Let's Compete! The Influence of Human-Agent Competition and Collaboration on Agent Learning and Human Perception

Ornnalin Phaijit
l.phaijit@unsw.edu.au
The University of New South Wales
Sydney, Australia

Claude Sammut
c.sammut@unsw.edu.au
The University of New South Wales
Sydney, Australia

Wafa Johal
wafa.johal@unimelb.edu.au
The University of Melbourne
Melbourne, Australia

ABSTRACT

In interactive agent learning, the human may teach in a collaborative or adversarial manner. Past research has been focusing on collaborative teaching styles as these are common in human education settings, while overlooking adversarial ones despite promising results in recent research. Moreover, agent performance has been the main focal point while neglecting the perspective of the human teacher, who is crucial to the instructional process. In this work, we examine the impact of competitive and collaborative teaching styles on agent learning and human perception. We conducted a study (N=40) for participants to demonstrate a task in different interaction modes for teaching a computer agent: collaboratively, competitively, or without interacting with the agent. Most participants reported that they preferred competing against the computer agent to the other two modes. Despite smaller numbers of demonstrations given from the user, the agent performance from the interactive modes (collaborative and competitive) was comparable to the non-interactive mode (solo). The agent was perceived as being more competent in the competitive mode than the collaborative mode despite the marginally worse in-task performance. These preliminary findings suggest that competitive types of interaction, when agents or robots learn from humans, lead to better human perception of the agent's learning when compared to collaborative, and better user engagement when compared to non-interactive learning from demonstrations.

CCS CONCEPTS

• **Human-centered computing** → **Interaction design**; **Empirical studies in HCI**.

KEYWORDS

human-computer interaction, interactive learning, agent learning, curriculum learning, learning from demonstration

HAI '22, December 5–8, 2022, Christchurch, New Zealand

ACM ISBN 978-1-4503-9323-2/22/12…$15.00
https://doi.org/10.1145/3527188.3561922

ACM Reference Format:
Ornnalin Phaijit, Claude Sammut, and Wafa Johal. 2022. Let's Compete! The Influence of Human-Agent Competition and Collaboration on Agent Learning and Human Perception. In *Proceedings of the 10th International Conference on Human-Agent Interaction (HAI '22), December 5–8, 2022, Christchurch, New Zealand*. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3527188.3561922

1 INTRODUCTION

Interactive training of agents has increased over the years due to the ability to integrate human expertise and individual preferences, thereby improving the performance of the system [13]. However, research in this field has mainly been looking into interaction styles which are advisory or collaborative in nature, such as giving feedback [7, 14, 16] or instructions [8, 21] to the agent learner. In comparison, there has been very little research into human adversary in agent training. In this context, human adversary refers to when the human trainer teaches the agent by interacting adversarially with it. Recent studies have shown that human adversarial teaching is able to produce robust agent performance [6, 9, 24]. Yet, all existing human adversarial agent training studies have been for reinforcement learning, which requires for the human to know exactly *when* and *how* to adverse/disturb the agent. Whereas, teaching by demonstration may place less mental burden on the teacher as s/he simply has to perform the task. However, to the best of our knowledge, there has not yet been any research on how human adversary may perform in learning from human demonstration. Although it is certainly intuitive for the teacher to interact with the agent in a supportive and advisory manner, there still remain questions on how collaborative and competitive ways of teaching compare with each other.

Moreover, current literature on interactive agent learning has been heavily focused on the agent's performance, while lacking the human teacher's perspective during the process. As, in this kind of teaching framework, the effectiveness of the agent's learning highly depends on the human teacher [1], it is important that the perspective of the teacher is studied in order to produce an effective solution to agent learning. Not only is user perception important for agent performance, it also influences user retention. For instance, an agent's perceived competency greatly impacts user trust in an agent [4]. By building user trust, this can increase the user's willingness to accept the agent [10], making the training process more pleasant for the teacher. Agent training can also typically be repetitive, time-consuming and disengaging for the human. As human disengagement has been negatively correlated with task performance [19], not only does this impact upon the user negatively, but the reduction in teaching quality may also negatively affect agent learning.

In this work, we conduct an exploratory study to investigate the impact of collaborative and competitive human teaching manners on both agent learning and human perception. We designed and developed a modified Pacman computer game as a task to compare between collaborative and competitive teaching styles. In the collaborative mode, the user performs demonstrations while collaborating with the agent on the task. In the competitive mode, the user demonstrates while competing against the agent on a task. As a baseline, we also implemented a solo mode where the user demonstrates the task without interacting with the agent. We conducted a user study (N=40) for participants to play all of the modes. Our results show that demonstrations from the interactive modes produced comparable results to those from the solo mode despite the smaller numbers of demonstrations performed. We found that for the solo and competitive modes, higher proportions of risks made by the participants significantly improved the performance of the agent. In regards to human perception, the participants reported that they liked playing competitively the most and perceived the agent in this mode as performing better than in the collaborative mode despite the agent's marginally worse in-task performance.

2 BACKGROUND

2.1 Learning from human teachers

Although learning from experience has shown to be effective in autonomous agents [18], the learning process can be very slow. Learning from demonstration, on the other hand, is a technique in which the agent learns by observing the human teacher performing the task. This provides examples for the agent to learn from, rather than simply leaving it to acquire a sufficient amount of experience. Similarly, when humans learn from observing and imitating others, the learning process is much faster than using trial-and-error. Behavioral cloning [2] is an approach which allows the agent to learn by using information on the state of the demonstrator's environment, and the corresponding actions of the demonstrator. With this knowledge, a classifier can be built to allow for the agent to perform an imitation of the demonstrator's behaviour.

