Neural Computing and Applications (2023) 35:16739–16752
https://doi.org/10.1007/s00521-022-07746-9

S.I.: HUMAN-ALIGNED REINFORCEMENT LEARNING FOR AUTONOMOUS AGENTS AND ROBOTS

Incorporating rivalry in reinforcement learning for a competitive game

Pablo Barros[1] · Özge Nilay Yalçın[2] · Ana Tanevska[1] · Alessandra Sciutti[1]

Received: 21 January 2022 / Accepted: 18 August 2022 / Published online: 11 November 2022

Abstract

Recent advances in reinforcement learning with social agents have allowed such models to achieve human-level performance on certain interaction tasks. However, most interactive scenarios do not have performance alone as an end-goal; instead, the social impact of these agents when interacting with humans is as important and largely unexplored. In this regard, this work proposes a novel reinforcement learning mechanism based on the social impact of rivalry behavior. Our proposed model aggregates objective and social perception mechanisms to derive a rivalry score that is used to modulate the learning of artificial agents. To investigate our proposed model, we design an interactive game scenario, using the Chef's Hat Card Game, and examine how the rivalry modulation changes the agent's playing style, and how this impacts the experience of human players on the game. Our results show that humans can detect specific social characteristics when playing against rival agents when compared to common agents, which affects directly the performance of the human players in subsequent games. We conclude our work by discussing how the different social and objective features that compose the artificial rivalry score contribute to our results.

Keywords Reinforcement learning · Rivalry · Competitive learning · Multi-agent reinforcement learning

1 Introduction

The social aspects of interaction are usually overlooked when optimizing an artificial agent through reinforcement learning [1]. Most of the training loop is done in an offline manner, or focuses on optimizing objective metrics that do not directly involve social aspects, for instance by using planners [2] or human annotation feedback [3]. Some common success metrics in this regard involve solving the task in fewer steps, reducing predicted values, or achieving some predefined intermediate objective goals. When interaction with humans is the main goal, these artificial agents are evaluated mostly based on their objective performance [2]. In the few examples where humans are present in the loop, the success measures are mostly related to the embodied interaction [4, 5], and not to the underlying decision-making process that these agents learned.

One scenario where these problems are very evident is competitive interaction. In a competitive game, an agent can learn to adapt toward its opponents by using reinforcement learning [6], even when these opponents are humans [7]. However, it is extremely difficult to measure the social aspects of this interaction, without relying on typical human–robot or human–computer interaction schemes [8]. Although providing important insight on some social aspects, these evaluations usually focus on controlled lab-scenarios [9] and on the production of different robotic behaviors [10] and dialogues [11]. This in turn neglects exploring how the agents' various learning strategies influence their explicit behaviors and their

✉ Pablo Barros
pablo.alvesdebarros@iit.it

Özge Nilay Yalçın
oyalcin@sfu.ca

Ana Tanevska
ana.tanevska@iit.it

Alessandra Sciutti
alessandra.sciutti@iit.it

[1] CONTACT Unit, Italian Institute of Technology, Genova, Italy

[2] School of Interactive Arts and Technology, Simon Fraser University, Vancouver, Canada

interaction with humans [12], despite it being one of their most important characteristics. This can be evidenced even in the new area of explainable reinforcement learning [13, 14].

In this study, we address the problem of including social aspects in the learning strategies of artificial agents in a competitive scenario. We propose an objective human-centered metric based on rivalry [15], to compose the reward function of the agents. Rivalry is a subjective social relationship arising between two actors, based on the competitive characteristics of an individual, as well as the increasing stakes and psychological involvement in the situation [16]. We chose rivalry as it showcases the competing relation between individuals, which often affects their motivation and performance during gameplay [16, 17]. We model rivalry as a function of objective factors (such as game performance) and subjective information (such as certain personality traits and competitiveness level), and evaluate our model using the Chef's Hat Card Game [18]

To obtain the social features of rivalry and map the intrinsic personality traits arising from human perception of learning agents, we run first an exploratory experiment where human players face artificial agents implemented using deep Q-learning (DQL) [19] and proximal policy optimization (PPO) [20]. Both learning agents are implementing COPPER [21], a continual learning adaptation for Chef's Hat agents. Using questionnaires, we collect how these agents impact the human players in terms of competitiveness, and how humans perceive the social characteristics of these agents.

Using the compiled information from this experiment, we measure how humans perceive such agents in terms of rivalry. We then attach different social characteristics to each artificial opponent and use them in a rivalry synthesizing mechanism, to calculate the rivalry of an agent toward an opponent. We then run two ablation studies to develop a social characteristic predictor that will be used by the agents both when perceiving their opponents, and for finding the best manner to integrate rivalry in the learning routines.

Finally, we run a second experiment, where each artificial agent synthesizes a rivalry score against human players. By collecting the same information from questionnaires as in the first experiment, we can contrast the impact of the rival agents on the game when compared with non-rival agents.

Our results demonstrate that both learning agents are perceived as different social agents when playing against a human, in particular when compared with the random agent. When the learning is modulated by the rivalry score, we observe a strong contribution of rivalry on the performance of the human players. We discuss these results in terms of the contribution of the social and objective features for the formation of the rivalry function, and how they impact human perception. Ultimately, we detail how the performance of each agent changes when using the rivalry modulation.

2 Related work

Reinforcement learning has received ample attention in the last years, in particular on the development of artificial agents. However, understanding the social role and the derived impact of social interaction within a learning mechanism is not yet fully explored. In particular, in multi-agent competitive scenarios, there is still a focus on performance-based metrics, which make it difficult to summarize, or even to verify, the social components which are directly affected when these models are deployed in scenarios involving humans. In this section, we detail the most relevant literature in this field, and the ones we based our entire explorations on.

2.1 Reinforcement learning in competitive games

In the late 1990s, several researchers tried to identify the impact of the Deep Blue artificial chess player [22] on the development of artificial intelligence [23, 24]. They all argue that beyond the technical challenge of beating a human, there is an underlying impact on how this agent affects the opponents' behavior during the entire interaction. Over time, these investigations were let aside by the mainstream community, which focused mostly on solving more complex problems. This vision is reflected by the recent development of deep reinforcement learning and the research on training artificial agents to play competitive games that have flourished since [25]. AlphaGo [26] demonstrated that these agents can play competitively against humans in very complex games. The recent development of agents that play the StarCraft computer game [7] pushes these boundaries even further. These agents learn how to adapt to dynamic environments, how to map hypercomplex states and actions, and how to learn new strategies [27]. Most of the studies, however, focus on the final goal of these agents: how to be competitive against humans. As such, none of them focus on understanding the impact that these agents have on human opponents.

