Competitive Physical Human-Robot Game Play

Boling Yang,[1] **Xiangyu Xie**[2] **Golnaz Habibi,** [3] **Joshua R. Smith** [1, 4]

[1] Paul G. Allen School of Computer Science & Engineering,
University of Washington
[2] Department of Mechanical Engineering,
University of Washington
[3] Department of Aeronautics and Astronautics,
Massachusetts Institute of Technology
[4] Department of Electrical & Computer Engineering,
University of Washington
bolingy@cs.washington.edu, shawnxie@uw.edu, golnaz.habibi@gmail.com, jrs@cs.washington.edu

Abstract

While competitive games have been studied extensively in the AI community for benchmarking purposes, there has only been limited discussion of human interaction with embodied agents under competitive settings. In this work, we aim to motivate research in competitive human-robot interaction (competitive-HRI) by discussing how human users can benefit from robot competitors. We then examine the concepts from game AI that we can adopt for competitive-HRI. Based on these discussions, we propose a robotic system that is designed to support future competitive-HRI research. A human-robot fencing game is also proposed to evaluate a robot's capability in competitive-HRI scenarios. Finally, we present the initial experimental results and discuss possible future research directions.

Introduction

Competition is one of the most common forms of human interaction, yet *competitive interaction* has rarely been discussed in the context of Human Robot Interaction (HRI). There has indeed been a large focus in HRI on *cooperative interaction*, such as human-aware motion planning, object handover actions, and collaborative manipulation. Conversely, the absence of studies in competitive robot interaction may be due to anxieties concerning the actions of robots whose interest do not necessarily align with our own. However, these fears should not prohibit us from considering positive impacts that competitive-HRI can yield, such as providing the participants with motivation, inspiring their potential, and more.

We initiate our research project in competitive-HRI by focusing on how to create a competitive robotic agent that can challenge human users in athletic performance and physical exercise. Physical exercise is essential to our physical and mental health. We hypothesize that a robot with adversarial behaviors can provide athletic practice or exercise sessions that are more personalized, effective, and enjoyable. We believe that athletic practice and physical exercise are scenarios in which nearly anyone can directly benefit from competitive-HRI. In this paper, we first motivate competitive-HRI by discussing psychological studies about how competition can influence participants in a posi-

Figure 1: A robot-based athletic training system based on a PR2 robot and a VIVE VR system. The human and the robot are playing a competitive fencing game. Detailed rules of the game can be found under the System Design & Implementation section. The upper figure shows the gameplay in the VR environment, where the left sub-window shows the game from a third-person perspective, and the right sub-window shows the human's perspective when wearing the VR headset. The lower figure shows the game in actuality. Please refer to this link for example gameplay videos.

tive manner and how physical exercise is more efficient and motivating when an appropriate amount of competition is introduced. We then survey the literature of reinforcement learning in competitive games, and discuss the challenges in competitive-HRI and algorithms that addresses these issues. Next, the design and implementation of our competitive-HRI robotic system are introduced, along with a competitive fencing game we designed to evaluate our system. Finally, we will present initial experimental results and discuss the insights drawn from the results.

Competition

A number of psychological research studies indicate that competition can effectively increase participants' performance, motivation, and ability to learn in various scenarios. Plass et al. (2013) compared how individual, cooperative, and competitive game modes affect learning, performance, and motivation in an educational mathematics video game. They found that players in both competitive and cooperative conditions solve significantly more problems during the game than those in the individual condition. Furthermore, competitive players may have developed better arithmetic skills in the game, such that they were able to complete more problems than those in other conditions. The quality of problem-solving strategies was also affected by different game modes, where competitive and individual players were less likely to employ inefficient problem-solving strategies. The study also found that competitive and collaborative games enhance enjoyment and situational interest. Higher enjoyment refers to a participant having a higher intrinsic desire and tendency to engage in an event (Salen, Tekinbaş, and Zimmerman 2004). Furthermore, higher situational interest refers to an event that is able to elicit more attention and effective reactions from the participants (Hidi and Renninger 2006).

The effect of competition in physical exercise and athletic training has also been well studied. Feltz et al. (2014) created a virtual exercise partner that always slightly outperforms the human participant in cycling or holding a plank exercise. Without knowing it was a manipulated virtual peer, the participants felt less capable than the peer and showed performance improvement during the study. Viru et al. (2010) explored the mechanisms of how exercise performance can be enhanced under competitive conditions. They found that, in a treadmill running test, the running duration was prolonged by an average of 4.2% in competition, and was accompanied by a significantly greater peak VO_2 (maximum rate of oxygen consumption) response. Inspired by these studies, this project aims to create a robotic exercise partner that is able to challenge a human user.

