Understanding Human-AI Cooperation Through Game-Theory and Reinforcement Learning Models

Beau G. Schelble[1], Christopher Flathmann[1], Nathan McNeese[1], Lorenzo Barberis Canonico [1],
[1]Human-Centered Computing, Clemson University
(bschelb,cflathm,mcneese,lorenzb)@g.clemson.edu

Abstract

For years, researchers have demonstrated the viability and applicability of game theory principles to the field of artificial intelligence. Furthermore, game theory has been shown as a useful tool for researching human-machine interaction, specifically their cooperation, by creating an environment where cooperation can initially form before reaching a continuous and stable presence in a human-machine system. Additionally, recent developments in reinforcement learning artificial intelligence have led to artificial agents cooperating more efficiently with humans, especially in more complex environments. This research conducts an empirical study to understand how different modern reinforcement learning algorithms and game theory scenarios could create different cooperation levels in human-machine teams. Three different reinforcement learning algorithms (Vanilla Policy Gradient, Proximal Policy Optimization, and Deep Q-Network) and two different game theory scenarios (Hawk Dove and Prisoners dilemma) were examined in a large-scale experiment. The results indicated that different reinforcement learning models interact differently with humans with Deep-Q engendering higher cooperation levels. The Hawk Dove game theory scenario elicited significantly higher levels of cooperation in the human-artificial intelligence system. A multiple regression using these two independent variables also found a significant ability to predict cooperation in the human-artificial intelligence systems. The results highlight the importance of social and task framing in human-artificial intelligence systems and noted the importance of choosing reinforcement learning models.

1. Introduction

Human-artificial intelligence (AI) systems have a massive potential to outperform either agent alone. Human-AI systems are characterized by at least one human and at least one AI interacting with one another in a shared environment or task, as seen in other recent work [1]. This potential began to be shown when IBM's artificial intelligence (AI) system Deep Blue defeated Kasparov [2], shifting the field of human-AI interaction. Human-AI systems' ability was shown explicitly in Kasparov's "advanced chess" tournament (where AIs, humans, and human-AI systems compete against each other), highlighting a human-AI system that could defeat both the top AI as well as the top human chess players. This team consisted of an amateur human and a mediocre AI [3]. The finding suggested that successful collaboration between humans and AI is certainly possible.

The enhanced ability behind many human-AI systems seen in recent research lies in leveraging either agent's strengths. As research published in Nature shows bots attempting to solve a graph coloring problem requiring high levels of coordination, fails to achieve a globally optimal solution [4]. The study points out a way to transcend the limitations of bots coordinating on a macro level by adding humans to the team [4]. Successful human-AI systems are capable of doing more than merely playing Chess and solving graph coloring problems; they extend even to the medical field. Specifically, in cancer detection, as teams of doctors partnering with machine learning algorithms outperformed both expert teams as well as state-of-the-art neural networks in diagnosing cancer [5]. This level of success can be attributed to the unique advantages that emerge from harvesting human and AI potential in a compatible and integrated way and should pave the way for more research of this kind.

However, these human-AI systems are not without their challenges. Prior research has shown that effective team behavior occurs when each team member seeks to model their teammates' thought process, which is inherently more challenging in a human-AI system [6]. Specifically, humans tend to distance themselves from teammates they perceive to be autonomous, and AIs tend to avoid wanting to cooperate with human agents who

URI: https://hdl.handle.net/10125/70652
978-0-9981331-4-0

do not share their thought process [7]. These challenges can make cooperation in human-AI systems difficult, especially when coupled with the fact that cooperation in these human-AI systems can be significantly altered by the nature of the task and the most mutually beneficial outcome, a problem compounded by the fact that task and social framing varies widely from system to system.

Task framing is the reason for taking action, while social framing is the context of the agents' relationship and the results of taking an action [8]. The importance of task design and group dynamics is mentioned in recent research agendas on human-machine interaction [9], emphasizing the importance of addressing these aspects of human-AI systems as the current study does. Choosing the right reinforcement learning model (RL) from the many available contributes to the problem, as the chosen model may impact their ability to cooperate.

RL agents become more prevalent in applied settings around the world as time goes on, serving in a variety of different industries [10, 11], the need for further research to clarify the dynamics behind human-AI systems is obvious. With the numerous RL models available to practitioners, there is a specific need to highlight the dynamics behind the RL model used on human-AI system dynamics like cooperation. Along with the various settings and contexts that human-AI systems are deployed to, it is vital to identify how the social and task framing of these systems may potentially alter system outcomes.

The current paper leverages the use of two similar but distinct game theory scenarios that specifically emphasize cooperation to construct strategic interactions. The experimental setups incorporate three state-of-the-art RL models whose strategic behavior illuminates their different receptiveness to specific incentive structures, while the two game theory scenarios emphasize differences in social and task framing for these same systems. In order to capture these differences, the current study focuses on answering the following research questions:

- **Research Question 1**: Does average overall cooperation in a human-AI system differ based on the reinforcement learning model used?

- **Research Question 2**: What effect does the game theory scenario have on average overall human-AI system cooperation?

- **Research Question 3**: Can the game theory scenario and reinforcement learning model used to predict overall human-AI system cooperation?