2.2 DQL and PPO playing behavior on the Chef's Hat card game

In the same development wave, it was recently investigated the design and development of reinforcement learning agents to play the four players Chef's Hat competitive card game [28]. These agents were based on deep Q-learning (DQL) [19] and proximal policy optimization (PPO) [20], and achieved success in learning how to win the game in different tasks: playing against random agents, self-play, and online adaptation toward the opponents. However, it was observed that these agents present different behavior during gameplay, while maintaining a similar objective performance measured by overall wins over a series of games. What has not yet been done is to measure the social impact that such strategies can have when the agents play against humans.

2.3 Human-centric analysis of RL

When analyzing the impact of artificial agents on humans, there are now decades of studies focused on Human–Robot [29] and Human–Computer [30] Interaction (HRI and HCI, respectively). Social and affective computing research suggests humans tend to treat computers as social actors, [31] where attributes such as personality and emotions are modeled to affect how these agents are perceived [32, 33]. In an interaction setting, these attributes can be used to change the behavior of humans or improve interaction in different contexts [34, 35]. In RL, most of these studies however focus on optimizing RL agents to solve a specific task, even social ones, without having much feedback on the social aspects of the task as part of their learning mechanism. Such agents are usually designed to learn an expected outcome, such as improving engagement [4], or imitating humans [6]. None of the most recent studies focus on extracting the intrinsic behavior bias that different learning schemes apply to the final agent.

2.4 Rivalry in cooperative and competitive games

One way to explain the behavior of an agent as a factor of its learning strategy is to measure its impact on humans. In cooperative scenarios, there exist several social metrics that take interaction into consideration [36, 37], but most of them focus on the subjective impressions humans have of the embodied interaction [38], the subjective quality of the interaction [39], or the efficiency of the interaction to solve a task [40].

In a competitive game however, one of the most informative metrics is the rivalry [15] between the human and the agents. Rivalry is defined as a competitive relation between individuals or groups, characterized by the subjective importance placed upon competitive outcomes (i.e., win or lose) independent of the objective characteristics of the situation (e.g., tangible stakes) [16, 17]. A proposed theoretical model of rivalry suggests that antecedents of rivalry are similarity factors, competitiveness, and relative performance of the agents [16]. The presence of a rivalry effect, in turn, affects the motivation of the individual and their performance [17]. We aim to evaluate how different agents affect the user's perception and their performance due to the increased competitiveness and rivalry effects.

In competitive games, rivalry is a central concept that directly affects the opponent's behavior through their motivation in play. In human-to-human scenarios and economics, a healthy rivalry is considered to be an important factor that can positively affect the performance of opponents, while in other situations, it can also contribute to unnecessary risk-taking behavior [41].

In human-in-the-loop online learning scenarios for competitive games, the absence of rivalry or competitiveness might result in the human opponent to lose motivation to play the game and show sub-optimal performance during gameplay. The agents who are learning actively from the human are bound to learn from this bad performance input, which in turn would result in sub-optimal learning. By introducing the notion of rivalry, we expect to increase motivation and engagement in the game and thus also the user performances.

3 Proposing artificial rivalry

Our rivalry modulation acts directly on two types of agents: a deep Q-learning (DQL) one and a proximal policy optimization (PPO) one. Both agents were recently adapted and optimized for the Chef's Hat game through the COPPER modulation [21]. COPPER introduces an opponent-specific experience-prioritizing memory used to improve the continual learning capabilities of each agent when playing against known opponents. We use these agents' implementation as our artificial opponents, and apply the rivalry modulation to their learning mechanism. It is important that our agents continually update their playing strategy when playing against the human opponents, so the rivalry score is created and updated accordingly.

3.1 A Chef's Hat agent

The DQL and PPO implementations of the agents were chosen due to their success on learning different strategies [28], and their good performance when playing against human players [21]. Both agents are implemented as COPPER-based agents, and are set to keep learning during all of our experiments.

The Chef's Hat game, illustrated in Fig. 1, is a multiplayer competitive card game. At the beginning of the game, each player receives 17 cards, and the player that discards all of them first wins the match. The entire rules of

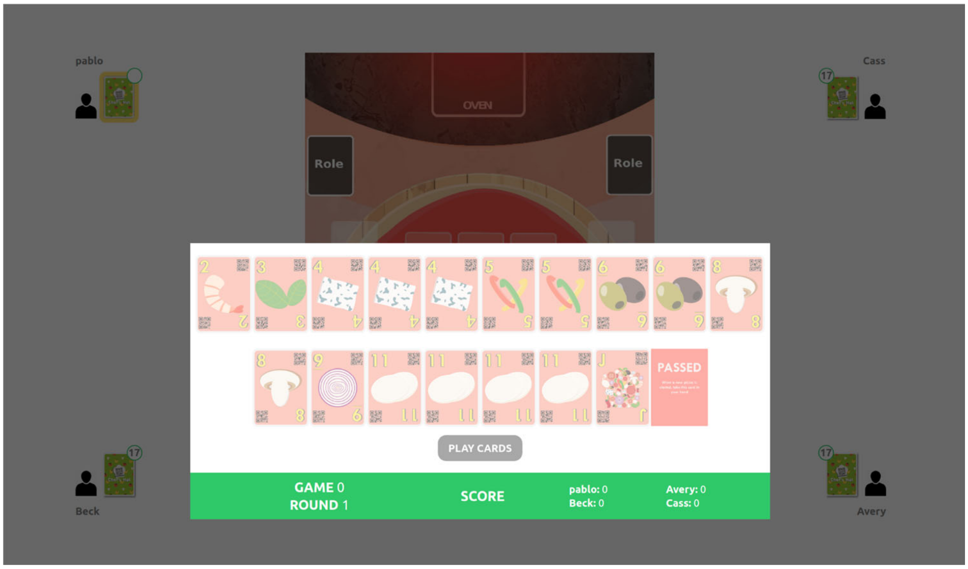

Fig. 1 Illustration of the Chef's Hat card game environment used for all our experiments. Available at: https://github.com/pablovin/ChefsHatGYM

the game and the capability of agents to learn different strategies were recently explored in different studies [18, 21, 28]. The game state is composed of 28 values, referring to the 17 possible cards each player has in their hands, and 11 cards on the game board. The action space is represented by 200 different discard actions an agent can do, which reflects the complexity of the strategy formation in this game.

3.1.1 Deep Q-learning

Our deep Q-learning (DQL) agent implements a Q function as:

$$Q : SxA \rightarrow \mathbb{R} \tag{1}$$

where S is the state, in our case represented by the 28 values composed by the cards at hand and the cards at the board. The actions, A, are expressed using the 200 discrete values for all the possible actions.

