Better Understanding of Humans for Cooperative AI through Clustering

Edward Su
School of Computer Science
University of Technology Sydney
Sydney, Australia
edward.su@student.uts.edu.au

William Raffe
School of Info Technology
Deakin University
Melbourne, Australia
william.raffe@deakin.edu.au

Luke Mathieson
School of Computer Science
University of Technology Sydney
Sydney, Australia
Luke.Mathieson@uts.edu.au

YuKai Wang
School of Computer Science
University of Technology Sydney
Sydney, Australia
YuKai.Wang@uts.edu.au

2024 IEEE Conference on Games (CoG) | 979-8-3503-5067-8/24/$31.00 ©2024 IEEE | DOI: 10.1109/CoG60054.2024.10645647

***Abstract*—Cooperative AI and AI alignment research are increasingly important fields of study as machine learning models are becoming more prevalent in society. Applications such as self-driving cars, realistic AI in games, and human-AI teams, all require further advancement in cooperative and alignment research before more widespread applications can be achieved. However, research in these fields has typically lagged behind other machine learning applications due to the difficulty of creating models that are robust to and can adapt to novel human partners. We attempt to address this through the creation of a framework that uses Archetypal Analysis, a unique clustering algorithm that finds extremal 'archetype' points in a dataset and expresses each other point as a convex combination of these archetypes. This framework creates understandable archetypes of players which a reinforcement learning agent can use to adapt accordingly to unseen partners. We show that this framework not only results in performance comparable to other cooperative benchmark models but also achieves higher levels of perceived cooperativeness without the need for human involvement during the training process. As such, we demonstrate that the use of clustering techniques to better model different types of human behaviour and strategies can be an effective approach in improving the ability of AI models to adapt to and improve cooperation with novel partners.**

***Index Terms*—Cooperation, Reinforcement learning, Archetypal analysis, Clustering, Multi-agent**

I. Introduction

Reinforcement Learning (RL) has been used to produce models capable of competing with humans in high-skill games such as Chess, Go and Dota 2 [1]–[3]. However, much of this success has been focused on the domain of zero-sum problems, where AI models are designed to perform optimal actions and compete with their opponents. As machine learning tools become increasingly complex and integrated into society, evidence shows that more work needs to be done on improving their capacity to cooperate with humans. [4].

Early attempts at developing AI capable of cooperation took the form of multiagent reinforcement learning (MARL), where multiple RL agents are trained simultaneously with one another [5]–[7]. In this approach, agents attempt to maximize a reward function similar to traditional RL but have the additional property that agents can interact with one another and share knowledge, communicate, and perform joint actions [8]. MARL has found success in numerous areas spanning tasks in games [9], economic interactions [10], and joint-decision making [11], but these have largely been between AI agents. It is often the case that RL models generalize poorly to human partners in cooperative settings [12].

Proposed to address this problem is the use of techniques that aim to replicate human behaviour as closely as possible, such as imitation learning, behaviour cloning, and inverse reinforcement learning, which attempt to replicate human action as closely as possible [12], [13]. Another approach that has found success is known as fictitious co-play, a novel framework that involves having an agent training with a pool of deep reinforcement learning (DRL) models that are checkpointed at different stages during training to better represent a diverse range of skills and playstyles [14]. These models likely found more success due to having exposed the DRL models with partners with more diverse behaviours that make them more robust and generalizable to human behaviours. Making more robust agents is a step towards better cooperative models but we believe that generalisability is not equivalent to better cooperation or alignment with humans. We assume that to improve cooperation and alignment, AI models need to have a better understanding of the different playstyles humans have manifested by their unique personalities and backgrounds.

Games have made use of clustering techniques in the past to better understand their players or to enhance the capabilities of traditional AI systems [15]–[17]. We hypothesize that there may be potential to use similar techniques alongside RL models to enhance cooperation and alignment. One clustering technique that stands out is archetypal analysis (AA), which

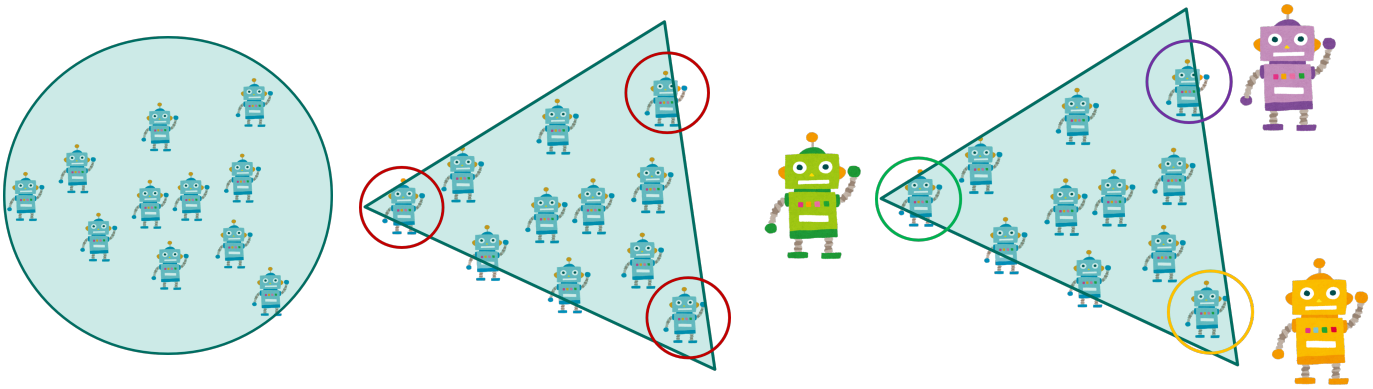

Fig. 1: The framework of our AA agent. In the **left-most image**, we use fictitious co-play to train models and have the models complete episodes in the environment while tracking play data. Illustrated in the **centre image**, we then performed AA on the dataset to generate archetypes representing outlier playstyles. We then train separate PPO models to cooperate with each archetype as shown in the **right-most image**.