2. Related Work

Game theory is the study of the decision-making process of self-interested agents in strategic situations. Emerging from the intersection between mathematics and economics, it functions as a highly appropriate framework to conduct AI research in cooperation as it provides a mathematical common ground that humans can understand, and AI can train to be experts in. Hence, a reward-maximizing agent embodies the definition of a "rational" player within these specific contexts; however, this does not mean this mathematical rationality would extend to other contexts as complexity, and environmental factors could change. While a human could achieve this rationality, AI would have greater consistency in being rational in game theory specific scenarios. While it is out of the scope of this study, human rationality in different environments could be further explored through concepts such as bounded rationality, which would provide a more human-centered definition of rationality, especially in contexts outside of these simplistic game theory scenarios. However, the game theory rationality studied here is essential in reaching a Nash Equilibrium, where each individual's strategies can converge and mutually respond to each other [12]. The use of game theory models, and therefore game theory scenarios imply that players are going to converge to this equilibrium.

2.1. Game Theory

While the Nash Equilibrium can be present in a wide variety of scenarios, the optimal equilibrium can differ from scenario to scenario. Due to the existence of scenario-specific optimal strategies, research efforts have created Matrix Game Social Dilemmas (MGSDs), which allow game theory principles to be applied to a variety of scenarios to elicit multiple factors in creating group strategy, including group reciprocity, norm enforcement, and social network effects [13]. The design of these scenarios has resulted in the creation of games where individual players are not able to succeed solely through an individualist mindset, but rather through group strategy [14].

Due to the implicit goal of reaching a group strategy, game theory provides a potentially beneficial lens for viewing human-AI interaction through; most notably, game theory can be used to evaluate the team's ability to coordinate and cooperate [15]. Specifically, MGSDs provide a powerful method for evaluating human-AI cooperation as they provide a variety of social and task contexts that can be used to engage joint strategy within a team [16]. The formation of these strategies depends heavily on important teaming factors, like fairness, coordination, reciprocity, and cooperation, which is the focus of this study [17]. Due to the factors that contribute to efficient game theory models, which are essential to teams in general, game theory serves as a powerful tool for observing human-agent interaction.

Due to the limitations of AI in representing the expansive real world, limiting human-AI system interaction around MGSDs provides a capable and fair environment for observing and understanding human-AI interaction. The matrix design of MGSDs and their social nature create a platform that can put humans and AI on an even playing field. This methodology, and the benefits outlined above, can be extended to the context of human-AI systems, specifically involving AI built with RL, to observe human-AI cooperation and the potential of using game theory to predict and plan human-AI strategy.

2.2. Reinforcement Learning

RL is a class of machine learning algorithms that are based upon behavioral models that reward and punish behavior for inducing the discovery of a unique policy, mapping situations to actions as to maximize positive rewards over time [18]. These tradeoffs and strategies are balanced through a series of hyperparameters inherent to each RL model, giving unique advantages based on the underlying algorithm. Some of the most widely used modern RL models include Deep Q-Network (DQN) [19], Vanilla Policy Gradients (VPG) [20], and Proximal Policy Optimization (PPO) [21]. In terms of similarities and differences, VPG and PPO are more closely related to each other as they are both on-policy methods while the DQN is an off-policy method, which, simplistically, means the DQN maximizes the utility of target states while VPG and PPO maximize the utility of the current state. These optimization differences could lead to a higher level of convergence by the DQN, especially in the more simplistic environments used in game theory.

One of RL's greatest strengths is creating an understanding of an environment through simple board states, which has led to RL models being highly skilled in a variety of games, such as Go, Chess, and soccer [22]. This strength is made possible through self-play, where AI agents can repeatedly play many games over time to develop a sophisticated understanding of their environment, which can be represented as a simplistic board state and reward signals [23]. These skills are important to navigating game theory scenarios as they too can be represented as matrix-based board states that can be navigated by AI systems, which includes the ability to learn a cooperative strategy from the self-play of simulations of game-theory games.

While RL's strengths allow it to find optimal paths in more simplistic contexts, such as older video games, RL has shown a deficiency in understanding more complex environments where multiple solutions exist and the possibility of getting stuck in a local optimum increases [24]. Advances in RL have seen these problems begin to be mitigated; for example, J.W. Crandall's work in human-AI cooperation has produced a novel RL model that ensures payoffs are at a minimum of the game and also learns to cooperate [25]. These findings are in addition to more recent work showing the newly developed S# model to be fully capable of cooperating with human players and AI players in game-theoretical situations like Prisoner's Dilemma, which require intuition and are affected by cultural norms and emotion [26]. Due to the potential RL AI has shown in the past years for understanding complex environments, it would prove to be an essential tool in human-AI cooperation since teams are set in more realistic environments. However, despite these advancements, there is still a lack of empirical research on human-AI system's ability to converge and cooperate on optimal strategies when observed in differing social and task framing.