Teaching is a skill and the requirement for the human to learn how to teach on top of having to teach the agent can be adding to the human's mental load. It has been shown that increased cognitive load negatively affects human task performance [15]. By simply demonstrating rather than explicitly teaching the student, this places less burden on the demonstrator as s/he simply has to perform the actions without having to consider what inputs will be the most conducive to agent learning. This is convenient as it has also been found that human teachers tend to want to perform demonstrations for the learning agent rather than giving feedback only [11, 23].

By utilising our knowledge of the preferences of human teachers, a better teaching process may be formulated for both the human user and the agent learner. For these reasons, teaching by demonstration was chosen for agent training in our work.

2.2 Adversarial learning

Human-agent interactions in interactive agent learning can be collaborative or adversarial. While there has been ample research showing the efficacy of human advisory teaching, human adversarial teaching is still new. Duan et al. [6] discovered that human adversary helped to improve the grasping performance of the robot compared to the performance without human intervention. However, it was found that despite the participants being instructed to act adversarially, at times, they would not apply perturbations correctly as they found the act to be challenging. This led to the creation of Yoon et al.'s framework [24] which distinguished between collaborative and adversarial interactions such that the robot would only learn from adversarial interventions. They found that the performance improved when using the framework. In 2021, Hamaya et al. [9] found that mixing human advisory and human adversary in a robotic peg-in-hole learning task improved the agent performance when compared to scenarios with no or random interactions. Although these studies paved the way for research into using human adversary in agent learning, they did not compare the effects of human adversary and human *collaboration*. All existing research on agent training using human adversary has been in reinforcement learning which places the burden on the human to recognise *how* to hinder the agent. For instance, this requires watching and waiting for the agent, recognising *when* exactly to advise or disturb, and then recognising exactly *how* to correct or disturb it. Instead of giving feedback, translating this into learning from demonstration could mean that the human teacher simply has to carry out the task without being concerned with how to advise or disturb. In our work, the human teaches the agent by performing the task alongside the agent in collaborative and competitive scenarios.

2.3 Curriculum Learning

Typically, when humans learn, easier and more digestible concepts are presented to the learner first, before they are to encounter more advanced examples. This technique is generally used to ensure the learning process is as effective as possible. Likewise, this concept has been introduced to agent learning as *curriculum learning*. It has been shown in research that finding a purposeful ordering of examples for the machine learner improves its learning process [3, 17, 20]. One of the major issues, however, is figuring out how best to order the examples in an effective way. For a layperson with no knowledge of how the machine works, it can be difficult to evaluate the complexity of samples. Without classifying learning tasks by difficulty, [9] found that providing advisory interactions, followed by human adversarial ones, helped to minimise errors early on while providing exploratory learning in the later stages for the robot. Although this provided stepping stones in regards to curriculum learning using different interaction modes, other combinations have not yet been explored.

3 RESEARCH QUESTIONS

Combining current research gaps in interactive agent learning, we formulate questions to determine the advantages and disadvantages of the two interactive teaching modes (competitive and collaborative). In a game for the purpose of teaching an agent to perform a task, would playing collaboratively with the agent or playing competitively against the agent generate better agent performance? Would playing the game alone (i.e. solo mode) without the agent, instead, be the most effective?

To our knowledge, there has also been no research done on whether people would prefer to be performing tasks collaboratively or competitively with a learning agent. It is also unclear whether performing demonstrations while collaborating with the agent or competing against it would feel more challenging to the user, and whether these interaction modes would affect the human's perception of the agent.

Finally, one could also wonder if instead of making a binary choice between collaborative and competitive mode, a combination of these modes could be found beneficial for the agent's learning. Considering that curriculum learning has been shown to be effective in improving agent learning, it would be advantageous to investigate the impact of ordering different interaction modes on agent performance.

We therefore propose to explore four key research questions focusing on: agent performance, user preference, user perception of agent competency, and curriculum learning.

- Which mode is better for agent learning?
- What interaction mode do people prefer?
- How do different interaction modes influence people's perceptions of agent competency?
- How could using different interaction modes to construct curricula influence agent learning?

4 EXPERIMENT

4.1 Task Design

As we wanted to compare the effects of collaboration and competition on agent learning and human perception, a simple game scenario was chosen as the task in our experiment. Using the Unity game engine [22], we developed a simplified version of the original Pacman computer game from 1980. In our version, the player navigates through the 2D 7x7 maze to collect *Pacdots*, which are placed around the map, while avoiding collision with the *ghost* character (Figure 1). This map has been derived from UC Berkeley's smallGrid map [5]. Pacman gains points by colliding with (and therefore collecting) Pacdots. One Pacdot contributes one point to the score. Upon collision with the ghost, Pacman dies and can no longer play. This game was chosen as the task is fairly simple and can easily be generalised to real world non-gaming scenarios such as objective collection and navigation. This game was also designed to be simple enough to ensure participants could be regarded as task experts. The small map size was to ensure that the agent could go from having no knowledge to being able to compete against human experts within ten games. All characters – the ghost, the player's Pacman, and the agent's Pacman – had the same speed.