uniquely clusters data around outliers known as 'archetypes' as opposed to mean data points typical of other clustering techniques [18]. These archetypes are defined as extremal points in the data that create a boundary that encapsulates all other observations. The benefit of doing so is that the basis vectors are significantly different, which provides more meaningful information when contrasting strategies and makes it simpler to interpret the results achieved [19]. Though computationally inefficient in comparison to more developed clustering algorithms such as DBSCAN and spectral clustering, AA's approach to clustering better aligns with research into how human personality models in the form of the Five-Factor [20] and HEXACO models [21], where personality is determined to be a combination of extremal features.

In this paper, we propose a framework for developing RL models with the ability to cooperate with novel human partners which we refer to as the AA agent. To test the framework, we use the 'Overcooked' environment developed by Carroll et. al. [12] and take inspiration from the fictitious co-play techniques proposed by Strouse et. al. [14]. Our framework works by creating a dataset of game playthroughs representative of human players and extracting 'archetypal' playstyles using AA offline as seen in Fig. 1. For each archetype, an agent is trained to optimally cooperate with them, which we will refer to as a cooperative model. During runtime, a linear least-squares algorithm is used to calculate a human partner's alignment to each archetype, and the relevant cooperative model action is stochastically chosen for the AA agent to perform, weighted by the strength of the alignment. We call the resulting model, the AA agent.

In our experiments, we found that the AA agent performed similarly to self-play and human-trained models regarding the score it achieved in the game. However, the AA agent was perceived as significantly more cooperative than the self-play model and similar to the human-trained model according to participant feedback and ratings. Though achieving similar results to human-trained models, we find that the AA agent demonstrates a lot of potential having done so without the use of human data and demonstrating more adaptive behaviour.

II. Previous Work

For AI models to improve in their ability to cooperate with humans, there is a need for them to be able to understand the intentions of humans such as through mental models [22]–[25]. Much research has been conducted in measuring and discerning human intentions in robotics, however, the techniques have yet to be applied to cooperative AI or RL settings in general [26]–[28]. Existing methods in robotics for collecting information on human intent generally involve the use of wearable technology [29], [30] or sensors [27] which can enable the collection of more data suitable for informing the actions of AI in cooperative contexts.

An approach addressing the construction of mental models of humans for cooperative settings is human-AI Shared Mental Models (SMMs) which proposes that for humans and AI agents to cooperate, they must both have an understanding of their individual and shared goals [22]. Though they have found some success in improving cooperation, the approaches are context-specific and/or context-sensitive often requiring specialized architecture which limits more widespread adoption and testing. These methods remain relatively unexplored with only a few minor experiments in the past decade [23]–[25], largely due to how abstract they are, with minimal concrete benefits. More developed approaches to model and predict human intentions and behaviours include Dynamic Function Allocation and Adaptive Automation [31]–[33]. Though these methods have found success in areas such as aerospace [34] and navigating vehicles [33], they often require bespoke solutions for different contexts, struggle to generalize to different partners and only consider unidirectional adaptation with the AI adapting to the human.

Though applications of human profiles in the context of RL have been minimal, there has been much research into the features that make up personalities, a crucial piece of

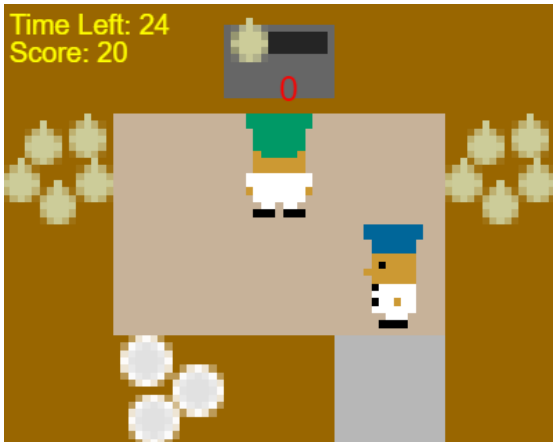

Fig. 2: The 'Cramped Room' layout we built upon, and conducted experiments with, from the 'Overcooked ai' environment developed by Carroll et al. [12].

knowledge when determining what techniques are best for capturing it. The main approaches to model these personalities are the Five-Factor model [20] and the HEXACO model [21], which though differ from one another in the choice of traits, each express human personality as a combination of various extremal traits or features, rather than a single, fixed trait. The result is that when the model is applied to individuals, they will have a unique combination of different degrees of these traits which when combined, define their characteristics such as how they approach learning and process information [35].