To update the Q-values, the algorithm uses the following function:

$$Q'(s_t, a_t) = Q(s, a) + \alpha \times (\mathrm{TD}) \tag{2}$$

where t is the current step, α is a pre-defined learning rate and TD is the temporal difference function, calculated as:

$$\mathrm{TD} = r_t \times \gamma \times \max Q\left(s_{t+1,a_t}\right) - \max Q\left(s_{t1,a_t}\right) \tag{3}$$

where r_t is the obtained reward for the state (s_t) and action (a_t) association, γ represents the discount factor, a modulator that estimates the importance of the future rewards, and $\max Q\left(s_{t1,a_t}\right)$ is the estimate of Q-value for the next state.

This agent also implements a target model, which is a time-delayed policy which receives a snapshot of the original policy after a certain number of training steps.

The target model is used to obtain the target Q learning when calculating the updated TD:

$$\mathrm{TD} = r_t \times \gamma \times \max Q\left(s_{t+1,a_t}\right) - \max Q_t\left(s_{t1,a_t}\right) \tag{4}$$

where $\max Q_t\left(s_{t1,a_t}\right)$ is the Q-values obtained from the target network.

3.1.2 PPO

Our proximal policy optimization (PPO) [20] agent implements an actor-critic model:

$$A(s_t, a_t) = r_t + \gamma \times V(s') - V(s) \tag{5}$$

where $V(s)$ represents the critic value for a given state, and $V(s')$ for the future state. The advantage function is used to stabilize training the actor-network, while the critic network uses the discounted rewards as the target.

The actor-critic base model is updated by implementing adaptive penalty control, based on the Kullback–Leibler divergence, to drive the updates of the agent at each interaction. It showed to be important to learn different and more optimized strategies [28], probably due to how this method simplifies the necessity of having a large-memory replay [42, 43].

3.2 COPPER

Both agents implement the COPPER modulator [21] that expands the prioritize experience replay (PER). Traditional PER can be expressed as:

PER(i) is calculated based on the network's loss after calculating TD in a forward pass of the network (using an input i):

$$\mathrm{PER}(i) = \frac{p_{\mathrm{a}}^{i}}{\sum_k p_{\mathrm{a}}^{k}} \tag{6}$$

where a indicated how much we want to rely on the priority, p is the priority, and k the total number of saved experiences. COPPER, on the other hand, introduces a new opponent-specific term (o):

$$\mathrm{COPPER}(i) = o\frac{p_{\mathrm{a}}^{i}}{\sum_k p_{\mathrm{a}}^{k}} \tag{7}$$

In recent experiments, COPPER was shown to be much more effective on learning new strategies against recurrent opponents, in particular when playing against humans [21].

3.3 Modeling rivalry from a human perception

To optimally define the impact each agent has on the players, we used a standard formalization of rivalry [16]. Rivalry can be defined as a subjective social relationship arising between two actors based on the competitive characteristics of an individual, as well as the increasing stakes and psychological involvement in the situation. Thus, a proposed theoretical model of rivalry, illustrated in Fig. 2, suggests that antecedents of rivalry are similarity factors (S_{a}), competitiveness (C), and relative performance (P) of the agents [16].

Rivalry research suggests that individuals tend to evaluate their abilities by comparing their performance to persons who have similar characteristics to themselves. The similarity can be measured in demographics [16], gender [44], personality [16], perceived traits [45] and rank in competition [17] to assign social or behavioral attributes to the other.

For this work, in addition to the ranking and performance similarity, we included trait similarity as a factor consisting of competence, agency, and communion traits as the indicators of relevant behavioral attributes. These traits are frequently used in social stereotypes research (e.g., gender, nationality, age, status) [46] and have been shown to affect liking [47]. Moreover, social perception [48] has been previously shown to affect competitive behavior [49]. Agency, competence, and communion traits have also been used in virtual agent research to examine users' judgments of expected agent behavior [50]. These traits were chosen as representative ones due to their comprehensibility, which facilitates the self and other assessment [51]. We collect the agency (ag), competence (ct), communion (cm), and competitiveness (C) assessments of each of the agent players, from a human perspective, through an exploratory experiment. Once these values are calculated, we can define a rivalry score as (R_{a}):

$$S_{\mathrm{a}} = \sqrt{(\mathrm{ag_h} - \mathrm{ag_a})^2 + (\mathrm{ct_h} - \mathrm{ct_a})^2 + (\mathrm{cm_h} - \mathrm{cm_a})^2} \tag{8}$$

$$C = C_{\mathrm{h}} \tag{9}$$

$$P_{\mathrm{a}} = (\mathrm{points_h} - \mathrm{points_a})/15 \tag{10}$$

$$R_{\mathrm{a}} = (S_{\mathrm{a}} + C_{\mathrm{h}} + P_{\mathrm{a}})/3 \tag{11}$$

where the index a defines one of the opponents, the index h defines the human ratings about the social behavior of the opponents, obtained in a prior human study. The individual performance of each player is calculated by the Chef's Hat environment[1] as the sum of the average score of each game (given by the sum of the points in each round, divided by the number of rounds in the game), divided by the total number of played games.

The rivalry score (R_{a}) is used as our main evaluation metric on how the agents impact the human players, thus we expect it to change accordingly to the game development. As the participants evaluate the entire game behavior of an agent, and the only difference among artificial players is the way they play the game, the measure of rivalry reflects directly the user's perception of the outcome of the reinforcement learning strategies.

3.4 Rivalry as a learning modulator

Once rivalry can be defined from a human perspective, we need to model it from an agent perspective, and use it as a learning modulator. This will make the agents create a rivalry sense against the other players, and most important, will help the agents to develop a behavior that can make it be identified by a human as a rival.

To achieve a rivalry modulator, we calculate rivalry in a similar manner as we do from a human perspective. To predict similarity, though, the agents used a similarity predictor trained based on the humans' responses collected during our exploratory experiment. The agent-player similarity predictor ($\mathrm{pr_a}(h)$) matches state + action chosen by humans (h) with the human-assessed competence, agency and communion traits ($\mathrm{ag_a}$, $\mathrm{ct_a}$ and $\mathrm{cm_a}$), and it follows the

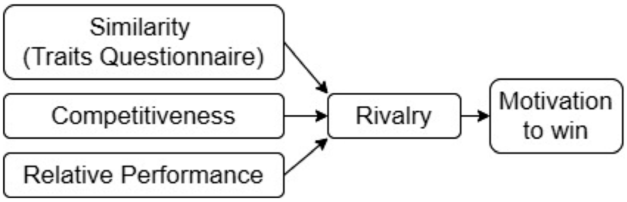

Fig. 2 The theoretical model of rivalry used for our new rivalry metric proposition, based on the framework proposed by [16]

[1] https://github.com/pablovin/ChefsHatGYM.

survey scales used in virtual agent research in terms of calculation of the scores for each trait [50].