4.1.1 The Second Pacman. In contrast to the original Pacman game where there is only one Pacman character, in our game, we introduce another Pacman character, but in green, as our agent to play with the human player and learn from her/him.

As the agent needs to learn by observation very quickly within ten games, we employed behavioural cloning. We used Microsoft's

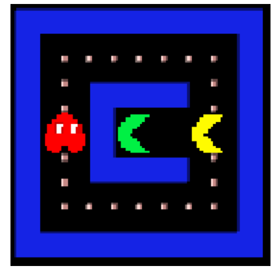

Figure 1: The Game is a 7x7 map of walls, "Pacdots", a ghost and Pacman characters.

LightGBM [12] framework for the decision tree construction for agent learning. Nine features of the player's environment were used to build the classifier to dictate the action of the agent: whether to move left, right, up, or down. Four of the nine features selected were related to the following: Pacman's distance to the closest Pacdot 1) from the position to the left, 2) from the position to the right, 3) from the position above, and 4) from the position below. Features 5-9 were related to Pacman's distance to the ghost: 5) from the position to the left, 6) from the position to the right, 7) from the position above, and 8) from the position below. Lastly, the final feature was whether the ghost was three steps away.

These features were chosen based on UC Berkeley's features used in their approximate Q-learning algorithm: distance to the closest food and the number of ghosts one step away [5]. The distance to the closest food, or Pacdot, for each action (move left, move right, move up and move down) was therefore included in our features 1-4. In our game, the players are humans whose reaction times can be delayed (unlike how the player in UC Berkeley's version is an agent). We therefore used the distance from the ghost (features 5-8) to help the agent determine the appropriate action, and increased the threshold from one step to three steps (feature 9) to account for human reaction times. Additionally, since there is only one ghost in our version, feature 9 is simply whether the ghost is three steps away or less.

After each game finishes, the decision tree is built based on the accumulated demonstrations from the previous games, and is used as a classifier for the agent's actions in the next game.

4.1.2 Game Modes. We designed the game such that there are three modes. As a baseline, we implemented the *solo* mode, in which the user is playing as the yellow Pacman with no other Pacman characters in the game. The player wins if all Pacdots on the map are collected without Pacman dying. The game ends when all Pacdots have been collected or if the player dies.

In order to compare between collaborative and competitive teaching interactions, we developed:

- The *collaborative* mode, where the user and the agent are playing as Pacman characters and working together as a team. They share the same score. The team wins when all of the 26 Pacdots are collected by either of the team members. The game ends when all Pacdots have been collected or if both the human and the agent are dead.

- The *competitive* mode, where the user and the agent are playing as Pacman characters but are now on different teams. They have their own scores. The Pacman that has collected more Pacdots by the end of the game wins, regardless of deaths. There is a tie if both Pacman characters end up with the same score.

As the agent has no training data to learn from in the first game in collaborative and competitive modes, it plays Game 1 stochastically. As more games are completed, the agent accumulates more training data to play in the next game.

4.2 Procedure

After obtaining the participants' consent, they were explained the goal of the game without being told the ghost's moves were random. They were then asked to play ten games of the solo mode first. This was to allow for the participants to familiarise themselves with the game and to obtain baseline data. After the baseline, we introduced to the participants the AI-based Pacman in green. We used a within-subject experimental design to test the collaborative and competitive modes. Twenty of participants were then asked to play 10 games of the collaborative mode, followed by 10 games of the competitive mode. The other 20 participants played 10 games of the competitive mode before playing 10 games of the collaborative mode. All game moves made by participants were recorded. After the games were finished, we carried out post-task experiments: agent learning performance, agent final performance, and curriculum learning. The procedure's timeline can be observed in Figure 2.

4.3 Participants

We recruited 40 participants (23 M, 17 F) to play our Pacman game and fill out the questionnaire before and after playing. The age range was between 18 and 32 (M=23.88, SD=3.69). The number of hours participants spent playing video games per week ranged between 0 to 50 hours (M=8.06, SD=11.03).

4.4 Questionnaire

Participants were asked to answer the questionnaire before and after playing.
Pre-game demographic question: How many hours do you play video games in a typical week? Along with this information, the participant ID and the mode that each participant started (collaborative or competitive) after playing the solo mode were recorded.

A post-game questionnaire was administrated to the participant after the end of the experiment.

Post-game questions:

- Did you feel the AI improve from the first game?
- Which mode (collaborative or competitive) did you feel the AI was better in?
- Which mode (collaborative or competitive) did you feel more engaged in?
- Which mode (collaborative or competitive) did you find more challenging to clear the maze?
- Which mode (collaborative or competitive) did you personally prefer playing?
- Which (solo or with/against the agent) did you enjoy playing more?[1]
- Did you use different strategies in the two modes (collaborative and competitive)?

5 RESULTS & DISCUSSION

5.1 In-Task Agent Performance

Our results showed that during the games with the participants, the agent obtained mean scores of 14.28 (SD=7.46) and 13.70 (SD=7.42) in the collaborative and competitive modes respectively. The large variances were likely due to the highly stochastic nature of the game as the ghost's moves were random in a small map. Although our results do show that the mean in-task score of the agent was slightly greater in collaborative than that in competitive, the difference was found to be statistically non-significant ($p = .21$) upon conducting the Wilcoxon signed-rank test. The agent's rates of death were also similar in the collaborative and competitive modes, at 48.25% (193 deaths) and 49.25% (197 deaths) respectively. Overall, the in-task performance of the agent did not differ significantly between the two interactive modes.