Related Work

There has been very limited human-robot interaction research that explores competitive interactions. Kshirsagar et al. (2019) studied how a human's performance was effected by a robot "co-worker" working in the same workspace when competing for a real monetary prize. Human participants were slightly discouraged when competing against a high performance robot. Another observation was that people would hold a more positive attitude toward a robot with lower performance. Mutlu et al. (2006) compared the perceptions of an ASIMO robot when it was playing a video game cooperatively or competitively with human participants. Their results suggested that male participants were more engaged by competitive gameplay, but the cooperative agent was more socially desired. Short et al. (2010) found that, when a robot cheated in a "rock-paper-scissors" game, human participants had a greater degree of social engagement and made greater attributions of mental state during the game.

On the other hand, there have been some attempts to use robotic systems to assist in physical exercise. Fasola and Mataric (2010) developed a socially assistive robot to provide real-time coaching and encouragement for a seated arm exercise. Most participants did find the robot to be helpful in their exercise and considered it to be an exercise instructor. Süssenbach et al. (2014) created an interactive action-based motivation model for an indoor-cycling activity. In order to motivate the user throughout the exercise section, the robot employed communication strategies according to the user's physical state and condition. Their system successfully increased users' workout efficiency and intensity. Sato et al. (2017) created a robotic system to assist the training of volleyball players. The system has sufficient mechanical capability to imitate the motion and strategy of top volleyball blockers.

Each of these studies focused on one or a very small set of competitive scenarios. In addition, most of these scenarios only require very limited and simple robot motions. In this work, we examine the concepts from game AI that we can adopt for competitive-HRI. Furthermore, we design a robotic system that can potentially play various physically competitive games against a human player.

Modeling Competitive-HRI As Games

In this section, we first formulate competitive-HRI tasks as a multi-agent Markov games problem. With such a problem setting in mind, we then survey the literature on competitive games and discuss the existing frameworks and algorithms that can be applied to competitive-HRI tasks. However, the game representation doesn't capture every challenge in competitive-HRI. Therefore, we will also discuss the additional challenges introduced by the embodied agent.

Multi-agent Markov Games

In this paper, we only consider 2-player zero-sum game scenarios in which one human user is competitively interacting with a single robot. In order to formalize the representation for the subset of competitive-HRI problems that we are interested in, we choose to consider them as multi-agent Markov games(Littman 1994). A Markov game between a human and a robot is a partially observable Markov decision process defined by a tuple $(\mathcal{S}, \mathcal{O}^h, \mathcal{O}^r, \mathcal{A}^h, \mathcal{A}^r, \mathcal{T}, \mathcal{R})$. Here, $\mathcal{S}$ is a set of states describing the state of the game. $\mathcal{O}^h$, $\mathcal{O}^r$, $\mathcal{A}^h$, and $\mathcal{A}^r$ are the sets of observations and actions of the human and robot respectively. A transition function

$\mathcal{T} : S \times \mathcal{A}^h \times \mathcal{A}^r \rightarrow \mathcal{S}$ maps the state and actions to a subsequent state. Assuming r_t^r and r_t^h are the instantaneous reward at time t for the robot and human, $r_t^h = -r_t^r$ in the zero-sum game setting. The human player will try to maximize his/her long term reward $\mathcal{R}$ for a finite time horizon T, $\mathcal{R} = \sum_{t=0}^{t=T-1} \gamma^t r_t^h$, while the robot will try to minimize it.

RL Algorithms for Markov Games

Under the multi-agent Markov Games setting, we will discuss the existing solutions and their applicability to competitive-HRI. Creating a transition model for human players via actual human demonstration data is hard. Creating a transition model can be extremely resource consuming, and the resulting models are most likely not transferable from one task to another. Training the robotic player with model-based or supervised learning methods is not the most appropriate approach, because these methods require access to a comprehensive state-action transition model. Therefore, we will focus on reviewing reinforcement learning (RL) methods that do not require massive real-world datasets.