A promising approach to developing better mental models of humans, reflective of the aforementioned research of personality is through the use of clustering techniques such as Archetypal Analysis which has already seen applications in contexts outside of cooperative AI [18], [19], [36] and allows human partners to be represented as combinations of archetypal behaviours which can be effective in informing AI models on how to best adapt to them.

III. Implementation

A. Environment

An environment that has seen increasing use for developing cooperative AI models is 'Overcooked', a multi-agent environment that challenges multiple agents in their ability to cooperate with one another [12], [14]. In this game, players act as chefs responsible for delivering as many dishes to customers within a given time frame. In order to do so, they must coordinate with one another in the kitchen by delegating tasks, sharing ingredients and avoiding running into one other.

Overcooked serves as a great testbed for our purposes as it has a high level of complexity, with multiple tasks to be performed simultaneously and others that may force coordination between agents. This complexity enables the emergence of more diverse play patterns, which in turn leads to more meaningful archetypes when applying archetypal analysis. In addition, the multi-agent nature of the environment is ideal for assessing what archetypes may perform better with others, which can be invaluable when attempting to coordinate with an unseen player. This is supported by the sparse nature of rewards that encourages agents to work together to achieve long-term goals as opposed to acquiring fast rewards they could get themselves.

We used the environment implementation developed by Carroll et. al. [12], as seen in Fig. 2, with a few adjustments in order to facilitate our custom AA agent. This included a few data structures to hold information of each episode as well as information on the archetypal profile of each agent.

B. Archetypal Analysis

Our approach in integrating AA to develop cooperative AI models is by using it as a heuristic in an ensemble framework, where it can be used to determine the most appropriate model action to use. In order to achieve this, we first created a dataset of playthroughs by representative human players to run AA on and find archetypal playstyles. Once this was developed, we would then train an RL model with each archetypal playstyle so that they were optimised to cooperate with them. This would leave us with N number of RL models where N is the number of archetypal playstyles which we would shift between to select actions during runtime using ensemble learning.

To produce the initial dataset of playstyles, we chose to create RL models that were representative of human playstyles as opposed to recording real human playthroughs. This was as it would be expensive and time-consuming to have a human play through a large number of levels to create the dataset. To create the RL models, we took inspiration from Strouse et al. [14] and their use of fictitious co-play models and take a similar approach in creating this initial dataset to avoid the use of human data. This involved training 5 self-play agents using a PPO policy that are checkpointed during the process to represent levels of player skill. These models then play with one another and the data from each playthrough is saved. Through observations collected during recordings of multiple playthroughs of the overcooked environment by human players, we found that the most relevant features conducive to determining the archetype of a player were:

- Number of objects placed
- Number of objects boiled
- Number of soup delivered
- Number of soup plated

Each of the 5 PPO agents were initiated with a random seed and trained for 10,000 timesteps and checkpointed after timesteps 2500, 5000 and 7500, producing 20 different models. These models would then play a random number of playthroughs with themselves in an overcooked level ranging from 100-150 times each. In the end, we produced 2647 playthroughs which served as our dataset representing different approaches to the environment.

We then perform archetypal analysis on this dataset, which provides us with K number of archetypes, where K is an arbitrary integer we choose. To aid us in choosing an effective number of archetypes, we calculated the explained variance of different numbers of archetypes, which can be seen in Fig. 3.

TABLE I: Archetype Profiles

Archetype	Features			
	objects placed	*objects boiled*	*soup delivered*	*soup placed*
A1	0.000000	0.770684	0.582924	1.000000
A2	0.000000	0.846997	0.834530	0.000000
A3	0.954288	0.000000	0.000000	0.295429

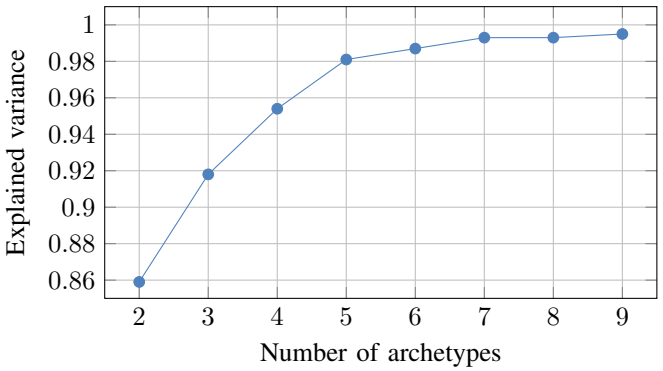

Fig. 3: Comparison of different number of archetypes

Evidently, benefits of additional archetypes tapers off after 5 archetypes with 3-5 archetypes explaining most of the variance in the dataset. Using fewer archetypes would fail to encode all the information from the dataset, while using more archetypes would take away the benefits of the dimensionality reduction effects afforded by the algorithm. With the original feature set having a dimensionality of 4, we chose to proceed with 3 archetypes to take advantage of the dimensionality reduction benefits of archetypal analysis. The profiles of the archetypes generated with K = 3 archetypes are seen in Table I and the data-points in the feature-set can be expressed as a convex combination of these 3 archetypes shown in Fig. 4.