2.3. Human-AI Systems

Human-AI systems involve two or more agents which consist of at least a single human, and a single AI. The principal obstacle has been avoiding limiting each agent to only local information to make processing a complex environment tractable. Markovian games have emerged as the primary model for human-AI systems that include RL because they enable a distributed decision-making process and a stochastic environment [27]. The dominant approach in those settings relies on the Nash-Q algorithm for general-sum stochastic games, which enables RL agents to converge towards stable strategies in zero-sum and common payoff games. Such restrictive parameters have made Nash-Q challenging to generalize from [28]. Furthermore, such tight game structures do not lend themselves to optimal solutions for games with multiple NEs, making learning in such settings nontrivial [29].

Prior human-AI research has focused on independent RL, where each RL agent is not aware of the other agents and instead senses them as part of the interactions with its environment. Such non-stationary environments violate the Markovian property, which undermines the generalizability of the policies the agents learn [30]. Specifically, human-AI system strategies rarely converge towards an optimal equilibrium because, as long future rewards are highly discounted, agents may not risk deviating from a suboptimal equilibrium [31]. Alternatives to independent RL involve creating special-purpose algorithms (WoLF, JAL, AWESOME) that privilege rationality and joint-action in many cooperative scenarios [32]. However, recent research has pointed out that such algorithms cannot shape the learning behavior of the opponents to obtain higher payouts at convergence, especially over repeated games [22].

The current study, however, does not focus on hyperparameter tuning, algorithm development, or other technical advancements, but instead on the human aspects affecting the dynamics behind these human-AI systems, hoping to understand better and predict those systems outcomes.

3. The Current Study

The current research study reports on an experiment in which a human-AI system played two game theory scenarios, Hawk Dove and Prisoners Dilemma (detailed below in a later section). Each game theory scenario focuses around cooperation in order to achieve the optimal expected reward, but the motive and concept of the two scenarios are unique. Participants also interacted with three different reinforcement learning models: DQN, VPG, and PPO. These two variable groups represented the two independent variables (IV) manipulated for this experiment: 1) game theory scenario (Prisoners Dilemma, Hawk Dove), and 2) reinforcement learning model (DQN, VPG, PPO). All independent variables were examined resulting in a 2x3 factorial design conducted between subjects. Based on the experimental design and previous research the following hypothesis can be considered regarding RL algorithms: (1) Due to the algorithmic design of DQN models, we would expect them to achieve higher levels of cooperation. Regarding scenario choice, this study elects to take a more exploratory approach to the effects the social framing of each scenario could have on cooperation rather than hypothesizing the specific superiority of either scenario.

3.1. Participants

This experiment recruited 226 participants from Amazon Mechanical Turk to participate in the experiment, resulting in 226 human-AI systems completing the experiment. Participants' demographics were as follows: gender: 145 Male, 80 Female, 1 Other, Age: 63 between 18-25 years, 94 between 26-35 years, 41 between 36-45 years, 14 between 46-55 years, 14 between 56-65 years. The Prisoners Dilemma condition consisted of 103 human-AI systems, while the Hawk Dove condition consisted of the remaining 123 human-AI systems. The number of human-AI systems per reinforcement learning model is shown below in Table 1. The imbalance in the number of human-AI systems completing the tasks was a result of systems being dropped from the analysis for incomplete data recording during the completion of the game theory task, which was the result of client-side connectivity issues.

Table 1. Participant Numbers

Prisoners Dilemma: **103**		
DQN: **43**	VPG: **27**	PPO: **33**

Hawk Dove: **123**		
DQN: **43**	VPG: **39**	PPO: **41**

3.2. Task

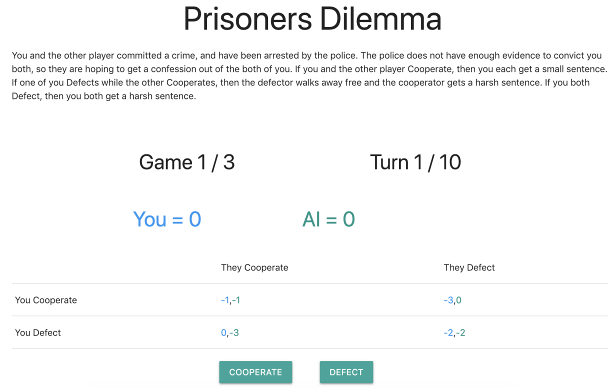
The cooperative game theory scenarios known as Prisoners Dilemma and Hawk Dove were selected to provide a broad analytical base to identify the extent to which different factors affect the willingness to cooperate with both the human players and the reinforcement learning agents. While both scenarios target cooperation between the two players, the fundamental motivations and concepts are unique, making a detailed description of each scenario necessary.

3.2.1. Prisoners Dilemma The Prisoners Dilemma is an ideal scenario in game theory where two players are posited to have been arrested by authorities for committing a crime. Once apprehended, each player is separated from the other so that they are unable to communicate. Because the police do not have sufficient evidence to convict both players, they offer each player the opportunity to confess to gain a lighter sentence at the expense of the other player.

Prisoner's Dilemma's core result is that the Nash Equilibrium induces both players to confess, leading to the collectively worst outcome for both players. However, in experimental settings, this dynamic often changes when the Prisoners Dilemma is played iteratively. The iterative nature is because a sequential Prisoner's Dilemma creates the opportunity for players to punish one another for defecting from an agreement to remain silent, thus creating a reasonable expectation of cooperation.