The topic of Multi-agent Markov games is well-explored. Since games like Chess, Checkers, and Go provide an interactive environment with a clear scoring system, researchers can use these environments to benchmark the ability of algorithms to train an agent to learn, reason, and plan (Schaeffer et al. 2007; Campbell, Hoane Jr, and Hsu 2002; Silver et al. 2017). Many algorithms that solve multi-agent Markov games are designed under a multi-agent reinforcement learning scheme, where agents develop emergent and complex behavior through interacting with each other and co-evolving together. Hillis (1990) was an early experiment with competitive co-evolution. Stanley and Miikkulainen (2004) experimented with evolutionary strategies in a simple 2D competitive environment. He et al. (2016) used a deep Q neural network to model both the transition function and the opponents' policy in competitive games. Tampuu et al. (2017) applied deep-Q learning to train agents to play the Pong game in both cooperative and competitive settings. Silver et al. (2018) achieved superhuman performance in the game of Go, chess, and shogi by reinforcement learning from competitive self-play. The LOLA (Foerster et al. 2017) and LOLA-DiCE (Foerster et al. 2018) algorithms are policy gradients methods that update the policies by calculating gradients over the parameter space of both the current training agent and all other opponents. These two algorithms can achieve better learning stability in a competitive environment. However, they assume the dimension of all agents' parameter space are identical, and all agents have access to all the other agents' parameters, which limit their applicability in many cases.

Although the methods above can achieve great performance in certain scenarios, they have only been tested in board games, video games, 2D particles, or similar environments. These environments usually have a much simpler state transition mechanism compared to the competitive-HRI problem we are interested in. Because of the kinematic and dynamic complexity of a robot and the human body, their transition functions are highly non-linear. Furthermore, controlling a robot requires a model to learn motor skills under complex dynamics. The following works have successfully extended game RL to learn complex motor skills. Pinto et al. (2017) proposed an adversarial training method that can be applied to most existing RL algorithms. Their algorithm relies on a simple iterative training mechanism, which makes it very easy to implement. However, this algorithm can suffer from convergence and stability problems that commonly exist in multi-agent learning problems (Mertikopoulos, Papadimitriou, and Piliouras 2018; Mazumdar, Jordan, and Sastry 2019). Through the use of PPO and a very large training batch size, Bansal et al. (2017) created humanoid and quadrupedal agents that play physically competitive games like soccer, wrestling, and more in simulation. Lowe et al. (2017) proposed a centralized action-value function that takes the actions of all agents as input in order to stabilize the DDPG algorithm (Lillicrap et al. 2015) in multi-agent settings. All three approaches use neural networks to model the complex environment changes, and they all update their policy according to the actor-critic framework. In this paper, we will evaluate two of these approaches and compare their strengths and weaknesses under our competitive-HRI environment. Nevertheless, a traditional game representation does not capture the challenges of using embodied agents, which we will discuss in the next paragraph.

Challenges Outside of Game Frameworks

While reinforcement learning algorithms for games are often designed for virtual agents acting in simulated environments, controlling an embodied agent to interact with human beings can be much more challenging. First, these reinforcement learning methods explore and exploit the policy space by generating a large number of rollouts in simulation. Yet, because of the mismatch of model parameters between the simulation and the real world, a policy that performs well in simulation may perform poorly in reality. In addition, compared to a virtual environment, the embodied agent will typically have noisier observations, lower control precision, and larger uncertainty. Finally, a robot needs to prioritize the human user's safety at all times when interacting with them competitively, an issue that does not arise in pure simulation settings.

System Design & Implementation

We created a robotic system to explore whether a robot can effectively improve human athletic training under competitive conditions. Some of the major challenges of competitive-HRI problems were raised in the last section. In this section, we will address these challenges and propose possible solutions for our system. A preview of the system pipeline is shown in Fig.3.

Evaluation Environment

A zero-sum fencing game environment was created to train and evaluate our agents. It is an attack and defense game where the human is the antagonist and the robot is the protagonist. A screenshot of the fencing game in the Mujoco simulation environment is shown in Fig. 2. Because