Afterward, we trained a RL model with a PPO policy to cooperate with the model that had the closest alignment to each archetype for 10,000 timesteps. The result of this was 3 policies that were each tailored to cooperate with one of the 3 player archetypes produced previously. During runtime, the AA agent observes the players alignment to one of the aforementioned archetypes through a least-squares algorithm comparing the players current play data scaled by time. Using the resulting vector representing alignment, the agent stochastically selects the appropriate cooperative model action weighted by the players proximity to the given archetypes. For example, based on a players actions, performing a least-squares operation may result in a distribution of [0.001, 0.87, 0.129] representing the convex combination of 3 archetypes. The model will then select the 1st archetype with 0.001 probability, the 2nd archetype with 0.87 probability and the 3rd archetype with 0.129 probability.

C. Benchmarks

To assess the capabilities of our custom cooperative agent framework, we compare it with other models that have been used for cooperation. A variety of agents were used in the experiment including 2 benchmark models from Carroll et al. [12]:

- Self-play agent: This benchmark agent was trained from scratch with itself using Proximal Policy Optimization (PPO). It had access to information about its state, such as what object it was holding, as well as the state of the environment, such as the number of objects in the pot. To expedite training, it was given rewards for positive intermediate actions, including putting onions into the pot and picking up soup with a dish.
- Human-aware PPO agent: This benchmark agent was developed by first training a model to act as similarly to a human as possible using behaviour cloning, a technique where the model learns a policy from demonstrations. The agent then trains with this model using PPO and implements a model-based planner that uses a hierarchical A* search to act and strategize optimally in response to the policy of their partner.
- Random action agent: This agent selects an action to perform at complete random. It is not expected that this agent will perform well and it largely acts as a point of comparison for participants to evaluate the cooperativity of a partner.
- AA agent: This ensemble agent was trained using our custom framework described previously under the Implementation section. It makes use of archetypal analysis to select the appropriate model action in response to their partner's perceived playstyle.

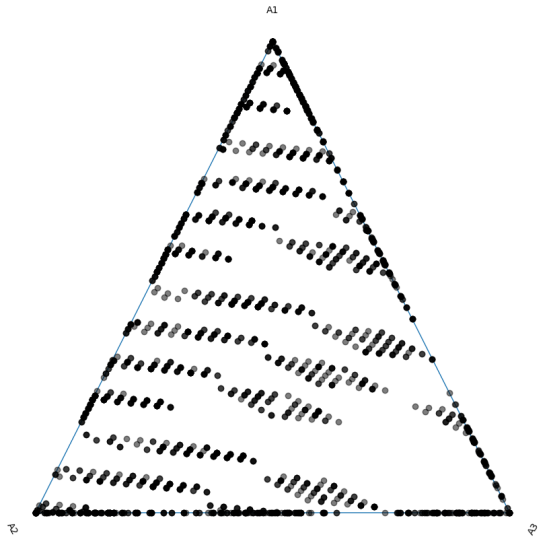

Fig. 4: Datapoints expressed as a convex combination of 3 archetypes.

IV. Experiment

A. Procedure

The experiment was held in person at a university computer lab. Upon entry, participants were greeted and informed of

the experiment's procedure. They would then be required to complete a written consent form and be given an initial safety brief. They were then directed to a computer with instructions for how to play the game while a host would answer any questions they had. To ensure that players understand the directions, they would play a test game that was not recorded with a random AI agent. They were allowed to play as many test games as necessary until they felt confident in playing the game.

Once ready, the host will randomly pair the participant with an AI agent among the list of agents listed above, who they would then play a game with. Each game lasted for 30 seconds, after which they were prompted to complete the relevant sections of the questionnaire and the experiment host would note down the score that was achieved and any observations or opinions they had.

Upon completion, the host would randomly change the AI agent to one that had not already been chosen and then prompt the participant to begin again when they were prepared. This procedure would continue until the participant had been partnered with each AI agent, at which point they would be debriefed and the experiment was concluded.

B. Participants

Participants were recruited through announcements distributed across university channels, including club communications, and on local game industry forums and groups. These recruitment messages were uniformly disseminated across public forums to maintain a non-personalized approach, thereby minimizing potential biases and ensuring voluntary participation. Our goal was to acquire approximately 20 participants, however recruitment concluded with 16 participants.

The participants we acquired fell under the 20-30 year old range with 75% of participants between 20-25 and the remaining 25% between 25-30. Participants had a diverse range of experiences in playing cooperative games with 12.5% rating themselves as beginners to cooperative games, 56.25% as having intermediate experience, and 31.25% as having advanced experience. As a result, we saw a good variety of perspectives on cooperative behavior and several different strategies to succeed in the game.

C. Metrics

When comparing the performance of Cooperative RL models, researchers have employed various quantitative metrics to assess their effectiveness such as cumulative reward and average rewards per episode [12], [14]. Though these approaches see common usage and can be useful to compare the performance between RL agents, they overlook factors unique to cooperative contexts such as perceived cooperation. For instance, in human-robot interaction scenarios, the perception of cooperation from users can significantly influence their satisfaction and engagement with the system, regardless of individual agent performance metrics [37], [38]. Therefore, incorporating measures of perceived cooperation, such as user feedback or subjective evaluations, alongside traditional performance metrics can provide a more comprehensive understanding of the effectiveness of RL models in cooperative settings.