Our similarity predictor was built as a multi-layer perceptron (MLP) neural network mapping an interval of action/spaces with a set of similarity scores. The similarity predictor of each agent inferred during gameplay the traits of each human the agents played against. Each type of agent (DQL and PPO) also have a single set of traits associated, derived from the human judgments provided in the exploratory study. Thus, the similarity based on an agent's (a) perspective (S_{h}) is given as:

$$S_{\mathrm{h}} = \left(\sqrt{(\mathrm{pr_a}(h) - (\mathrm{ag_a}, \mathrm{ct_a}, \mathrm{cm_a}))^2}\right) \tag{12}$$

The performance measure of the agents was given by their own assessment of their actions. To achieve this, each agent computed the introspective confidence (ic_{a}) [52] of each action, which focused on scaling the selected Q-value of an action toward the final goal, using a logarithm transformation which computes the probability of success. In our case, this was the probability of winning the game. The introspective confidence gives us a self-assessment of the agent's actions, based on its own game experience. We use the introspective confidence as the agent's competitiveness:

$$C_{\mathrm{a}} = \frac{\sum ic_{\mathrm{a}}(\mathrm{act})}{\mathrm{totalActions}} \tag{13}$$

where act, is each of the actions the agent took during the game.

The relative performance (P_{h}) is calculated similarly to the human's perspective but considers the agent's perspective. Thus, the predicted rivalry (R_{h}) is defined as the mean of the previous three factors:

$$R_{\mathrm{h}} = (S_{\mathrm{h}} + C_{\mathrm{a}} + P_{\mathrm{h}}). \tag{14}$$

To guarantee the agent learns how to be a rival of an opponent, we include the predicted rivalry (R_{h}) into the final reward of the agent by a simple weighted average, optimized in an ablation study described below.

4 Experimental setup

Our rivalry modulation is directly related to a very specific reinforcement learning task: multi-agent competitive interaction. In this regard, the Chef's Hat Card game was chosen as our primary investigation environment. As our experiments involve a good mix of human-based studies and artificial agents optimization, we separate them into four categories: first, we perform our exploratory study to understand how the PPO and DQL agents are perceived by humans. Our second experiment uses the information collected from the exploratory study to train and validate the similarity predictor neural network. The third experiment optimizes and evaluates the rivalry learning modulator. And finally, we run a human-based study to identify the impact of the rival agent on human perception.

4.1 Chef's Hat environment

The Chef's Hat environment is an OpenAI GYM-based implementation of the Chef's Hat card game[2]. It includes all the rules and mechanics of the original game and can be used to train and evaluate artificial agents. The game itself is based on turns, and on each turn a player can do a discard action or a pass action. For every match played, the players gain points based on their finishing position, with the winner gaining 3 points. A full game consists of several matches until one of the players reaches 9 points.

Following our previous experiments with the Chef‘s Hat environment, we use a global reward scheme that only gives a full reward once the agent performs an action that leads to it winning the game. We also start the game with pre-trained agents, which learn how to play the game using a self-play strategy [28] available through the Chef's Hat Player's Club repository[3].

To collect human data, and to allow humans to play against the agents, we use the Chef's Hat Online software[4], which is a browser-based interface for the Chef's Hat Environment and allows human players to participate in the game.

4.2 Exploratory study: understanding the agents

In our first experiment, we perform an exploratory study using the Chef‘s Hat Online environment. The goal of this study is twofold: we want to understand and measure the impact of each learning agent on the rivalry attribution. Also, we collect the human attributions to these agents, and use them to describe both agents socially, when synthesizing rivalry.

For this study, we implement three agents, a DQL, a PPO, both with COPPER, and a naive agent that only performs random actions during the entire game. A human plays a 9-points game against these agents, and we collect the entire game status over the entire experiment, which includes all players' performance. We also run two questionnaires, one at the beginning of the game to collect the human players' self-assessment, and one by the end of the game to investigate the participants' perspective about the agents.

[2] https://github.com/pablovin/ChefsHatGYM.

[3] https://github.com/pablovin/ChefsHatPlayersClub.

[4] https://github.com/pablovin/ChefsHatOnline.

In our questionnaire, we follow the standard traits questions to measure competence, agency, and communion [46]. In particular, we measure competence score as an average of the scores of the items related to ambition, courage, decisiveness, and aggressiveness; agency as an average of intelligence, innovation, organization, and compassion; and communion as an average of compassion, affection, emotional response, and sensitiveness. Each of these terms is represented with a Likert scale that varies from 1 (Not at all) to 5 (Very).

We also asked humans to self-assess their competitiveness and the perceived competitiveness of each agent, once the game was over. Each agent was only identifiable by one of three names (Evan, Dylan and Frankie), in order to make the entire evaluation based on their game behavior alone. The questionnaires are available on our Appendix. In total, 28 different persons played the game.

Using the collected information, we calculate the rivalry score from the human perspective for each agent. We then proceed with a statistical test to identify the contribution of each of the terms that compose rivalry (similarity, relative performance and competitiveness).

4.3 Rivalry ablation: similarity predictor

The goal of this experiment is to achieve a reliable similarity predictor, to be used by the agents. We implement it as a fully-connected multi-layer perceptron (MLP) neural network, that receives as input a sequence of pairs of action and cards on board, and predicts a similarity label. We fine-tune it and evaluate it using the data obtained from the previous exploratory study.

To optimize this network we used a Tree-Parzen Optimizer (Hyperopt [53]), and a cross-validation with 70% of the collected data for training and 30% for testing. We report the optimization search-space and the final architecture in the Appendix Section A.3. We calculate the performance in terms of mean accuracy over 30 runs.

4.4 Rivalry ablation: rivalry learning modulator

To maximize the effect of the rivalry modulator in the agent's learning, but without losing the focus on winning the game, we run an optimization study to find the best weight when adding rivalry in the agent's reward function. Our rivalry score is defined in a way that it would increase when an agent plays against itself, as the social behavior, performance, and game strategy of the agents are the same. So, we optimize the rival weight toward the reward by maximizing the rival score while maintaining a similar performance against each other.

We run 1000 simulation games, where an agent plays against three other agents, among which one is another instance of itself. During all the games, we add the rival score weight as an updatable parameter of the network, and optimize it toward the opponent that has the same implementation of the agent. We stop the training when the rival score of both agents' instances toward each other is maximized. We track the rivalry evolution over time, together with the agents' performance, to guarantee that the optimization reaches a satisfactory state.

4.5 Rivalry impact: playing against humans

Once the rival agents are implemented, and we found the optimal reward weight, we deploy them into the same scenario as our first exploratory experiment. We repeat the same settings, but now the human plays the game against a rival agent, a non-rival COPPER agent that continues learning and a non-rival agent that does not learn during the game. In this experiment, as we are interested on the differences within these agents, we only implemented a DQL-based agent. The goal of this experiment is to verify the impact of the rivalry in terms of perceived game play. We collect the same data, using the same questionnaires, and calculate the rivalry score of all the participants toward the agents.