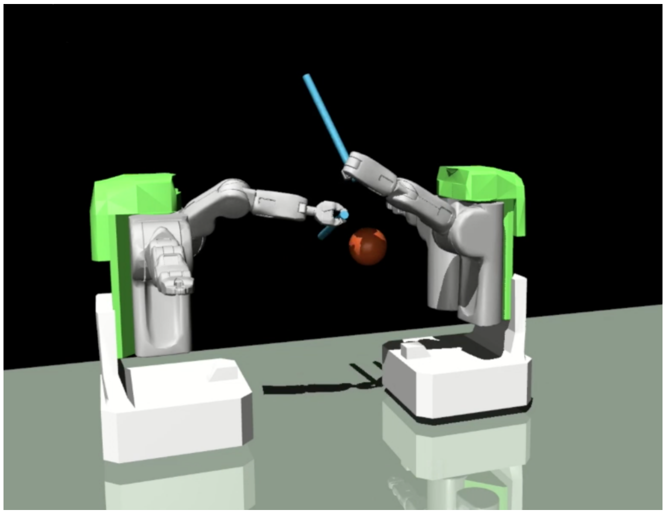

Figure 2: The fencing game in Mujoco simulation environment. The left agent simulates the human antagonist player, and the right agent simulates the robot protagonist player.

the overall form-factor of a PR2 robot is similar to a human being, both agents are represented by a PR2 model in the simulation. The antagonist agent on the left scores by placing its bat within the orange spherical(target) area located between the two agents. The antagonist's score will increase by 1 for every 0.01 seconds that its bat is placed within the target area without contacting the opponent's bat. However, the antagonist will receive -10 points of score deduction if its bat is placed within the target area and makes contact with the protagonist's bat simultaneously. Meanwhile, the protagonist agent on the right's goal is to minimize the antagonist's score in a game. Moreover, the protagonist will lose the game immediately if its bat is placed within the target area for more than 2 seconds. Each game will last for 20 seconds. The observation space for both agents is represented by the following tuple:

$$\mathcal{O}^h = \mathcal{O}^r = (p_h, p_r, v_h, v_r, t)$$

Where p_h and p_r are the Cartesian pose of the bat frame for the human player and robot player respectively. v_h and v_r describe the velocity humans' and robots' bat. t specifies the game time in seconds.

Hardware Components

The physical system is designed and built based on a PR2 general-purpose robotic research platform. The PR2 robot has two 7 degrees of freedom(DoF) arms, and its overall form-factor is similar to a human adult. It is therefore a suitable embodied agent for physical interaction with a human. All joints of the robot support 100Hz real-time control, allowing the robot to quickly respond to human actions. An HTC VIVE VR headset and two controllers serve two purposes in this system: First, the use of the headset and controllers allows the robot to easily track the pose of the human's head and hands. Second, VR technology provides an immersive experience to users, and it is commonly used in HRI research that requires a particular interaction scene (Li et al. 2019; Matsas and Vosniakos 2017). By synchronizing the actual robot's location and joint angles to the PR2 model within the VR environment, human users will be able to see the robot's behavior in real-time. Consequently, this system has the advantages of both the real and virtual environment, where actual physical contacts, robot motions, and ambient sound can be perceived by the human user, yet, the experimental environment can be easily and quickly modified based on the requirements of different tasks.

Learn to Play Games

We evaluated the performance of two multi-agent actor-critic algorithms trained in the fencing game environment. In the first approach, we combined the iterative training structure from (Pinto et al. 2017) and the classic PPO algorithm(Schulman et al. 2017). The detail of this approach is presented in Algorithm 1.

Algorithm 1: Iterative Two-Agent Training

Input: Environment $\mathcal{E}$; Stochastic policies μ and ν
Initialize: Parameters θ_0^μ for μ and θ_0^ν for ν
for $i = 1, 2, ..N_{iter}$ **do**
  $\theta_i^\mu \leftarrow \theta_{i-1}^\mu$
  for $j = 1, 2, ..N_\mu$ **do**
    rollout $\leftarrow roll(\mathcal{E}, \mu_{\theta_i^\mu}, \nu_{\theta_{i-1}^\nu}, N_{traj})$
    $\theta_i^\mu \leftarrow$ PPO_Update(rollout)
  end
  $\theta_i^\nu \leftarrow \theta_{i-1}^\nu$
  for $j = 1, 2, ..N_\nu$ **do**
    rollout $\leftarrow roll(\mathcal{E}, \mu_{\theta_i^\mu}, \nu_{\theta_i^\nu}, N_{traj})$
    $\theta_i^\nu \leftarrow$ PPO_Update(rollout)
  end
end

The protagonist's parameters θ_i^μ will first be trained by collecting trajectories that result from playing against the antagonist with a static policy. This continues until the protagonist's policy achieves good performance against the antagonist's current policy. The antagonist will then be trained against the protagonist with a static policy in order to find a policy ν with parameters θ_i^ν that the protagonist's policy is not robust to. This training sequence is repeated for N_{iter} iterations. Both agents are updated by optimizing the PPO clipped surrogate objective. Since this method could suffer from stability problems when N_{iter} is large, we ran the algorithm with $N_{iter} = 2$, and receive a robot agent that can play the fencing game sufficiently well.