In terms of winning, the agent's mean in-task win rate for the competitive mode (M=48.50% SD=15.58%) was much lower than that of the collaborative mode (M=97.75%, SD=4.74%). This was unsurprising as it had to play against, instead of collaborating with, the human *expert*.

5.2 Post-Task Agent Performance

To evaluate the agent's learning from user demonstrations, we used the recorded moves made by the participants during the games to construct classifiers (using Microsoft's LightGBM framework) for the agent. We then allowed the agent to play in isolation. As we wanted to examine the learning process of the agent, we documented the score of the agent with an increasing number of expert samples used to build its classifier.

Figure 3 shows the game score of the agent with the number of expert samples used from the solo, collaborative and competitive modes. It can be seen that, for a given number of demonstration samples, the agent's performance was comparable between the solo and competitive modes. However, learning was noticeably slower when using the collaborative mode's expert samples. This could be due to the missing context derived from when the participants were playing as a team with the agent. The players could have intentionally allowed for the agent to contribute to the team score. In contrast, in the competitive mode, the players were each playing for themselves rather than accommodating for another Pacman, and hence there was less context missing. The participants would have been attempting to make optimal moves at any point in time.

Next, we examined the final agent performance using the full game samples. We allowed for the agent to play the game in each of the three modes 100 rounds each, using each participant's samples. The mean agent score was the lowest at 21.96 (SD=3.40) for the collaborative mode, while the mean scores for the solo and competitive modes were similar to each other at 23.21 (SD=2.04) and 23.05 (SD=2.67) respectively. We performed Wilcoxon signed-rank tests

[1] this question was added after 19 participants had already undertaken the experiment.

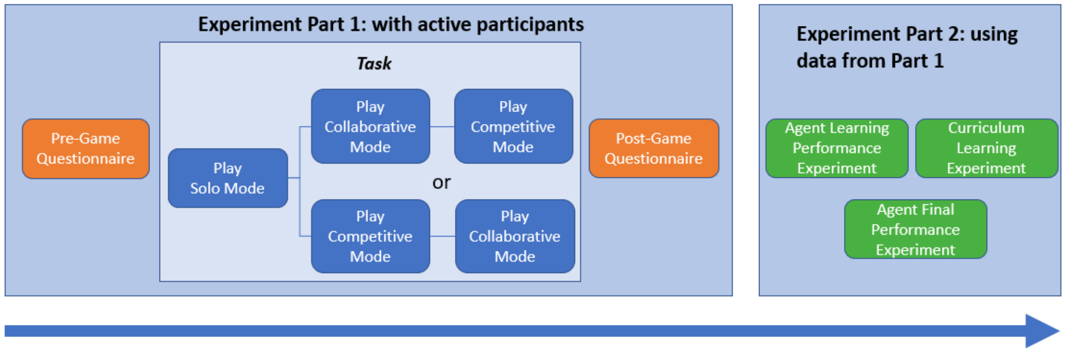

Figure 2: The experimental procedure timeline

between pairs of the three modes and found $p = .24$ between the collaborative and competitive modes, $p = .36$ between the solo and collaborative modes, and $p = .91$ between the solo and competitive modes. These results reveal that the score differences between the modes are non-significant. It is also important to note that the mean number of expert samples in one solo match is 41.77 (SD=5.90) and the mean numbers of expert samples in a collaborative match and a competitive match are 33.17 (SD=6.55) and 31.79 (SD=6.57) respectively (Figure 4). From conducting Wilcoxon signed-rank tests, we discovered that there were significant differences between the numbers of samples in a solo game and a collaborative or competitive game ($p < .001$) and no significant difference between the collaborative and competitive modes ($p = .30$). The number of moves required to finish a game in the collaborative and competitive modes were respectively on average 20.59% and 23.88% less than the number of moves required for a solo game. Even with smaller numbers of samples, demonstrations from the interactive modes still allowed for the agent to perform comparably to the solo mode's samples with no significant differences.

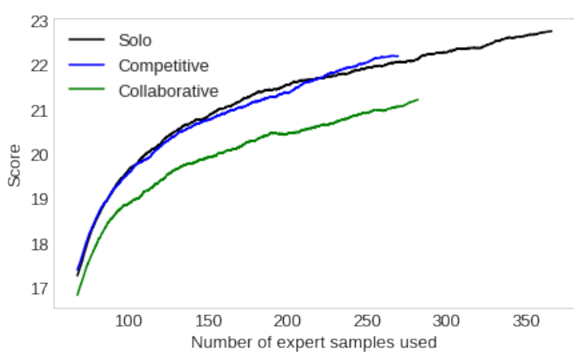

Figure 3: Post-task agent's performance from learning from expert samples from the solo, collaborative and competitive modes.