We then run statistical tests to identify the impact of the rival agent in the game, in particular when compared to the non-rival agents. This experiment aims to identify the contribution of rivalry in terms of social perception, but also on the final performance of the agents.

5 Results

5.1 Exploratory study

A total of 28 games were completed and were therefore included in all subsequent evaluations in this experiment. Of these games 13 were played in English, 14 in Portuguese, and 1 in Italian. All of them, however, played using the same Chef's Hat Online platform, so following the same game rules. Of the participants who played these games, 54% were between the ages 31 and 50, 36% reported to be between 18 and 30, 7% reported to be aged over 50, and one participant chose not to disclose their age.

On average, each game was played for 3.34 matches (SD = .47), with an average score of 2.25 (SD = .71). We calculated the rivalry scores for each participant toward each of the three agents, based on the proposed equations (see Sect. 3.3), by using the Similarity, Competitiveness and Performance measures.

Similarity scores were computed using Eq. 9 (see Sect. 3.3). An analysis of the results showed that agents were rated as significantly different in terms of their overall similarity scores with participants (Friedman test: $\chi^2 = 6.82$, $df = 2$, $p = 0.033$). Further analysis using Durbin multiple comparisons [54] showed that the random agent was perceived significantly different than the DQL ($p = 0.009$) agent. Moreover, there were significant differences in the similarity scores for particular items in the traits questionnaire, namely "Decisive" (Friedman test: $\chi^2 = 6.40$, $df = 2$, $p = 0.041$) and "Innovative" ($\chi^2 = 6.05$, $df = 2$, $p = 0.049$) traits, and scores showed a tendency to differ for "Creativity" ($\chi^2 = 5.43$, $df = 2$, $p = 0.066$).

The relative performance scores stand of the three agents showed significant difference ($\chi^2 = 28.1$, $df = 2$, $p < 0.001$), where both the DQL (Durbin multiple comparisons, $p < 0.001$) and PPO ($p < 0.001$) agents were significantly different from the random agent. No significant difference was seen between the DQL and PPO agent ($p = 0.23$). Finally, we used the self-attributed values for the competitiveness scores for all the users ($M = .89$, SD = .17). Rivalry scores calculated using these three measures were significantly different among the agents ($\chi^2 = 25$, $df = 2$, $p < 0.001$). Pairwise comparisons revealed that both DQL ($p < 0.001$) and PPO ($p < 0.001$) agents were significantly different than the random agent. No difference was found between the DQL and PPO agents ($p = 0.791$).

5.2 Similarity predictor

In our First Experiment, we collected a total of 16,000 action/cards on board pairs, and associated them with the "Decisive", "Innovative" and "Creative" labels self-assessed (from humans), or given by (from the agents) during the human data collection, using the average of them as our final similarity score. We report the optimized search-space and the final architecture in Table 1.

The final performance of the similarity predictor after running 30 cross-validation routines, when trained with the data collected from humans and agents, is 83% accuracy, with a standard deviation of 0.5.

Table 1 Search space and final architecture, in bold, used to optimize the similarity predictor

Parameter	Search space
Concatenated actions	[3,**5**,10, 15]
Number of layers	[1,**2**,3]
Units per layer	[16, 32, **64**, 128, 256, 512, 1024]

5.3 Rivalry learning modulator

After running 1000 games, the final optimized weighted value for the reward is as follows:

$$R = R_{\mathrm{o}} + 0.2 * R_{\mathrm{h}} \tag{15}$$

where R_{o} is the original reward of the agent, and R_{h} is the rivalry score. We also tracked the performance of the agent over time. Without rivalry, when playing in the same setting, a DQL agent obtained an average of 1.3 (game score), with a standard deviation of 0.1. When the rivalry modulation was used, the average wins stayed at 1.2, with a standard deviation of 0.2.

5.4 Rivalry impact

In this experiment, a total of 25 games were completed. From the final set of the completed games, 9 were played in English, 9 in Portuguese, 3 in Spanish and 4 in Italian. From the participants who played these games, 55% reported to be aged between 31–50 and the remainder reported to be between 18–30.

On average, the games were played for 3.65 matches (SD = .48), and with an average human score of 3.06 (SD = .68) which is significantly higher than the participants' scores from Study 1 ($t(46) = -3.98$, $p < 0.001$). Similar to the previous study, the rivalry scores for each participant were calculated using the proposed equations in Sect. 3.3, by using the Similarity, Competitiveness and Performance measures.

An analysis of the similarity scores, computed using Eq. 9 (see Sect. 3.3), showed that the agents were not rated as significantly different in terms of their overall similarity with participants (Friedman test: $\chi^2 = 1.57$, $df = 2$, $p = 0.457$). The relative performance scores of the three agents showed significant differences ($\chi^2 = 23.7$, $df = 2$, $p < 0.001$), where Durbin–Conover multiple comparisons test revealed that all agents performed significantly differently than each other. Pairwise comparisons showed the rival DQL agent performed the best, being significantly better than both COPPER DQL ($p = 0.036$) and offline DQL agents ($p < 0.001$). The COPPER DQL agent was also significantly better ($p < 0.001$) than the DQL agent.

Finally, the rivalry scores calculated using these two measures and the self-attributed competitiveness scores of the participants ($M = .79$, SD = .18) were significantly different among the agents ($\chi^2 = 15.7$, $df = 2$, $p < 0.001$). Pairwise comparisons revealed that the rival DQL agent had significantly higher rivalry scores compared to both the COPPER DQL ($p = 0.036$) and DQL ($p < 0.001$) agents. Further, the COPPER DQL agent had significantly higher rivalry scores ($p = 0.009$) than the DQL agent. However, the participants' motivation to play with agents were not significantly different among different types of agents ($\chi^2 = 0.2133$, $df = 2$, $p = 0.9$). Table 2 shows the values for the relative performance and rivalry scores of the three agents.

6 Discussions

In our study, we are mostly interested on how humans perceive the impact of different learning strategies when interacting with artificial agents. In particular, we evaluate if we can modulate this perception in a controlled manner using the rivalry term. Our experiments demonstrate that we can indeed achieve such manipulation, although in a limited manner. In this section, we discuss these findings in more detail.

6.1 Are learning agents perceived as rivals?

Our experimental results provided insights to identify whether agents trained with different RL strategies yield distinct rivalry and whether including rivalry in the reward function enable agents to modulate human's response on a rivalry scale. Our first exploratory study showed that agents trained with DQL and PPO strategies yield distinct rivalry if compared to a random agent when playing the Chef's Hat card game against humans, which confirms that their capability of learning a strategy is indeed perceived by humans. However, we failed to find a significant difference between DQL and PPO agents, which mirrors the lack of significant difference between their relative performances and their scores for the "Decisive," "Innovative" and "Creative" traits.