In the second approach, we directly apply the algorithm in (Bansal et al. 2017) to our environment. This method also trains both agents in an iterative scheme. Nevertheless, when training each of the agents, instead of having such an agent face against the latest opponent's policy, this algorithm proposed to randomly load a previous version of the opponent's policy from history. When using a very large batch size, this algorithm demonstrated better stability during a long sequence of training.

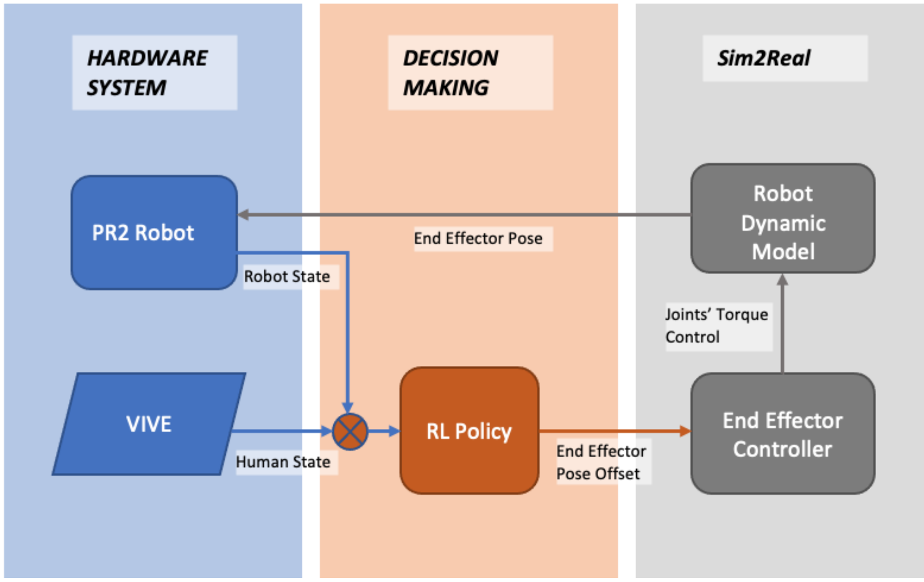

Figure 3: A block diagram demonstrating the pipeline of the proposed robotic system

Sim2Real

System identification(systemID) and domain randomization techniques are typically used to handle the kinematic and dynamic mismatch between simulation and actuality. Since domain randomization introduces extra variance to our learning task, it could worsen stability and convergence issues that already exist in multi-agent environments. Our empirical study shows that the combination of a Jacobian Transpose end-effector controller and systemID provide a great Sim2Real solution for our PR2-based system. Instead of specifying the torque values for each joints, the policy outputs an offset or end-effector pose to be 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 damping, armature, and friction loss for all joints of the simulated robot. θ_c represents the proportional gains and derivative gains of the end-effector controller. T_r and T_s are two trajectories sampled from the real robotic system and simulation 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. This systemID process tunes end-effector dynamics of the simulated robot such that it is similar to the dynamics of a real PR2 robot. We set an upper bound for the controller's torque output to prevent possible human injury. To reduce the amount of computation, only the protagonist agent that represents the actual PR2 robot uses the end-effector controller.

Initial Result & Discussion

In this section, we will first discuss the problems that arose during our development and what we have achieved in this project so far. While we are still preparing a number of experiments to evaluate our system and gaining further insight into competitive-HRI, some initial results with the fencing game environment will be presented. In the end, we will propose some possible future research direction from both algorithmic and HRI perspectives.

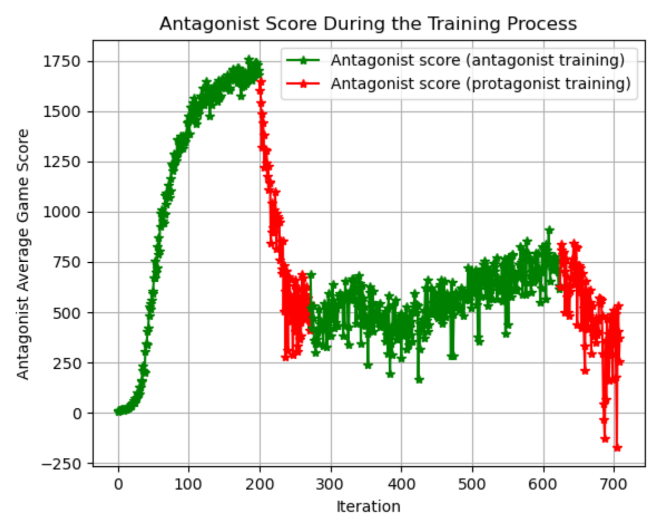

Figure 4: Antagonist score in the fencing game during the iterative training in Algorithm 1. The antagonist agent is trained first, and $N_{iter} = 2$.

