimidating", or "Boring". However, future work could study how the perceived ease of use affects the participants' effort in competitive games.

Enjoyment and Difficulty. We found that perceived enjoyment and difficulty do not significantly vary as the amount of player experience increases, but the amount of data used to train the corresponding policy does have an affect. Fig. 6 shows the average ranking comparison for enjoyability and difficulty across both policies and experiment sections. Among the utilized policies, no significant variance was observed across the three characterized polices for both enjoyability($p = 0.40$)

and difficulty($p = 0.35$). Paired-T tests show that the warm-start policy is less enjoyable and difficult than characterized policies ($p < 0.01$), except for the enjoyability of the warm-start policy and policy 2 ($p = 0.13$). The result is similar in the time domain - no significant variance was found in the last three sections, in which characterized policies were randomly ordered (enjoyability: $p = 0.5$, difficulty: $p = 0.42$). Section one, which only uses the warm-start policy, is less enjoyable and difficult than other sections($p < 0.05$). Across all sections, the perceived difficulty is positively correlated to perceived enjoyment with a moderate coefficient of 0.6.

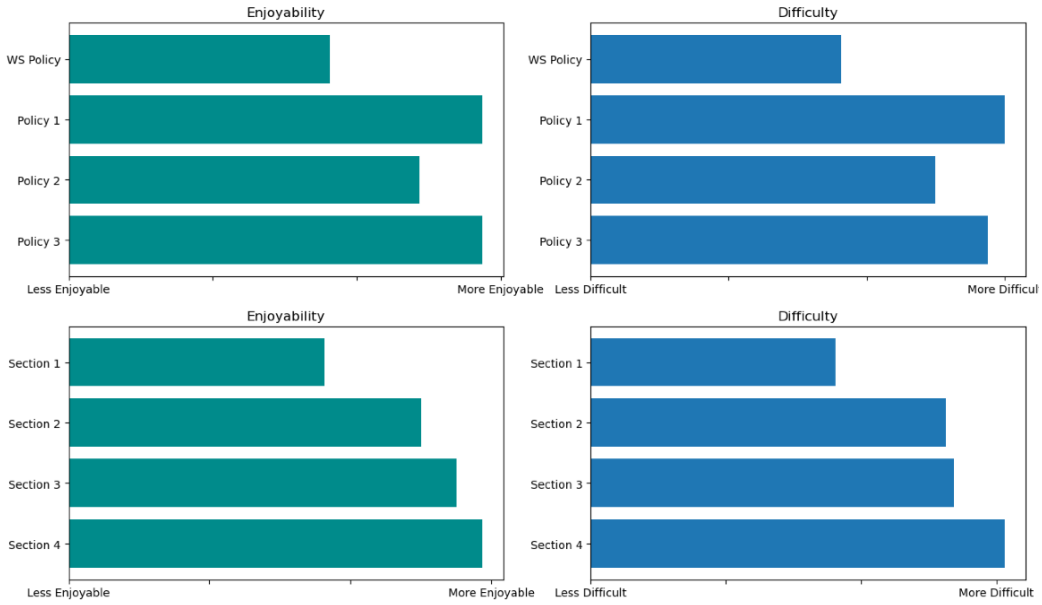

Figure 6: The average ranking comparison for enjoyability and difficulty across both policies and experiment sections.