As such, the data we collected during the experiment included observational notes on player behavior and trends, scores they achieved with each AI, and a questionnaire that the participants filled out. The questionnaire was filled by players after each playthrough with questions varying from those assessing their intrinsic motivation as applied in previous experiments evaluating the experiences of playing games [39], to open questions that assess the levels of confidence and trust that players have of their team partners [40]. The questionnaire also asks participants to rate the cooperativity of each agent they partner with using a 5-point Likert scale, giving us a qualitative measure in which to compare the perceived level of cooperation for each agent. The questions for each agent included:

- How cooperative did you feel your partner was on a scale of 1-5?
- What factors contributed to you reaching the above conclusion

General questions that were asked include:

- How experienced are you with playing cooperative video games?
- What differences stood out between playing with a human and an AI?
- Any final comments you would like to add?

V. Results and Discussion

In this section, we provide a summary of the results we achieved during the experiment. This includes quantitative comparisons between the scores each AI agent achieved as well as the cooperative rating that participants gave them. Qualitative data such as observations of participants and the responses they provided in the questionnaire will also be summarised.

A. Quantitative Results

The following section covers the quantitative results that were collected through participant playthroughs and the rating participants gave to agents in the questionnaire. We conducted Bayesian, non-parametric ANOVA tests on the two quantitative measures of agent performance, that being the scores they were able to achieve with human partners as well as the cooperativity rating participants gave each agent in the questionnaire. Though we also experimented with a standard non-parametric ANOVA test, we found the Bayesian equivalent had similar results but provided additional information as a result of the Bayes factor. We then performed paired sample tests using a Bayesian Wilcoxon signed-rank test to gather more specific data on comparisons between models. We first begin with the tests on the scores that each model was able to achieve and then continue to discuss the equivalent for the cooperative ratings each model received.

From the non-parametric ANOVA test shown in Table II, the BF (Bayes Factor) when comparing AA to Human-Trained,

TABLE II: Post Hoc Comparisons - Agent Type

		$BF_{10,U}$	error %
AA	SelfPlay	0.294	0.013
	Random	35.997	7.802×10^{-7}
	Human_Trained	1.545	3.122×10^{-6}
SelfPlay	Random	8673.877	2.548×10^{-7}
	Human_Trained	1.387	3.761×10^{-6}
Random	Human_Trained	12344.200	1.359×10^{-7}

TABLE III: Descriptives

Agent Type	Mean	SD	SE	CoV
AA	47.500	17.701	4.425	0.373
SelfPlay	50.000	12.649	3.162	0.253
Random	28.750	10.247	2.562	0.356
Human_Trained	56.250	15.000	3.750	0.267

TABLE IV: Bayesian Wilcoxon Signed-Rank Test

Measure 1		Measure 2	BF_{10}	W	Rhat
AA	-	SelfPlay	0.527	8.000	1.000
	-	Random	37.910	55.000	1.002
	-	Human_Trained	1.898	10.000	1.000
SelfPlay	-	Random	374.649	105.000	1.018
	-	Human_Trained	1.421	4.000	1.000
Random	-	Human_Trained	377.061	0.000	1.012

TABLE V: Post Hoc Comparisons - Agent Type Rating

		$BF_{10,U}$	error %
AA	SelfPlay	0.414	8.686×10^{-7}
	Random	0.414	3.079×10^{-7}
	Human_Trained	0.414	0.019
SelfPlay	Random	0.414	7.208×10^{-7}
	Human_Trained	0.414	2.228×10^{-7}
Random	Human_Trained	0.414	2.155×10^{-7}

TABLE VI: Descriptives Rating

Agent Type	Mean	SD	SE	CoV
AA	3.750	0.856	0.214	0.228
SelfPlay	2.750	0.683	0.171	0.248
Random	1.875	0.719	0.180	0.383
Human_Trained	4.063	0.772	0.193	0.190

TABLE VII: Bayesian Wilcoxon Signed-Rank Test Rating

Measure 1		Measure 2	BF_{10}	W	Rhat
AA	-	SelfPlay	7.729	90.000	1.004
	-	Random	1450.546	120.000	1.007
	-	Human_Trained	0.521	20.000	1.000
SelfPlay	-	Random	16.759	62.000	1.006
	-	Human_Trained	60.059	0.000	1.002
Random	-	Human_Trained	173.140	0.000	1.021

AA to Self-play, and Self-play to Human-Trained agents were all equivocal, with a BF10 < 3 representing insufficient evidence to prove that they are not equal. This suggests that there are no significant differences in their performance given the sample size. Following the same measure for significance, we found that there was sufficient evidence that random-action agents were significantly different from the other 3 agents.

Looking at the distribution of results in Figure III, we found there was substantial variability in the score of the AA agent, which had a standard deviation of 17.7, as well as the human-trained agent with a standard deviation of 15. This is in comparison to the random and self-play agents which had noticeably lower standard deviations. We believe this was a result of poorer performance with partners with more even archetypal mixtures which we expand upon in section V-B.

We then conducted a Bayesian Paired Samples T-Test as seen in Table IV to find additional details between the agents, which reinforced the findings we previously found with the ANOVA test. We did not include the comparisons between self-play and human-trained models with random agents as they were not the focus of the experiment.