For the policy training, we were able to create a robotic agent with reasonable performance by using Algorithm 1. Fig. 4 shows the antagonist's score throughout the iterative training process. The inclining green curves indicate the antagonist learns to score higher while avoiding the penalization from contacting the protagonist's bat. The declining red curves indicate the protagonist learns to minimize the antagonist's score. Each iteration is stopped when convergence is detected or a timeout occurs. The algorithm from (Bansal et al. 2017) provides more stable training in a longer training sequence, and the resulted policies are seemingly more sophisticated comparing to those resulted from Algorithm 1. Due to the time limitation, we will compare and evaluate the two learning methods in our future works.

Although we are still experimenting and improving the RL learning process, the other parts of the robotic system are ready for experiments. We performed an initial experiment with a policy that has been trained with Algorithm 1. In this experiment, one human participant was asked to play five consecutive games with the PR2 robot. Table 1 presents the participant's score and average heart rate in each game.

	Game 1	Game 2	Game 3	Game 4	Game 5
Human Score	-541	503	329	830	754
Average Heart Rate (BPM)	124.9	129.0	144.6	150.8	144.5

Table 1: Game scores and average heart rate of the human participant during five consecutive competitive human-robot fencing games.

The human participant got a negative score in the first game because he made contact with the robot's bat frequently. However, the participant was able to quickly make adjustments and scored increasingly higher in the consecutive games. Meanwhile, the participant's heart rate also increased when playing the games with the robot. The participant has a baseline average heart rate of 83 BPM when resting and 111 BPM when walking. The participant's faster heart rates during the competitive games indicate that he exerted greater physical effort compared to the two baseline scenarios.

Since this project is still in the development stage, there are still many interesting questions that need to be explored and answered. From the RL perspective, finding algorithms that can stably create more sophisticated control policies for competitive-HRI problems is essential. On the other hand, our initial experiment data shows that the human participant's performance was gradually increasing during the games. Creating a robot agent that can learn from small amounts of real world data and quickly improve its policy can help the human's performance improve faster. Moreover, it will be useful if the robot can help a participant achieve certain training goals. From the user experience perspective, it is important to understand what aspects of competitive-HRI are enjoyable to human users, and what can possibly results in negative emotions.

References

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

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

Fasola, J.; and Mataric, M. J. 2010. 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*, 416–421. IEEE.

Feltz, D. L.; Forlenza, S. T.; Winn, B.; and Kerr, N. L. 2014. 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.

Foerster, J.; Farquhar, G.; Al-Shedivat, M.; Rocktäschel, T.; Xing, E.; and Whiteson, S. 2018. DiCE: The Infinitely Differentiable Monte Carlo Estimator. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, 1524–1533. Stockholmsmässan, Stockholm Sweden: PMLR. URL http://proceedings.mlr.press/v80/foerster18a.html.

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

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

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

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

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

Li, R.; van Almkerk, M.; van Waveren, S.; Carter, E.; and Leite, I. 2019. Comparing human-robot proxemics between virtual reality and the real world. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 431–439. IEEE.

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

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

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

Matsas, E.; and Vosniakos, G.-C. 2017. Design of a virtual reality training system for human–robot collaboration in manufacturing tasks. *International Journal on Interactive Design and Manufacturing (IJIDeM)* 11(2): 139–153.

Mazumdar, E. V.; Jordan, M. I.; and Sastry, S. S. 2019. On finding local nash equilibria (and only local nash equilibria) in zero-sum games. *arXiv preprint arXiv:1901.00838* .

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

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

Pinto, L.; Davidson, J.; Sukthankar, R.; and Gupta, A. 2017. Robust adversarial reinforcement learning. *arXiv preprint arXiv:1703.02702* .

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

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

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

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

Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; and Klimov, O. 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347* .

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

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

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

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

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

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

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