Figure 1 is taken from the Prisoners Dilemma interface of the custom experimental platform. The payoffs for mutual cooperation are -1 for each player, the payoffs for mutual defection are -2 for each player, and the payoff for successfully defecting on a cooperative player are 0 and -3, respectively. In a Prisoner's Dilemma, defecting is the dominant strategy because both players are better off defecting, given what they expect the other player to do. Axelrod's famous tournament that included both human and computer-generated solutions found the tit for tat strategy to be most beneficial, effectively repeating the partner's previous decision [33].

Figure 1. Interface for the Prisoners Dilemma game theory scenario



3.2.2. Hawk Dove The Hawk Dove game is a more dynamic version of a Prisoner's Dilemma where each player is faced with a decision of whether to attack or to remain peaceful. Hawk Dove is symmetric, so each player abides by the same incentive structure that rewards peace (0 payoff) over war (-2 payoff). The only situation in which any player is better than being peaceful is by successfully attacking when the other player selects to be peaceful.

The Hawk Dove scenario looked the same as Figure 1 but the scenario, title, and reward square was modified to match the Hawk Dove scenario (exact reward square defined in measures). A successful attack occurs when one of the players decides to remain peaceful, and results in one point being transferred from the peaceful player to the attacker. This payoff is significant because it results in a smaller loss for the peaceful player when attacked than when mutually attacking. This aspect is essential because it creates a somewhat powerful incentive to remain peaceful, implying that attacks result from an essential zero-sum mentality driving the player.

3.3. Materials and Equipment

A custom experimental platform was developed to accommodate the current study consisting of an interface that supported each of the experimental conditions. The interface for Prisoners Dilemma is shown below in Figure 1, and the interface for Hawk Dove use the same format with slight modification to the content provided. Each move by both players was recorded by the application and stored on a server. Each player plays three rounds of the game theory scenario for ten turns, with the score re-setting every round.

The open-source RL framework *TensorForce* was utilized to implement the RL agents. The TensorForce library is focused on providing explicit APIs, readability, and modularization to deploy RL solutions both in research and practical applications [34, 35].

3.4. Procedure

Participants were recruited through the Amazon Mechanical Turk (MTurk) platform, a platform that allows researchers to recruit participants worldwide in return for monetary compensation [36]. The MTurk platform is highly reliable and hugely representative of the population compared to typical university subject pools [37].

For the current study, participants were randomly assigned to conditions. The first thing participants did was give informed consent to participate in the study, which, after consenting to participate in the study, saw the experiment begin automatically. The participants were shown one of two interfaces depending on the game theory scenario they were grouped into, (Figure 1 shows the basic layout). Directions to the scenario were shown at the top of the interface, with the players' scores directly below, followed by the outcome table and possible decisions. After playing the game for ten turns the interface reinitialized and began a new game. Participants played three games for a total of thirty turns, with data being automatically collected and stored by the platform. Once participants had completed the three games, they were directed to a Qualtrics survey for demographics collection and quality assurance. Quality assurance involved the participants answering a question to prove they were a human and were taking the experiment seriously. Upon completing the experiment, participants were paid $1 for their time (roughly ten minutes).

3.5. Measures

Average cooperation was the only dependent variable recorded by the current experiment. The operationalization of cooperation is necessary to clearly define due to the unique differences in the reward structure of the Prisoners Dilemma and Hawk Dove scenarios. As can be seen in Figure 1, the participants are shown the reward structure for their decisions (Hawk Dove reward structure square starting from top left to right: -2,-2, 1,-1, -1,1, 0,0). The 2x2 decision tree consists of four outcomes, regardless of the score displayed. These outcomes include the following: 1) Player A and Player B both do not cooperate, 2) Player B cooperates while Player A does not, 3) Player A cooperates while Player B does not, 4) Both players cooperate. Accordingly, the experimental platform's result was a number between 1 and 4, with 1 being low levels of cooperation and 4 being high levels of cooperation. This result was recorded for each turn in all three of the ten turn games. The results were then averaged for average overall cooperation in the human-AI system. Outcome 2 is considered lower cooperation than outcome 3 because the AI is rewarded based on the global performance, making it's tendency to cooperate inherently higher than the human agent's tendency to cooperate. This ordering also aligns with the general philosophy of game theory as the AI is much more likely to play rationally.

4. Results

Due to the unequal sample sizes, the assumption of homogeneity of variances was violated for this data set, making the use of non-parametric tests necessary; however, normality of the data set was maintained. Accordingly, the recommendations of current literature in statistical analysis were followed [38]. Notably, in order to minimize the chances of committing a Type 1 error, independent groups were compared using the Mann-Whitney U test, while the Kruskal-Wallis test was used to compare three independent groups. Post-hoc comparisons to Kruskal-Wallis tests were completed using Mann-Whitney U tests with Bonferroni adjusted p values. Finally, to do more than compare independent group means and determine predictability, the current analysis utilized a heteroskedasticity-consistent standard error estimator for ordinary least squares regression, as detailed by Hayes and Cai [39]. The results of these analyses are detailed in the following section, organized by research question. Additionally, gender and age data revealed no differences when used as control variables.