We examined to see if there would be a correlation between the post-task performance of agents using a specific participant's samples and the number of hours the participant spends each week on gaming. For each of the 40 participants, and for each mode, the participant's moves were used to build a classifier for the agent to play the game in isolation. The procedure was repeated 100 times for each mode and for each participant. We found that there was little to no correlation between the resulting agent's performance and the number of gaming hours upon examining the Pearson correlation coefficients (solo mode: $r = .08, p = .63$, collaborative mode: $r = -.03, p = .87$, competitive mode: $r = -.03, p = .87$). The lack of correlation could be due to the task being so simple that the amount of gaming the user regularly does had little to no effect. This was intended as we wanted the players to be deemed as experts at the task.

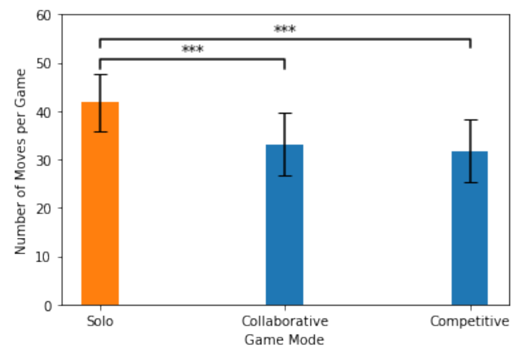

Figure 4: Mean numbers of player moves executed in one game.

5.3 Curriculum Learning

In this section, we examine the effects of combining different modes of the game as curricula for the agent's learning. To investigate this, we created six different learning curricula for the agent where the first half of the samples were from one mode and the second half from another. The combinations we chose were: 1) collaborative-solo, 2) competitive-solo, 3) solo-competitive, 4) solo-collaborative, 5) collaborative-competitive, and 6) competitive-collaborative. As we wanted to compare these curricula to the solo mode, the total number of samples used for each curriculum was the same as the number of solo samples for that particular participant.

We plotted the learning performance of the agent using these six curricula, using a moving average of 5500 samples in Figure 5. All of the curricula, except ones starting with the collaborative mode samples, had similar learning rates as that of the solo mode samples. From the moving average graph, it can be observed that the scores of all mixed curricula but two were converging towards slightly higher scores than the pure solo mode curriculum. The two curricula which did not converge towards as high scores were the collaborative-competitive and competitive-collaborative curricula. Interestingly, these were the ones without solo mode samples. Their mean final scores were also the lowest compared to the other mixed curricula.

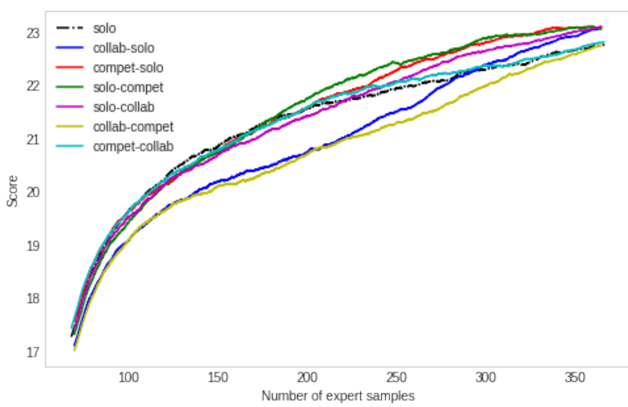

Figure 5: The Pacman agent's learning performance using different curricula

Apart from the collaborative-competitive curriculum, all of the mean scores obtained from using curriculum learning were higher than the mean score of samples from the solo mode only (Table 1). In general, however, there was no statistical significant difference among the curricula. The only statistically significant difference was between the collaborative-competitive and solo-competitive curricula (p=.038) which had the highest and lowest mean agent scores respectively.

These results from curriculum learning did not show much significant difference. This may be due to the fact that the ordering of the samples was not based on the difficulty level of the task. In the three game modes, the task complexity level remained the same. Hence, this method of sample sequencing by game modes was ineffective.

5.4 User Perception of Agent Improvement and Performance

During our experiment, the mean number of Pacdots the agent collected in Game 1 was 3.18 and 2.90 for the collaborative and competitive modes (SD=2.41 and SD=2.13 respectively). The agent's mean number of Pacdots collected was increased to 14.88 and 14.58 for the collaborative and competitive modes respectively by the end of Game 10 (SD=6.63 and SD=6.97). Although it is clear that the performance of the agent improved from Game 1 in both modes, the answers from the questionnaire did not reflect this. Only 10 (25%) of the participants felt the agent improved in both modes. Fifteen (37.5%) of the participants felt the agent only improved in the competitive mode, eight (20%) felt the agent only improved in the collaborative mode and the rest (seven, or 17.5%) of the participants reported that they did not feel the agent improve from the first game in either of the modes. During the questionnaire, some participants voiced that they had not been paying attention to how well the agent was performing in the first games. The players were more focused on their own performance than the agent's. It was also to be expected that the participants would be more aware of the opponent's performance during the first game of the competitive mode since the participant was having to compete against the agent. Hence, a sizeable portion of the participants felt the agent had improved in the competitive mode. On the other hand, some may have noticed the improvement of the agent during the collaborative mode more as the agent was directly affecting the player's score and was being a 'deadweight' during the first game. Some of the participants would voice frustration with the collaborative agent's performance.