From the non-parametric ANOVA test shown in Table V, we found that when it came to ratings of cooperativity, the AA and human-trained models were equivalent with a BF10 < 3. Using the same measure, the self-play model performed significantly worse than AA and human-trained models but better than the random action model which performed overwhelmingly poorly in comparison to the other models.

Analysing the details of the cooperative rating results in Table VI, we found that similarly to the score results, the AA agent had the highest variance in their ratings, though they were closer in line with the variance of the other agents this time.

We then similarly conducted a Bayesian Paired Samples T-Test as seen in Table VII to find additional details in the pair-wise comparisons of the cooperative ratings received by different agents.

In summary, the AA agent performed similarly to self-play and human-trained models regarding the score they were able to achieve during the experiment. Additionally, the AA agent achieved similar results to the human-trained model in cooperative ratings it received but was significantly better than self-play and random action models. In both score and cooperative rating measures, the AA agent had the greatest variance in its results.

B. Qualitative Results

The AA agent received mostly positive feedback from participants, with many comments recognizing that the agent made distinct attempts to cooperate with its partner's strategy. A few comments also mentioned that the AA agent would change its strategy multiple times throughout a playthrough to best match the strategy of the participant. The general theme of negative feedback regarding the AA agent was that it would sometimes suffer from periods of indecision. This often took the form of performing an action conducive to a strategy and then proceeding to perform an action towards another strategy which was frequently counter-productive. Through

observations, we found that this generally occurred to players whose archetypal mixture was close to an even distribution of each archetype, as due to the stochastic ensemble approach to action selection, the likelihood of multiple actions being conducive to a single strategy is low.

The Human-trained agent received similarly positive feedback in general, with a majority of comments by participants praising its ability to adapt to their strategy. The main complaints by participants with the Human-trained agent were that it would often block them from performing actions during playthroughs. Through observation, this would often occur when the player and the AI agent need to pass each other to get to a given object, with the Human-trained agent frequently taking on a more assertive personality, not backing down on its current trajectory and forcing the player to change paths should they wish to avoid a stalemate where both of them remain stationary.

The Self-play agent generally received negative feedback from participants regarding its ability to cooperate with them as partners. One major theme in the responses was that the Self-play agents tried to do everything themselves, completely disregarding the player, forcing them to try and adapt. Through observation, they act quite stubbornly and scripted, which leads to lower cooperativity ratings but still perform quite well in regards to the score they achieve as it forces their partner to adapt to them. This would frustrate many of the participants who would attempt to assist the AI initially but lose motivation once they realized that they were being ignored.

The Random-action agent expectedly received poor feedback from participants due to not performing many actions conducive to the success of the team. The main points of feedback were that the Random-action agent did not know what to do and just simply walked around cluelessly. A few responses instead mention that the AI took a passive approach to cooperation, expecting the participant to initiate. Through observation, we found that the Random-action agent would sometimes have high cooperative ratings due to its tendency to avoid collisions with their partner.

To summarise overarching qualitative trends in the questionnaire responses and the observations made during the experiment by experiment hosts:

1) AI obstruction of player actions had a significant effect on the cooperativity ratings an agent was given and would greatly frustrate participants. This would often overshadow otherwise great performance and good decision-making processes by the AI agent as there were situations where the human-trained agent or AA models that performed well in score and cooperation received poorer ratings in cooperativity if they obstructed player movement or actions.
2) Participants with more experience playing cooperative games would pay greater attention to the actions of their AI partner, while in comparison, those with less experience would largely focus on their actions. This would result in occasionally higher than-expected opinions by less experienced players for poor-performing agents such as the Random-action agent as well as agents that appear competent such as the Self-play agent.
3) The AA and Human-trained agents were able to significantly better adapt to their partner's strategies compared to the Random-action and Self-play agents. The Human-trained agent was typically more robust and smooth in its adaptations to the player while the AA agent made more distinct shifts in strategy in comparison at the cost of more instability.

VI. Conclusion

Our use of archetypal analysis as a heuristic for ensemble frameworks to better adapt to human partners is flexible, generalizable, and can easily be included in most existing model designs. Throughout the development, we found that there were a variety of future directions that could be taken to improve the model including smoothing the transition between strategies and the implementation of techniques to identify important features to observe in the environment rather than simply choosing what we believed to be the most relevant. Furthermore, there is potential for there to be more meaningful adaptations to non-archetypal partners by scaling the weights of the model directly rather than weighting stochastic action selection.

Overall, we proposed a novel method that is simple to include in existing DRL approaches and that has demonstrated promising results in making AI agents more cooperative with human partners. This was achieved by taking advantage of a clustering algorithm called archetypal analysis to better understand human partners and adapt to their actions. In doing so, we have established a promising direction for future research on improving the cooperative capabilities of deep reinforcement learning models.

References

[1] T. Cazenave, "Residual Networks for Computer Go," *IEEE Transactions on Games*, vol. 10, no. 1, pp. 107–110, Mar. 2018, conference Name: IEEE Transactions on Games.