6.2 Agents optimization

When developing the agents, we needed to adapt them toward using rivalry as a learning modulator. Our experiments demonstrate that in our simulation environment, the inclusion of the rivalry term does not affect the general performance of the agent, and helps in changing their underlying behavior. This behavior can be explicitly demonstrated when a DQL rival agent plays a game against a non-rival version of itself, a PPO agent and a random agent. Figure 3, illustrates a match between a DQL agent, with rivalry, a DQL agent without rivalry, a random agent, and a PPO agent. We observe that the rivalry value increases constantly for the DQL agent, as they share the same strategy and social traits. Against the PPO agent, the rivalry score increases at a different pace, which demonstrates that the rival agent can detect the strategy of the PPO agent, and associate a different social trait set to it. A random agent however does not have a strategy, and the rival score fluctuates for each action it takes as expected.

Table 2 Relative performance and rivalry scores of the three agents

Agent type	Performance (SD)	Rivalry (SD)
Rival DQL	1.50 (.50)	0.44 (.08)
COPPER DQL	0.63 (.24)	0.35 (.08)
Offline DQL	0.41 (.32)	0.30 (.36)

6.3 The role of rivalry in Chef's Hat

Our last experiment, where we measure the impact of rivalry, showed that agents trained with a predicted rivalry as a part of their reward function can modulate human responses on a rivalry scale, and in turn yield significantly higher rivalry results. Moreover, the rivalry scores for the rival, COPPER-based, and offline learning agents were all significantly different from each other, with the rival agent exhibiting the highest score. Similarly to the exploratory study, these differences mirrored the relative performances of the agents. The similarity results instead were not significantly different for any of the agents. This suggests that agents' behaviors were not providing enough information to be perceived as different in terms of their agency, communion and competence traits.

Our experiments confirmed the importance of agent's performances to trigger a sense of rivalry. Moreover, the agents learned to modulate their behavior enough to yield a significantly higher rivalry. However, their performances did not reach those of the participants, suggesting there is still room for improvement.

The addition of rivalry and the better performances of the agents seemed to increase the motivation of participants to play better, as it can be seen in the increased average scores from the first exploratory to the second one. Although we cannot exclude that this might depend also on a difference between the two players' samples, such result would be expected from rivalry literature [16]. We saw that the significant difference in rivalry scores did not had a significant effect on participants' motivation to play with different agents. Motivation to play again could be further investigated by allowing participants to only play with one type of agent in future studies. Another observation is that the rival RL agents were not associated to different judgments in terms of social traits such as agency, competence or sense of communion, on which the similarity estimation is based. This might be due to the scenario of the game, where the agent's behavior could be appreciated only from its choices in the game. This point could be further investigated in a study where agents have more indicators

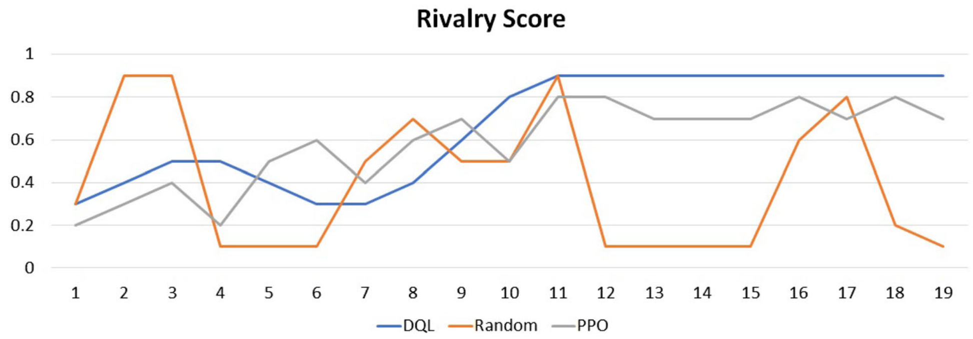

Fig. 3 Example of the rivalry score calculation of a rival DQL agent playing against a non-rival DQL, a PPO and a random-based agent

of social traits such as an embodiment, gestures or emotional expressions.

However, even in such a socially impoverished setting, the rival agent succeeded in triggering a higher rivalry in the human participants, revealing the potentiality of introducing rivalry in social reinforcement learning research.

In comparison with the first human experiment, the subjects of the second human experiment played slightly more turns, but achieved a significantly higher average score. This might be an indication that the rival agent made the game more motivating, and more challenging. Observing these characteristics from the behavior of each agent alone is a strong indication that when adding explicit social traits to the agent's behavior, the rivalry modulation could be a game-changer on social reinforcement learning research.

7 Conclusion

In this work, we examined the inclusion of social aspects in the learning strategies of reinforcement learning (RL) agents in a competitive card game scenario. We proposed the social metric of rivalry based on the background research in social psychology, and trained RL agents with a reward function that reflects this metric. The resulting agents, trained with the rivalry metric, were successful in yielding significantly different play styles, distinguishable in terms of a main antecedent of rivalry: relative performance. However, the similarity scores calculated based on social traits—another main factor in rivalry—failed to show differences. We plan to further address this issue in future studies by equipping RL agents with distinguishable social signals. Our results suggest that using social concepts such as rivalry shows promise in training agents to perform in human interaction contexts.

Appendix 1: Questionnaires

We used two questionnaires in our experiments. One before the game start, to collect the self-assessment information of the players, and one after the game finishes, to identify the players' perception of the agents. The trait similarity questions consisting of competence (Questions 9 to 12), agency (Questions 5 to 8), and communion (Questions 13 to 16) traits as the indicators of relevant behavioral attributes, as used as indicators of stereotypical traits [46] and previously used in virtual agent research [50].

Pre-game questionnaire

1. Which was the nickname you choose?
2. What is your mother tongue?
3. How old are you?
    - Less than 18 years old
    - Between 18 and 30
    - Between 31 and 50
    - More than 50
    - Do not want to disclose
4. How competitive you are when playing games?
5. How ambitious are you when playing games?
6. How courageous are you when playing games?
7. How decisive are you when playing games?
8. How aggressive are you when playing games?
9. How creative are you when playing games?
10. How Intelligent are you when playing games?
11. How innovative are you when playing games?
12. How organized are you when playing games?
13. How compassionate are you when playing games?
14. How affectionate are you when playing games?
15. How sensitive are you when playing games?
16. How emotional are you when playing games?