Table 1: Final Agent Scores from Different Curricula

Curriculum	Final Agent Score	
	Mean	Standard Deviation
Solo	23.21	2.04
Collab-Solo	23.43	1.50
Compet-Solo	23.65	1.11
Solo-Compet	23.71	1.16
Solo-Collab	23.55	1.32
Collab-Compet	23.20	1.60
Compet-Collab	23.24	1.70

As for the comparison of the agent's performance between the two modes, the average numbers of Pacdots collected by the agent were 14.28 (SD=7.46) and 13.70 (SD=7.42) for the collaborative and competitive modes respectively. Interestingly, however, 32 (80%) of the participants felt that the agent was better in the competitive mode than the collaborative mode. Although the mean scores and death rates of the agent between the two interactive modes were very similar (with the collaborative score even being slightly higher than the competitive one, and the collaborative death rate being slightly lower than the competitive one), the overwhelming majority of the participants felt that the agent was better in the competitive mode. This could have been due to the fact that players were more likely to view the agent's competency as higher when the agent was actively being adversarial towards them.

5.5 User Engagement and Preference

The majority of the participants (33 out of 40) reported that they felt more engaged in the competitive mode than the collaborative mode. 76.2% of the participants said they preferred playing collaboratively/competitively over playing the solo mode. It is apparent that users find it more engaging playing with an opponent as opposed to playing alone. These results may help to suggest that in the tedious tasks of the teacher having to perform demonstrations for an agent learner, it may benefit the teacher to perform these demonstrations while interacting with the student competitively.

In terms of user preference between the two interactive modes, 72.5% of the participants preferred playing competitively while just over a quarter (27.5%) preferred playing collaboratively. We discovered that the overwhelming majority of the players (92.5%) found the competitive mode more challenging; however, the majority also reported that they preferred playing this mode. This would suggest that the participants enjoy more challenging tasks when interacting with an agent. Although it is clear that at this level of difficulty people tend to prefer the challenge, it would be advantageous to, in the future, investigate the level of difficulty at which this would no longer hold true.

Some of the participants reported that they preferred a no-stress environment, and hence they preferred the collaborative mode. This would be another factor that should be taken into consideration when designing a platform for human teaching.

5.6 User Behaviour

From our experiment, we found that the death rates of the human players were 4.25%, 4.75% and 11.25% for the solo, collaborative and competitive modes respectively (Table 2). The death rate in the competitive mode more than doubled those of the other modes. This may have been due to the players' higher likelihood of making risky plays in order to have a higher score than the agent. Additionally, it could also have been due to participants not caring as much once they had defeated the agent. In any case, the fact that agent performance from competitive samples was found to be comparable to using the solo mode's samples despite the higher rate of deaths should be considered.

As one of the features used for classifying the action for the Pacman agent was whether or not the ghost was within three steps from the agent, we used this feature as a way to measure the percentage of "risky moves". We defined the percentage of risky moves as the proportion of the states that the agent was within three steps away from the ghost. Using the Pearson correlation coefficient, there is a positive correlation between the percentage of risky moves and the final agent performance in the solo mode ($r = .59, p < .001$) and the competitive mode ($r = .34, p = .034$). However, there is a weak negative correlation with a non-significant relationship in the collaborative mode ($r = -.18, p = .26$). This could be explained by the fact that in the solo and competitive modes, players would take risks in an attempt to play optimally. However, in the collaborative mode, since the agent would generally be making moves that were good, in order to simply assist the agent in collecting points, the player would be making risky sub-optimal moves.

For each of the three modes, we divided the participants into two groups: risk-averse and risk-seeking (Figure 6), based on their risky move rates. This was done using K-Means clustering. We then calculated the mean agent performance for each group. It can be observed in Figure 7, that for solo and competitive samples, risk-seeking individuals generated better agent performance (M=24.73 SD=3.33 and M=23.91 SD=4.48 respectively) when compared with risk-averse participants (M=22.11 SD=5.78 and M=22.47 SD=6.03 respectively). However, this was not the case for the collaborative mode. This shows that although risk-taking moves informed the agent what to do when in undesirable situations, not all risk-taking moves were of the same usefulness. Risks performed in competition or isolation could be more valuable for learning that those in collaboration.

Although our results indicate that having a higher percentage of moves where the player is within three steps of the ghost produces better agent performance for two game types, we did not investigate to what extent risk-taking would continue to be beneficial. Moreover, in terms of risks, we did not investigate the length of time the player spent near the ghost or how risks varied with the distance from the ghost. The number of steps being three to constitute a feature was chosen somewhat arbitrarily - it is a small number, yet accounts for some reaction time.

The rates of risky moves made in the three modes can be observed in Table 2. The percentage of risky moves was highest in the competitive mode, as expected, and lowest in the solo mode. During the collaborative mode, participants would intentionally approach closer to the ghost as they saw that the agent was already playing safely, and hence they would collect the Pacdots in riskier locations. The same can be said for the competitive mode. Players were more likely to get closer to the ghost as their aim was to collect more objectives than the agent. On the other hand, in the solo mode, players were less likely to be making risks since they did not feel a sense of competition and did not have anyone to help.

We define *maze completion rate* as the percentage of games where all the objectives in the map had been collected by either the player or the agent. As expected, the completion rate was the highest in the collaborative mode as the two Pacman characters could help each other in collecting all the Pacdots. Interestingly, the maze completion rate was the lowest in the competitive mode despite there being two Pacman characters. This could have resulted from players making riskier plays in the competitive mode and dying a lot more.

6 CONCLUSION

In this paper, we examined how demonstration samples from different interaction modes affect agent performance and human perception, such as user preferences and behaviours. We also looked at the impact of mixing task modes to provide learning curricula for the agent.