[2] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel, and D. Hassabis, "Mastering the game of Go without human knowledge," *Nature*, vol. 550, no. 7676, pp. 354–359, Oct. 2017, number: 7676 Publisher: Nature Publishing Group. [Online]. Available: https://www.nature.com/articles/nature24270

[3] OpenAI, C. Berner, G. Brockman, B. Chan, V. Cheung, P. Dębiak, C. Dennison, D. Farhi, Q. Fischer, S. Hashme, C. Hesse, R. Józefowicz, S. Gray, C. Olsson, J. Pachocki, M. Petrov, H. P. d. O. Pinto, J. Raiman, T. Salimans, J. Schlatter, J. Schneider, S. Sidor, I. Sutskever, J. Tang, F. Wolski, and S. Zhang, "Dota 2 with Large Scale Deep Reinforcement Learning," *arXiv:1912.06680 [cs, stat]*, Dec. 2019, arXiv: 1912.06680. [Online]. Available: http://arxiv.org/abs/1912.06680

[4] A. Dafoe, E. Hughes, Y. Bachrach, T. Collins, K. R. McKee, J. Z. Leibo, K. Larson, and T. Graepel, "Open problems in cooperative AI," *arXiv*, no. August 2019, 2020, _eprint: 2012.08630.

[5] C. Neary, Z. Xu, B. Wu, and U. Topcu, "Reward Machines for Cooperative Multi-Agent Reinforcement Learning," 2020, _eprint: 2007.01962. [Online]. Available: http://arxiv.org/abs/2007.01962

[6] H. Zhang and S. Zhang, "Multi-agent reinforcement learning," *Deep Reinforcement Learning: Fundamentals, Research and Applications*, vol. 73, pp. 335–346, 2020, iSBN: 9789811540950.

[7] N. Jaques, A. Lazaridou, E. Hughes, C. Gulcehre, P. A. Ortega, D. J. Strouse, J. Z. Leibo, and N. de Freitas, "Social influence as intrinsic motivation for multi-agent deep reinforcement learning," *36th International Conference on Machine Learning, ICML 2019*, vol. 2019-June, pp. 5372–5381, 2019, iSBN: 9781510886988 _eprint: 1810.08647.

[8] M. Sierhuis, J. M. Bradshaw, A. Acquisti, R. van Hoof, and R. Jeffers, "Human-Agent Teamwork and Adjustable Autonomy in Practice," p. 8.

[9] B. Baker, I. Kanitscheider, T. Markov, Y. Wu, G. Powell, B. McGrew, and I. Mordatch, "Emergent Tool Use From Multi-Agent Autocurricula," *arXiv:1909.07528 [cs, stat]*, Feb. 2020, arXiv: 1909.07528. [Online]. Available: http://arxiv.org/abs/1909.07528

[10] S. Parsons and M. Wooldridge, "Game Theory and Decision Theory in Multi-Agent Systems," *Autonomous Agents and Multi-Agent Systems*, vol. 5, no. 3, pp. 243–254, Sep. 2002. [Online]. Available: https://doi.org/10.1023/A:1015575522401

[11] Y. Chevaleyre, U. Endriss, J. Lang, and N. Maudet, "A Short Introduction to Computational Social Choice," in *SOFSEM 2007: Theory and Practice of Computer Science*, ser. Lecture Notes in Computer Science, J. van Leeuwen, G. F. Italiano, W. van der Hoek, C. Meinel, H. Sack, and F. Plášil, Eds. Berlin, Heidelberg: Springer, 2007, pp. 51–69.

[12] M. Carroll, R. Shah, M. K. Ho, T. L. Griffiths, S. A. Seshia, P. Abbeel, and A. Dragan, "On the utility of learning about humans for human-AI coordination," *arXiv*, no. NeurIPS, 2019, _eprint: 1910.05789.

[13] S. Feng and J. Boyd-Graber, "What can AI do for me?" pp. 229–239, 2019, iSBN: 9781450362726.

[14] D. J. Strouse, K. R. McKee, M. Botvinick, E. Hughes, and R. Everett, "Collaborating with Humans without Human Data," *arXiv:2110.08176 [cs]*, Oct. 2021, arXiv: 2110.08176. [Online]. Available: http://arxiv.org/abs/2110.08176

[15] C. Bauckhage, A. Drachen, and R. Sifa, "Clustering Game Behavior Data," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 7, no. 3, pp. 266–278, Sep. 2015. [Online]. Available: https://ieeexplore.ieee.org/document/6975073/

[16] M. Swiechowski and D. Slezak, "Grail: A Framework for Adaptive and Believable AI in Video Games," *Proceedings - 2018 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2018*, pp. 762–765, 2019, iSBN: 9781538673256 Publisher: IEEE.

[17] C. Rosenthal and C. B. Congdon, "Personality profiles for generating believable bot behaviors," *2012 IEEE Conference on Computational Intelligence and Games, CIG 2012*, pp. 124–131, 2012, iSBN: 9781467311922 Publisher: IEEE.

[18] A. Cutler and L. Breiman, "Archetypal analysis," *Technometrics*, vol. 36, no. 4, pp. 338–347, 1994.

[19] C. Thurau and A. Drachen, "Introducing Archetypal Analysis for Player Classification in Games Categories and Subject Descriptors," *Proceedings of the EPEX 11 Workshop*, 2011, iSBN: 9781450308045.