17. How experienced are you with competitive card games?

1(Not at all)	2	3	4	5 (Very)

1(Not at all)	2	3	4	5 (Very)

1(Not at all)	2	3	4	5 (Very)

1(Not at all)	2	3	4	5 (Very)

1(Not at all)	2	3	4	5 (Very)

1(Not at all)	2	3	4	5 (Very)

1(Not at all)	2	3	4	5 (Very)

1(Not at all)	2	3	4	5 (Very)

1(Not at all)	2	3	4	5 (Very)

1(Not at all)	2	3	4	5 (Very)

1(Not at all)	2	3	4	5 (Very)

1(Not at all)	2	3	4	5 (Very)

1(Not at all)	2	3	4	5 (Very)

1(Not at all)	2	3	4	5 (Very)

Post-game questionnaire

1. Which was the nickname you choose?
2. How competitive were your opponents?
3. How ambitious were your opponents?
4. How courageous were your opponents?
5. How decisive were your opponents?
6. How aggressive were your opponents?
7. How creative were your opponents?
8. How intelligent were your opponents?
9. How innovative were your opponents?
10. How organized were your opponents?
11. How compassionate were your opponents?
12. How affectionate were your opponents?
13. How sensitive were your opponents?
14. How emotional were your opponents?
15. Would you like to play with these opponents again?

	1(Not at all)	2	3	4	5 (Very)
Dylan					
Frankie					
Evan					

	1(Not at all)	2	3	4	5 (Very)
Dylan					
Frankie					
Evan					

	1(Not at all)	2	3	4	5 (Very)
Dylan					
Frankie					
Evan					

	1(Not at all)	2	3	4	5 (Very)
Dylan					
Frankie					
Evan					

	1(Not at all)	2	3	4	5 (Very)
Dylan					
Frankie					
Evan					

	1(Not at all)	2	3	4	5 (Very)
Dylan					
Frankie					
Evan					

	1(Not at all)	2	3	4	5 (Very)
Dylan					
Frankie					
Evan					

	1(Not at all)	2	3	4	5 (Very)
Dylan					
Frankie					
Evan					

	1(Not at all)	2	3	4	5 (Very)
Dylan					
Frankie					
Evan					

	1(Not at all)	2	3	4	5 (Very)
Dylan					
Frankie					
Evan					

	1(Not at all)	2	3	4	5 (Very)
Dylan					
Frankie					
Evan					

	1(Not at all)	2	3	4	5 (Very)
Dylan					
Frankie					
Evan					

	1(Not at all)	2	3	4	5 (Very)
Dylan					
Frankie					
Evan					

	Yes	No
Dylan		
Frankie		
Evan		

Appendix 2: Similarity predictor optimization

To optimize this network, we used a Tree-Parzen optimizer (Hyperopt [53]) based on the search space defined in Table 3.

Table 3 Search space used to optimize the similarity predictor neural network

Parameter	Search space
Similarity predictor	
Number of concatenated actions	[3,5,10, 15]
Number of layers	[1,2,3]
Units per layer	[16, 32, 64, 128, 256, 512, 1024]

Acknowledgements A.S and A.T. are supported by a Starting Grant from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme. G.A. No 804388, wHiSPER.

Funding Open access funding provided by Istituto Italiano di Tecnologia within the CRUI-CARE Agreement.

Declarations

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

1. Liu Z, Wang J, Gong S, Lu H, Tao D (2019) Deep reinforcement active learning for human-in-the-loop person re-identification. In: Proceedings of the IEEE international conference on computer vision, pp 6122–6131
2. Modares H, Ranatunga I, Lewis FL, Popa DO (2015) Optimized assistive human–robot interaction using reinforcement learning. IEEE Trans Cybern 46(3):655–667
3. Churamani N, Kerzel M, Strahl E, Barros P, Wermter S (2017) Teaching emotion expressions to a human companion robot using deep neural architectures. In: 2017 international joint conference on neural networks (IJCNN). IEEE, pp 627–634
4. Papaioannou I, Lemon O (2017) Combining chat and task-based multimodal dialogue for more engaging HRI: a scalable method using reinforcement learning. In: Proceedings of the companion of the 2017 ACM/IEEE international conference on human–robot interaction, pp 365–366
5. Gao Y, Yang F, Frisk M, Hemandez D, Peters C, Castellano G (2019) Learning socially appropriate robot approaching behavior toward groups using deep reinforcement learning. In: 2019 28th IEEE international conference on robot and human interactive communication (RO-MAN). IEEE, pp 1–8