4.1. RQ1: Does Cooperation Change Based on the AI Model Used?

In order to determine if significant differences existed in the overall cooperation of the human-AI systems between independent groups with different RL models a Kruskal-Wallis test was ran on the data set in its entirety.

Table 2. Kruskal-Wallis Test on Game Theory Scenario and Overall Cooperation

Kruskal-Wallis Test			
n	*H*	*df*	*p*
226	36.22	2	$< .001$
Post-Hoc Tests			
RL Models	*U*	*Z*	*p*
DQN-VPG	1650	-4.42	$< .001$
DQN-PPO	1581.50	-5.48	$< .001$
VPG-PPO	2015.50	-1.78	.075

The Kruskal-Wallis test (see Table 2) showed that the RL model used significantly affected overall human-AI system cooperation, $H(2) = 36.22$, $p = <.001$. Post-hoc Mann-Whitney U tests using a Bonferroni-adjusted alpha level of .017 (0.05/3) was used to determine the significance of each pairwise comparison. The comparison between human-AI systems using the DQN RL model and VPG RL model was significant, $U(N_{DQN} = 86, N_{VPG} = 66) = 1650.00$, $Z = -4.42$, $p <.001$. The difference in overall cooperation between human-AI systems using the DQN RL model and the PPO RL model was significant, $U(N_{PPO} = 74, N_{DQN} = 86) = 1581.50$, $Z = -5.48$, $p <.001$. All other pairwise comparisons were not statistically significant, along with a follow up Kruskal-Wallis test between RL models and improvement over the three games was not significant. Finally, to test the interaction effect between game theory scenario and RL model a two-way ANOVA was used; however, the results of this test should be interpreted carefully as ANOVA's are robust to violations of homoskedasticity, but only with roughly equal sample sizes. ANOVA results revealed a significant interaction effect between game theory scenario and RL model, $F(2, 222) = 35.69$, $p < .001$, $\eta^2 = .25$.

Based on these results, we can tell that the AI model used did have a significant effect on overall cooperation within the human-AI systems. While the VPG and PPO RL models had very similar cooperation levels, the DQN RL model had much higher levels of cooperation, lending credence to the notion that the RL model AI use to train impacts their interactions with human agents. This result also supports the earlier hypothesis that the

DQN model would produce higher levels of cooperation in human-AI systems.

4.2. RQ2: Is Cooperation Affected by Game Theory Scenario?

To investigate whether overall human-AI system cooperation was affected by the game theory scenario the system completed, a Mann-Whitney U test was ran on the two independent groups data set (see Table 3).

Table 3. Mann-Whitney U Test on Game Theory Scenario and Overall Cooperation

Mann-Whitney U Test			
Variable	*U*	*p*	r_{pb}
Average Cooperation	4575.50	$< .001$	-.281
Mean and Standard Deviation			
Scenario	*n*	*Mdn*	*SD*
PD	103	2.7	.51
HD	123	2.98	.58

Descriptive statistics revealed that the overall cooperation of human-AI systems completing the Prisoners Dilemma scenario (*Mdn* = 2.7), were lower than those completing the Hawk Dove scenario (*Mdn* = 2.98). The Mann-Whitney U test indicated that this difference was statistically significant, $U(N_{PrisonersDilemma}$ = 103, $N_{HawkDove}$ = 123) = 4575.50, z = -3.60, p <.001, r_{pb} = -.281. A follow up Mann-Whitney test between game theory scenarios and improvement over the three games was not significant.

This analysis provides additional clarity to the importance of the task and social framing in human-AI systems. As stated previously, while the Prisoners Dilemma and Hawk Dove scenarios target cooperation, the two have unique differences in context and motivation conveyed to the two players. While the current study cannot answer with certainty that the different contexts and motivations are the driving force behind these differences in cooperation, the results emphasize their impact.

4.3. RQ3: Can the Game Theory Scenarios and AI Model Used Predict Cooperation?

In order to move beyond simply ascertaining whether overall cooperation differences between the independent groups are significantly different an ordinary least squares regression must be utilized.

Running this regression gives the current study the ability to determine if the RL model and game theory scenario can predict the human-AI systems overall cooperation. To accomplish this a heteroskedasticity-consistent standard error multiple

Table 4. Game Theory Scenario and AI Model Linear Regression for Cooperation

Model Fit			
R^2	*F*	*df*	*p*
.209	26.76	3, 222	$< .001$
Score			
Variable	Coefficient	Std. Error	*p*
Constant	2.46	.06	$< .001$
DQN	.47	.08	$< .001$
VPG	.09	.07	.230
HD	.33	.07	$< .001$
Setwise Hypothesis Test			
F	df_{num}	df_{den}	*p*
12.15	2	222	$< .001$

ordinary least squares regression was used with the HC3 estimator to predict a human-AI systems overall cooperation from the game theory scenario used and the RL model used (see Table 4). As all variables were nominal each was dummy coded for use in the regression. The model explained a statistically significant amount of variance in overall cooperation, *F*(3, 222) = 26.76, p = <.001, R^2 = .21, $R^2{}_{adjusted}$ = .20. AI type DQN was a significant predictor of overall cooperation, β = .474, *t*(223) = 5.81, p = < .001. A change in game theory scenario saw human-AI systems overall cooperation increase by 0.474 points, *B* = 0.474, 95% CI [0.315, 0.633]. AI type VPG was not a significant predictor of overall cooperation, β = .085, *t*(223) = 1.20, p = .230. Alternatively, the game theory scenario Hawk Dove significantly predicted overall cooperation, β = .325, *t*(223) = 4.75, p = <.001. A change in game theory scenario saw human-AI systems overall cooperation increase by 0.325 points, *B* = 0.325, 95% CI [0.191, 0.459].