We found that despite the smaller numbers of demonstration samples in the interactive modes, these generated comparable performance to those in the solo mode. The learning speeds of the agent in the solo and competitive modes were observed to be very similar. In addition, the participants reported that they preferred playing with/against an agent compared to playing alone. Out of the three modes, the participants liked playing competitively the most. The agent was also perceived to be better in this mode than the collaborative mode, despite the marginally worse in-task mean score and death rate. Our results show that the participants' perceived agent competency and improvement was inaccurate due to the influence of the human-agent interaction type. This could suggest that there should be additional work to improve the indication of agent performance to the user. Higher agent perceived competency can potentially induce higher user trust and therefore acceptance. Learning from demonstration procedures may therefore benefit

Table 2: Player Behaviour and Maze Completion Rate in Different Game Modes

Game Mode	Player Death Rate	Risky Move Rate	Maze Completion Rate
Solo	4.25%	8.28%	96.00%
Collaborative	4.75%	11.41%	97.75%
Competitive	11.25%	12.94%	94.50%

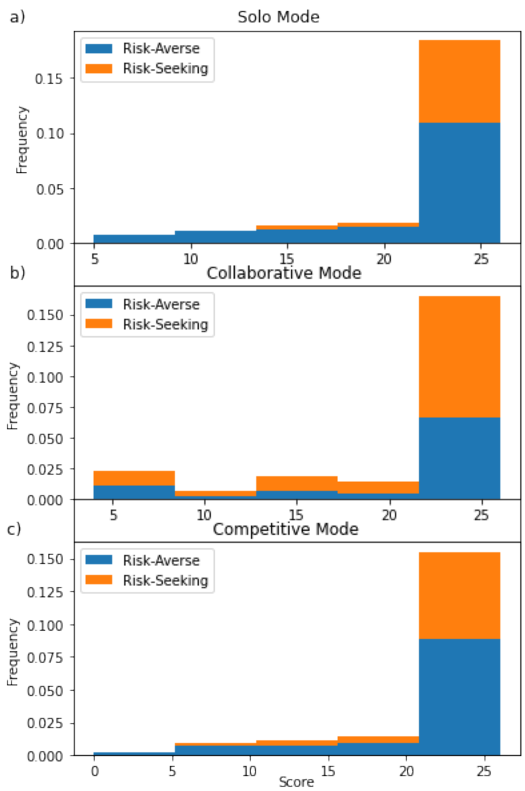

Figure 6: Score distribution between risk-averse and rise-seeking participants in a) solo mode, b) collaborative mode, and c) competitive mode

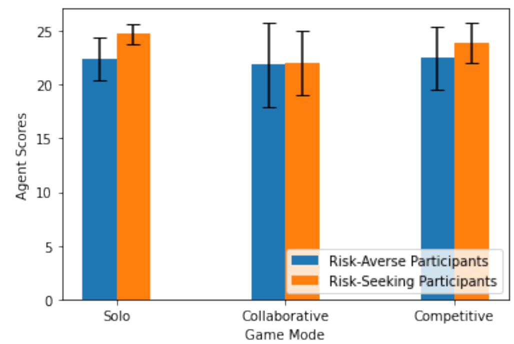

Figure 7: Mean agent scores from risk-averse and risk-seeking participants' samples

from implementing interactive modes, particularly competitive, in order to improve the teaching experience.

In terms of curriculum learning, we investigated how the ordering of the different modes would influence agent learning. Learning from demonstration in the collaborative mode early on in the curriculum could be observed to be noticeably slower than the other two modes due to the differences in the strategies used. We also found that there was only statistical significant difference between two mixed curricula (Solo-Competitive and Collaborative-Competitive). We concluded that building a curriculum through different modes only is ineffective as all three modes have the same level of task complexity. Our work also did not examine on the effects of sample proportions as we split the samples into two equal sized groups.

We found that the rate of risks made by the demonstrator in the solo and competitive modes improved the agent performance with statistical significance, while this was not the case for the collaborative mode. This is indicative that risks made in isolation or competition were more useful than those in collaboration for agent learning. Depending on the interaction type, the amount and type of risks to be taken in the demonstrations may need to be considered.

Although the task in this experiment was collecting objectives in a computer game, the knowledge from this work can be transferred to other non-gaming tasks such as physical objective collection or navigation as the mechanics are very similar. Human adversary could potentially be a suitable method for agent training, considering its positive influence on agent performance and the human perspective. Our work indicates that more research into different teaching interaction styles should be undertaken. The addition of competitive interactions could potentially be a way to improve the process for both the human and the agent.

ACKNOWLEDGMENTS

This research is partially supported by the Australian Research Council for their partial funding contribution (Grant No. DE210100858).

REFERENCES

[1] Saleema Amershi, Maya Cakmak, William Bradley Knox, and Todd Kulesza. 2014. Power to the People: The Role of Humans in Interactive Machine Learning. *AI Magazine* 35, 4 (Dec. 2014), 105–120.
[2] Michael Bain and Claude Sammut. 1995. A Framework for Behavioural Cloning. (1995), 103–129.
[3] Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. 2009. Curriculum learning. In *Proceedings of the 26th Annual International Conference on Machine Learning - ICML '09*. ACM Press, Montreal, Quebec, Canada, 1–8.
[4] De'Aira Bryant, Jason Borenstein, and Ayanna Howard. 2020. Why Should We Gender?: The Effect of Robot Gendering and Occupational Stereotypes on

Human Trust and Perceived Competency. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, Cambridge United Kingdom, 13–21.