6. Milani S, Topin N, Houghton B, Guss WH, Mohanty SP, Vinyals O, Kuno NS (2020) The minerl competition on sample-efficient reinforcement learning using human priors: a retrospective. arXiv preprint arXiv:2003.05012
7. Vinyals O, Babuschkin I, Czarnecki WM, Mathieu M, Dudzik A, Chung J, Choi DH, Powell R, Ewalds T, Georgiev P et al (2019) Grandmaster level in starcraft ii using multi-agent reinforcement learning. Nature 575(7782):350–354
8. Choudhury R, Swamy G, Hadfield-Menell D, Dragan AD (2019) On the utility of model learning in HRI. In: 2019 14th ACM/IEEE international conference on human–robot interaction (HRI). IEEE, pp 317–325
9. Khamassi M, Velentzas G, Tsitsimis T, Tzafestas C (2018) Robot fast adaptation to changes in human engagement during simulated dynamic social interaction with active exploration in parameterized reinforcement learning. IEEE Trans Cogn Dev Syst 10(4):881–893
10. Ritschel H, André E (2017) Real-time robot personality adaptation based on reinforcement learning and social signals. In: Proceedings of the Companion of the 2017 ACM IEEE international conference on human–robot interaction, pp 265–266
11. Cuayáhuitl H, Lee D, Ryu S, Cho Y, Choi S, Indurthi S, Yu S, Choi H, Hwang I, Kim J (2019) Ensemble-based deep reinforcement learning for chatbots. Neurocomputing 366:118–130
12. Tabrez A, Hayes B (2019) Improving human–robot interaction through explainable reinforcement learning. In: 2019 14th ACM/IEEE international conference on human–robot interaction (HRI). IEEE, pp 751–753
13. Madumal P, Miller T, Sonenberg L, Vetere F (2019) Explainable reinforcement learning through a causal lens. arXiv preprint arXiv:1905.10958
14. Sequeira P, Gervasio M (2020) Interestingness elements for explainable reinforcement learning: understanding agents' capabilities and limitations. Artif Intell 288:103367
15. Havard CT (2020) What is rivalry and where we go from here. In: Springer (ed) Rivalry in sport, 1st edn., vol 1. Palgrave Macmillan, Cham, Springer. Chap. 1, pp 9–35
16. Kilduff GJ, Elfenbein HA, Staw BM (2010) The psychology of rivalry: a relationally dependent analysis of competition. Acad Manag J 53(5):943–969
17. Kilduff GJ (2014) Driven to win: rivalry, motivation, and performance. Soc Psychol Personal Sci 5(8):944–952
18. Barros P, Sciutti A, Bloem AC, Hootsmans IM, Opheij LM, Toebosch RH, Barakova E (2021) It's food fight! designing the chef's hat card game for affective-aware HRI. In: Companion of the 2021 ACM/IEEE international conference on human–robot interaction, pp 524–528
19. Van Hasselt H, Guez A, Silver D (2016) Deep reinforcement learning with double q-learning. In: Thirtieth AAAI conference on artificial intelligence
20. Schulman J, Wolski F, Dhariwal P, Radford A, Klimov O (2017) Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347
21. Barros P, Bloem AC, Hootsmans IM, Opheij LM, Toebosch RH, Barakova E, Sciutti A (2021) You were always on my mind: introducing chef's hat and copper for personalized reinforcement learning. Front Robot AI 8. https://www.frontiersin.org/articles/10.3389/frobt.2021.669990/full
22. Campbell M Jr, Hoane AJ, Hsu F (2002) Deep blue. Artif Intell 134(1–2):57–83
23. DeCoste D (1997) The future of chess-playing technologies and the significance of kasparov versus deep blue. In: Deep blue versus Kasparov: the significance for artificial intelligence, pp 9–13
24. DeCoste D (1998) The significance of Kasparov versus deep blue and the future of computer chess. ICGA J 21(1):33–43
25. Shao K, Tang Z, Zhu Y, Li N, Zhao D (2019) A survey of deep reinforcement learning in video games. arXiv preprint arXiv:1912.10944
26. Silver D, Huang A, Maddison CJ, Guez A, Sifre L, Van Den Driessche G, Schrittwieser J, Antonoglou I, Panneershelvam V, Lanctot M et al (2016) Mastering the game of go with deep neural networks and tree search. Nature 529(7587):484–489
27. Shao K, Zhu Y, Zhao D (2017) Cooperative reinforcement learning for multiple units combat in starcraft. In: 2017 IEEE symposium series on computational intelligence (SSCI). IEEE, pp 1–6
28. Barros P, Tanevska A, Sciutti A (2020) Learning from learners: adapting reinforcement learning agents to be competitive in a card game. arXiv preprint arXiv:2004.04000
29. Rossi S, Ferland F, Tapus A (2017) User profiling and behavioral adaptation for HRI: a survey. Pattern Recogn Lett 99:3–12
30. Arzate Cruz C, Igarashi T (2020) A survey on interactive reinforcement learning: design principles and open challenges. In: Proceedings of the 2020 ACM designing interactive systems conference, pp 1195–1209
31. Reeves B, Nass C (1996) The media equation: How people treat computers, television, and new media like real people, vol 10. Cambridge, UK, p 236605
32. Ball G, Breese J (2000) Emotion and personality in a conversational agent. Embodied conversational agents. MIT Press, Cambridge, MA, USA, 189–219
33. Zhang J, Zheng J, Magnenat-Thalmann N (2016) Modeling personality, mood, and emotions. In: Springer (ed) Context aware human–robot and human–agent interaction, 3rd edn, vol 1. Springer, pp. 211–236. Chap. 1
34. Leite I, Martinho C, Paiva A (2013) Social robots for long-term interaction: a survey. Int J Soc Robot 5(2):291–308
35. Moerland TM, Broekens J, Jonker CM (2018) Emotion in reinforcement learning agents and robots: a survey. Mach Learn 107(2):443–480
36. Murphy RR, Schreckenghost D (2013) Survey of metrics for human–robot interaction. In: 2013 8th ACM/IEEE international conference on human–robot interaction (HRI). IEEE, pp 197–198
37. Perolat J, Leibo JZ, Zambaldi V, Beattie C, Tuyls K, Graepel T (2017) A multi-agent reinforcement learning model of common-pool resource appropriation. In: Advances in neural information processing systems, vol 30
38. Cañamero L (2020) Embodied affect for real-world human–robot interaction. In: Proceedings of the 2020 ACM/IEEE international conference on human-robot interaction, pp 459–460
39. Peltason J, Riether N, Wrede B, Lütkebohle I (2012) Talking with robots about objects: a system-level evaluation in HRI. In: 2012 7th ACM/IEEE international conference on human-robot interaction (HRI). IEEE, pp 479–486
40. Begum M, Serna RW, Kontak D, Allspaw J, Kuczynski J, Yanco HA, Suarez J (2015) Measuring the efficacy of robots in autism therapy: how informative are standard HRI metrics'. In: Proceedings of the tenth annual ACM/IEEE international conference on human–robot interaction, pp 335–342
41. Kilduff GJ (2010) The psychology of rivalry. PhD thesis, UC Berkeley
42. Bansal T, Pachocki J, Sidor S, Sutskever I, Mordatch I (2017) Emergent complexity via multi-agent competition. arXiv preprint arXiv:1710.03748
43. Kidziński Ł, Mohanty SP, Ong CF, Huang Z, Zhou S, Pechenko A, Stelmaszczyk A, Jarosik P, Pavlov M, Kolesnikov S et al (2018) Learning to run challenge solutions: adapting reinforcement learning methods for neuromusculoskeletal environments. In: Springer (ed) The NIPS'17 competition: building intelligent systems, 1st edn., vol 1. Springer, pp 121–153. Chap. 1

44. Miller CT (1984) Self-schemas, gender, and social comparison: a clarification of the related attributes hypothesis. J Pers Soc Psychol 46(6):1222
45. Tauer JM, Harackiewicz JM (1999) Winning isn’t everything: competition, achievement orientation, and intrinsic motivation. J Exp Soc Psychol 35(3):209–238
46. Eagly AH, Nater C, Miller DI, Kaufmann M, Sczesny S (2020) Gender stereotypes have changed: a cross-temporal meta-analysis of U.S. public opinion polls from 1946 to 2018. American Psychologist 75(3), 301–315. https://doi.org/10.1037/amp0000494
47. Wojciszke B, Abele AE, Baryla W (2009) Two dimensions of interpersonal attitudes: liking depends on communion, respect depends on agency. Eur J Soc Psychol 39(6):973–990
48. Cuddy AJ, Fiske ST, Glick P (2008) Warmth and competence as universal dimensions of social perception: the stereotype content model and the bias map. Adv Exp Soc Psychol 40:61–149
49. Fiske ST, Cuddy AJ, Glick P, Xu J (2002) A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. J Pers Soc Psychol 82(6):878
50. Nag P, Yalçın ÖN (2020) Gender stereotypes in virtual agents. In: Proceedings of the 20th ACM international conference on intelligent virtual agents, pp 1–8
51. Abele AE, Wojciszke B (2007) Agency and communion from the perspective of self versus others. J Pers Soc Psychol 93(5):751
52. Cruz F, Dazeley R, Vamplew P (2020) Explainable robotic systems: understanding goal-driven actions in a reinforcement learning scenario. arXiv preprint arXiv:2006.13615
53. Bergstra J, Yamins D, Cox DD (2013) Hyperopt: a python library for optimizing the hyperparameters of machine learning algorithms. In: Proceedings of the 12th Python in science conference, pp 13–20. Citeseer
54. Pohlert T, Pohlert MT (2018) Package ‘pmcmr’. R package version, vol 1(0)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.