The results of this regression analysis showcase the impact of RQ1 and RQ2 based on the significant predictive ability of game theory scenario and RL model used on cooperation. This finding further cements the point made in RQ1 and RQ2 that the social and task framing conveyed to both humans and AI models are highly relevant to both human-AI system outcomes.

5. Discussion

Our results show meaningful differences in the cooperative dynamics between humans and AIs across various settings. Instead of limiting ourselves to just one game theory model, such as the often-used Prisoner's Dilemma, we explored the additional Hawk Dove scenario for human-AI cooperation and compared the impact of the similar but unique social and task framing of each. The results from each game can be analyzed separately, but should also be understood as indicative of a broader behavior pattern. Beforehand, it is important to briefly discuss how the use of our selected game theory scenarios and RL models could have resulted in our observed results.

Specifically, the data from both scenarios have implications for both AI as well as humans' cooperative dynamics. Cooperation manifested differently based on the RL model that the AI teammate utilized. This finding may be the result of the distinct ways in which DQNs process strategic interactions compared to PPOs and VPGs. This would be expected as the design of DQNs generally leads to higher levels of convergence overtime, which could result in a more cooperative agent in these scenarios. Generally, the downside to this algorithm would be the time it takes to build the model; however, the simplistic nature of game theory scenarios allows DQN models to be trained quickly, resulting in higher degrees of cooperation forming in similar training times to the PPO and VPG. It is essential to understand, consider, and compensate for these differences when implementing AI alongside humans. These differences would need to be clarified for the specific task a team is conducting, which will allow a more intelligent and deliberate choice when deciding the back-end design of AI teammates. For instance, the DQN model's higher cooperation in these two tasks would suggest that it be utilized in similar contexts to the Prisoner's Dilemma and Hawk Dove scenario; however, choosing a PPO or VPG model due to ignorance towards models differences could significantly reduce the cooperation within the human-AI system. Without this knowledge and design, significant performance differences could be seen between different human-AI systems despite them existing in similar contexts and environments. Additionally, differences in cooperation levels over time were not significant between conditions, highlighting stable development of cooperation between all conditions.

It is also important to note that AI safety researchers should not assume that the willingness for an AI to cooperate with humans in one scenario necessarily generalizes to all situations. Our setup goes a long way in establishing a strong basis to investigate human-AI interactions by testing cooperative dynamics across two unique game theory scenarios where cooperation is in the collective interest of the multi-agent system. Using game theory in this setting is useful because sharp deviations from Nash Equilibria indicate the complex nature of the interactions clearly. However, limiting empirical research on human-AI cooperation to just one game would have only provided scant evidence about AI and human players' behavioral patterns. While both of these studies and scenarios look to evaluate cooperation in human-AI systems, the actual context and motive given to the participants varied based on the game. While we cannot say that these contextual differences are the reason for *cooperation differences*, the existence of these differences highlights the importance of the social framing of a task and scenario, which is suspected to be the reason behind varying levels of cooperation between games. Therefore, it is essential to consider both team, task, and evaluation contexts when looking to understand human-AI systems. A lack of understanding in these areas could significantly change the utility and viability of powerful tools, such as game theory, to help evaluate and coordinate aspects of human-AI systems.

While the bulk of contribution of this study is to the field of human-AI interaction, additional implications exist regarding the field of game theory. Specifically, this study further contributes to the literature regarding the applicability of game-theory to evaluating human-AI interaction. Different game theory scenarios were shown to change the level of cooperation possible during human-AI interaction. These findings demonstrate that the value game theory can have in the promotion and encouragement of cooperation. Furthermore, while identifying game theory as an evaluation tool is not entirely novel, the ability to use game theory as an encouragement and social scoping tool is highly important to future game theory research, especially regarding AI's interactions with game theory. Building scenarios and tasks that are scoped within the design of game theory scenarios, especially with the framing of the Hawk Dove, could create tasks and environments that demonstrate a greater and more apparent benefit from human-AI cooperating.

The more back-end consideration of algorithm selection and the more user-facing consideration of social and task framing show that understanding and designing human-AI systems are reliant on multiple layers of human-AI interactions. The continued pursuit of advancing human-AI systems will need to consider these features, especially in the field of research where algorithm and task selection could significantly affect the results human-AI systems exhibit. As research in this area continues, a complete understanding of the factors that affect human-AI interaction can be achieved.