[5] John Denero and Dan Klein. 2014. The Pac-Man Projects. http://ai.berkeley.edu/project_overview.html

[6] Jiali Duan, Qian Wang, Lerrel Pinto, C.-C. Jay Kuo, and Stefanos Nikolaidis. 2020. Robot Learning via Human Adversarial Games. *arXiv:1903.00636 [cs]* (Dec. 2020).

[7] Shane Griffith, Kaushik Subramanian, Jonathan Scholz, Charles L Isbell, and Andrea L Thomaz. 2013. Policy Shaping: Integrating Human Feedback with Reinforcement Learning. In *Advances in Neural Information Processing Systems*, Vol. 26. Curran Associates, Inc.

[8] Jonathan Grizou, Manuel Lopes, and Pierre-Yves Oudeyer. 2013. Robot learning simultaneously a task and how to interpret human instructions. In *2013 IEEE Third Joint International Conference on Development and Learning and Epigenetic Robotics (ICDL)*. 1–8.

[9] Masashi Hamaya, Kazutoshi Tanaka, Yoshiya Shibata, Felix von Drigalski, Chisato Nakashima, and Yoshihisa Ijiri. 2021. Robotic Learning From Advisory and Adversarial Interactions Using a Soft Wrist. *IEEE Robotics and Automation Letters* 6, 2 (April 2021), 3878–3885.

[10] Peter A. Hancock, Deborah R. Billings, Kristin E. Schaefer, Jessie Y. C. Chen, Ewart J. de Visser, and Raja Parasuraman. 2011. A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction. *Human Factors* 53, 5 (Oct. 2011), 517–527. Publisher: SAGE Publications Inc.

[11] Tasneem Kaochar, Raquel Torres Peralta, Clayton T. Morrison, Ian R. Fasel, Thomas J. Walsh, and Paul R. Cohen. 2011. Towards Understanding How Humans Teach Robots. In *User Modeling, Adaption and Personalization (Lecture Notes in Computer Science)*, Joseph A. Konstan, Ricardo Conejo, José L. Marzo, and Nuria Oliver (Eds.). Springer, Berlin, Heidelberg, 347–352.

[12] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems* 30 (2017).

[13] Jinying Lin, Zhen Ma, Randy Gomez, Keisuke Nakamura, Bo He, and Guangliang Li. 2020. A Review on Interactive Reinforcement Learning From Human Social Feedback. *IEEE Access* 8 (2020), 120757–120765.

[14] Zhiyu Lin, Brent Harrison, Aaron Keech, and Mark O. Riedl. 2021. *Explore, Exploit or Listen: Combining Human Feedback and Policy Model to Speed up Deep Reinforcement Learning in 3D Worlds*. Technical Report arXiv:1709.03969. arXiv.

[15] Luca Longo. 2018. Experienced mental workload, perception of usability, their interaction and impact on task performance. *PLOS ONE* 13, 8 (Aug. 2018), e0199661.

[16] James MacGlashan, Mark K. Ho, Robert Loftin, Bei Peng, Guan Wang, David L. Roberts, Matthew E. Taylor, and Michael L. Littman. 2017. Interactive Learning from Policy-Dependent Human Feedback. In *Proceedings of the 34th International Conference on Machine Learning*. PMLR, 2285–2294.

[17] Sanmit Narvekar, Bei Peng, Matteo Leonetti, Jivko Sinapov, Matthew E Taylor, and Peter Stone. 2020. Curriculum learning for reinforcement learning domains: A framework and survey. *arXiv preprint arXiv:2003.04960* (2020).

[18] Hai Nguyen and Hung La. 2019. Review of Deep Reinforcement Learning for Robot Manipulation. In *2019 Third IEEE International Conference on Robotic Computing (IRC)*. 590–595.

[19] Reinhard Pekrun, Thomas Goetz, Lia M Daniels, Robert H Stupnisky, and Raymond P Perry. 2010. Boredom in achievement settings: Exploring control–value antecedents and performance outcomes of a neglected emotion. *Journal of educational psychology* 102, 3 (2010), 531.

[20] Zhipeng Ren, Daoyi Dong, Huaxiong Li, and Chunlin Chen. 2018. Self-paced prioritized curriculum learning with coverage penalty in deep reinforcement learning. *IEEE transactions on neural networks and learning systems* 29, 6 (2018), 2216–2226.

[21] Shane Storks, Qiaozi Gao, Govind Thattai, and Gokhan Tur. 2021. *Are We There Yet? Learning to Localize in Embodied Instruction Following*. Technical Report arXiv:2101.03431.

[22] Unity Technologies. 2022. Unity Real-Time Development Platform | 3D, 2D, VR & AR Engine. https://unity.com/

[23] Andrea L. Thomaz and Cynthia Breazeal. 2008. Teachable robots: Understanding human teaching behavior to build more effective robot learners. *Artificial Intelligence* 172, 6 (April 2008), 716–737.

[24] Seung Hee Yoon and Stefanos Nikolaidis. 2020. Robot Learning in Mixed Adversarial and Collaborative Settings. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 9329–9336.