6. Limitations and Future Work

The following limiting factors should be taken into account when interpreting the results of this study. Response times were unable to be recorded during the game theory tasks to record the quality of the responses; quality checks could only be implemented in the post-task survey. Additionally, while the scoping of game theory used has identified advantages in observing strategic play, its use creates some partial limitations in this study. Firstly, the simplistic nature of the game theory scenarios used make it easier for cooperation to occur as the benefits can become apparent more quickly. Real-world environments and scenarios may not benefit from the same simplicity and may not be able to achieve high levels of cooperation in such a short amount of time. Secondly, game theory lends itself to a specific definition of rationality that can be viewed more simplistically and mathematically. However, rational behavior theories exist within humans, such as bounded rationality, which may go beyond the simplistic definition in the current study. These limitations do not mean that these results are not applicable to the real-world where humans have complex rationales, but it does mean that the relevancy of these results should not be considered without consideration for the game theory scoping used.

The two primary avenues for expanding upon this work involve the choice of RL models and game theory scenarios. For the former, RL's field is expanding so rapidly that new RL models have emerged that analyze strategic situations in different ways. This paper limited itself to DQNs, VPGs, and PPOs because they provide representative models from the classical, deep learning, and modern RL paradigms. The results strongly indicate that the RL model the agent operates by is not ancillary to the outcome of a human-AI interaction in a game theory setting; thus, it would follow that empirically testing additional models might also be useful to generate a complete picture of human-AI cooperation.

For the latter, many game theory scenarios would enable the exploration of human-AI cooperation under different incentive structures. Ideally, future research will focus on long-form games instead of iterative games since the cooperation's nature is different. This focus would expand game theory's viability to human-AI system interaction as specific game theory scenarios could be chosen based on the context and function of the human-AI system being evaluated. For example, the Centipede game, where participants take turns taking a slightly larger payoff or passing on a pot of rewards, would help identify how backward induction plays a role in human-AI cooperation. This type of scenario could be highly applicable in teams that mostly function asynchronously but are highly dependent on shared resources.

Overall, using cutting-edge RL models as well as context applicable and extended duration games can shed light on a different type of human-AI cooperation. While this study provides insight into the viability of using game theory to understanding human-AI systems, further research efforts are required to ensure broader applicability of game theory to real-world human-AI systems.

7. Conclusion

A significant question in AI safety and AI research as a whole for the years to come will be how to train humans and AIs to work together. Reinforcement learning is quickly becoming the dominant machine learning paradigm because of its generalizability. Thus, it is crucial to understand how the tasks human-AI systems face need to be framed and the task-specific benefits of differing agent design. To that end, this paper's methodology shows how different game theory models can be used to frame human-AI systems and better understand cooperation differences based on context. To that end, it is essential that the understanding of this research is further expanded to understand a more extensive variety of contexts, specifically related to human-AI interaction. As human-AI systems continue to progress into more environments, understanding the impact that task and social framing and context have on interaction, along with the underlying algorithms used for AI, will be vital to human-AI cooperation.

8. Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 1829008.

References

[1] M. H. Jarrahi, "Artificial intelligence and the future of work: Human-ai symbiosis in organizational decision making," *Business Horizons*, vol. 61, no. 4, pp. 577–586, 2018.

[2] F.-H. Hsu, *Behind Deep Blue: Building the computer that defeated the world chess champion*. Princeton University Press, 2004.

[3] C. Thompson, "Clive thompson on the cyborg advantage," *Wired Mag*, 2010.

[4] H. Shirado and N. A. Christakis, "Locally noisy autonomous agents improve global human coordination in network experiments," *Nature*, vol. 545, no. 7654, p. 370, 2017.

[5] D. Wang, A. Khosla, R. Gargeya, H. Irshad, and A. H. Beck, "Deep learning for identifying metastatic breast cancer," *arXiv preprint arXiv:1606.05718*, 2016.

[6] N. J. McNeese, M. Demir, N. J. Cooke, and C. Myers, "Teaming with a synthetic teammate: insights into human-autonomy teaming," *Human factors*, pp. 262–273, 2017.

[7] M. Demir, N. J. McNeese, and N. J. Cooke, "The impact of perceived autonomous agents on dynamic team

behaviors," *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2018.

[8] P. M. Miller and N. S. Fagley, "The effects of framing, problem variations, and providing rationale on choice," *Personality and Social Psychology Bulletin*, vol. 17, no. 5, pp. 517–522, 1991.

[9] I. Seeber, E. Bittner, R. O. Briggs, T. de Vreede, G.-J. De Vreede, A. Elkins, R. Maier, A. B. Merz, S. Oeste-Reiß, N. Randrup, *et al.*, "Machines as teammates: A research agenda on ai in team collaboration," *Information & management*, vol. 57, no. 2, p. 103174, 2020.

[10] K. Mason, P. Mannion, J. Duggan, and E. Howley, "Applying multi-agent reinforcement learning to watershed management," in *Proceedings of the Adaptive and Learning Agents workshop (at AAMAS 2016)*, 2016.

[11] L. Zheng, J. Yang, H. Cai, M. Zhou, W. Zhang, J. Wang, and Y. Yu, "Magent: A many-agent reinforcement learning platform for artificial collective intelligence," in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[12] J. F. Nash *et al.*, "Equilibrium points in n-person games," *Proceedings of the national academy of sciences*, vol. 36, no. 1, pp. 48–49, 1950.

[13] J. Z. Leibo, V. Zambaldi, M. Lanctot, J. Marecki, and T. Graepel, "Multi-agent reinforcement learning in sequential social dilemmas," in *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, pp. 464–473, International Foundation for Autonomous Agents and Multiagent Systems, 2017.

[14] M. L. Littman, "Markov games as a framework for multi-agent reinforcement learning," in *Machine learning proceedings 1994*, pp. 157–163, Elsevier, 1994.

[15] A. Bab and R. I. Brafman, "Multi-agent reinforcement learning in common interest and fixed sum stochastic games: An experimental study," *Journal of Machine Learning Research*, vol. 9, no. Dec, pp. 2635–2675, 2008.

[16] E. M. de Cote, A. Lazaric, and M. Restelli, "Learning to cooperate in multi-agent social dilemmas," in *AAMAS*, vol. 6, pp. 783–785, 2006.

[17] I. Erev and A. E. Roth, "Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria," *American economic review*, pp. 848–881, 1998.

[18] K. Tuyls and G. Weiss, "Multiagent learning: Basics, challenges, and prospects," *Ai Magazine*, vol. 33, no. 3, pp. 41–41, 2012.

[19] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," *arXiv preprint arXiv:1312.5602*, 2013.

[20] R. J. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning," *Machine learning*, vol. 8, no. 3-4, pp. 229–256, 1992.

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

[22] J. Foerster, R. Y. Chen, M. Al-Shedivat, S. Whiteson, P. Abbeel, and I. Mordatch, "Learning with opponent-learning awareness," in *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, pp. 122–130, International Foundation for Autonomous Agents and Multiagent Systems, 2018.

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

[24] J. Hu, M. P. Wellman, *et al.*, "Multiagent reinforcement learning: theoretical framework and an algorithm.," in *ICML*, vol. 98, pp. 242–250, Citeseer, 1998.

[25] J. W. Crandall and M. A. Goodrich, "Learning to compete, compromise, and cooperate in repeated general-sum games," in *Proceedings of the 22nd international conference on Machine learning*, pp. 161–168, 2005.

[26] J. W. Crandall, M. Oudah, F. Ishowo-Oloko, S. Abdallah, J.-F. Bonnefon, M. Cebrian, A. Shariff, M. A. Goodrich, I. Rahwan, *et al.*, "Cooperating with machines," *Nature communications*, vol. 9, no. 1, pp. 1–12, 2018.

[27] W. Zhang, L. Ma, and X. Li, "Multi-agent reinforcement learning based on local communication," *Cluster Computing*, pp. 1–10, 2018.

[28] Y. Shoham, R. Powers, and T. Grenager, "Multi-agent reinforcement learning: a critical survey," *Web manuscript*, 2003.

[29] X. Wang and T. Sandholm, "Reinforcement learning to play an optimal nash equilibrium in team markov games," in *Advances in neural information processing systems*, pp. 1603–1610, 2003.

[30] M. Lanctot, V. Zambaldi, A. Gruslys, A. Lazaridou, K. Tuyls, J. Perolat, D. Silver, and T. Graepel, "A unified game-theoretic approach to multiagent reinforcement learning," in *Advances in Neural Information Processing Systems*, pp. 4190–4203, 2017.

[31] C. Claus and C. Boutilier, "The dynamics of reinforcement learning in cooperative multiagent systems," *AAAI/IAAI*, vol. 1998, pp. 746–752, 1998.

[32] Y. Yang, R. Luo, M. Li, M. Zhou, W. Zhang, and J. Wang, "Mean field multi-agent reinforcement learning," *arXiv preprint arXiv:1802.05438*, 2018.

[33] R. Axelrod, "More effective choice in the prisoner's dilemma," *Journal of conflict resolution*, vol. 24, no. 3, pp. 379–403, 1980.

[34] M. Schaarschmidt, A. Kuhnle, B. Ellis, K. Fricke, F. Gessert, and E. Yoneki, "LIFT: Reinforcement learning in computer systems by learning from demonstrations," *CoRR*, vol. abs/1808.07903, 2018.

[35] A. Kuhnle, M. Schaarschmidt, and K. Fricke, "Tensorforce: a tensorflow library for applied reinforcement learning." Web page, 2017.

[36] J. J. Horton, D. G. Rand, and R. J. Zeckhauser, "The online laboratory: Conducting experiments in a real labor market," *Experimental economics*, vol. 14, no. 3, pp. 399–425, 2011.

[37] G. Paolacci, J. Chandler, and P. G. Ipeirotis, "Running experiments on amazon mechanical turk," *Judgment and Decision making*, vol. 5, no. 5, pp. 411–419, 2010.

[38] P. J. Rosopa, M. M. Schaffer, and A. N. Schroeder, "Managing heteroscedasticity in general linear models.," *Psychological Methods*, vol. 18, no. 3, p. 335, 2013.

[39] A. F. Hayes and L. Cai, "Using heteroskedasticity-consistent standard error estimators in ols regression: An introduction and software implementation," *Behavior research methods*, vol. 39, no. 4, pp. 709–722, 2007.