[20] R. M. Bagby and T. A. Widiger, "Five Factor Model personality disorder scales: An introduction to a special section on assessment of maladaptive variants of the five factor model," *Psychological Assessment*, vol. 30, no. 1, pp. 1–9, Jan. 2018.

[21] M. C. Ashton and K. Lee, "The HEXACO Model of Personality Structure and the Importance of the H Factor," *Social and Personality Psychology Compass*, vol. 2, no. 5, pp. 1952–1962, 2008, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1751-9004.2008.00134.x. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1751-9004.2008.00134.x

[22] R. W. Andrews, J. M. Lilly, D. Srivastava, and K. M. Feigh, "The role of shared mental models in human-AI teams: a theoretical review," *Theoretical Issues in Ergonomics Science*, vol. 24, no. 2, pp. 129–175, Mar. 2023, publisher: Taylor & Francis _eprint: https://doi.org/10.1080/1463922X.2022.2061080. [Online]. Available: https://doi.org/10.1080/1463922X.2022.2061080

[23] F. Gervits, A. Roque, G. Briggs, M. Scheutz, and M. Marge, "How Should Agents Ask Questions For Situated Learning? An Annotated Dialogue Corpus," in *Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue*. Singapore and Online: Association for Computational Linguistics, 2021, pp. 353–359. [Online]. Available: https://aclanthology.org/2021.sigdial-1.37

[24] N. Hanna and D. Richards, "The Impact of Multimodal Communication on a Shared Mental Model, Trust, and Commitment in Human–Intelligent Virtual Agent Teams," *Multimodal Technologies and Interaction*, vol. 2, no. 3, p. 48, Aug. 2018. [Online]. Available: http://www.mdpi.com/2414-4088/2/3/48

[25] Y. Zhang, "Role-based shared mental models," in *2008 International Symposium on Collaborative Technologies and Systems*, May 2008, pp. 424–431. [Online]. Available: https://ieeexplore.ieee.org/document/4543960

[26] C. L. R. McGhan, A. Nasir, and E. M. Atkins, "Human Intent Prediction Using Markov Decision Processes," *Journal of Aerospace Information Systems*, vol. 12, no. 5, pp. 393–397, 2015, publisher: American Institute of Aeronautics and Astronautics _eprint: https://doi.org/10.2514/1.I010090. [Online]. Available: https://doi.org/10.2514/1.I010090

[27] C. Nehaniv, K. Dautenhahn, J. Kubacki, M. Haegele, C. Parlitz, and R. Alami, "A methodological approach relating the classification of gesture to identification of human intent in the context of human-robot interaction," in *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.*, Aug. 2005, pp. 371–377, iSSN: 1944-9437.

[28] S. Holtzen, Y. Zhao, T. Gao, J. B. Tenenbaum, and S.-C. Zhu, "Inferring human intent from video by sampling hierarchical plans," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct. 2016, pp. 1489–1496, iSSN: 2153-0866.

[29] D. P. Losey, C. G. McDonald, E. Battaglia, and M. K. O'Malley, "A Review of Intent Detection, Arbitration, and Communication Aspects of Shared Control for Physical Human–Robot Interaction," *Applied Mechanics Reviews*, vol. 70, no. 1, Feb. 2018. [Online]. Available: https://doi.org/10.1115/1.4039145

[30] M. S. Erden and T. Tomiyama, "Human-Intent Detection and Physically Interactive Control of a Robot Without Force Sensors," *IEEE Transactions on Robotics*, vol. 26, no. 2, pp. 370–382, Apr. 2010, conference Name: IEEE Transactions on Robotics.

[31] A. R. Pritchett, S. Y. Kim, and K. M. Feigh, "Measuring Human-Automation Function Allocation," *Journal of Cognitive Engineering and Decision Making*, vol. 8, no. 1, pp. 52–77, Mar. 2014. [Online]. Available: http://journals.sagepub.com/doi/10.1177/1555343413490166

[32] N. R. Bailey, M. W. Scerbo, F. G. Freeman, P. J. Mikulka, and L. A. Scott, "A Brain-Based Adaptive Automation System and Situation Awareness: The Role of Complacency Potential," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 47, no. 9, pp. 1048–1052, Oct. 2003, publisher: SAGE Publications Inc. [Online]. Available: https://doi.org/10.1177/154193120304700901

[33] T. Lavie and J. Meyer, "Benefits and costs of adaptive user interfaces," *International Journal of Human-Computer Studies*, vol. 68, no. 8, pp. 508–524, Aug. 2010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1071581910000145

[34] T. Brunélis, P. Le Blaye, P. Bonnet, and N. Maille, "Distant mission management and dynamic allocation of functions," *Human Operating Unmanned Systems (HUMOUS)*, Sep. 2008.

[35] T. Bidjerano and D. Y. Dai, "The relationship between the big-five model of personality and self-regulated learning strategies," *Learning and Individual Differences*, vol. 17, no. 1, pp. 69–81, Jan. 2007. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S104160800700026X

[36] M. J. A. Eugster, "Performance Profiles based on Archetypal Athletes," *International Journal of Performance Analysis in Sport*, vol. 12, no. 1, pp. 166–187, Apr. 2012. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/24748668.2012.11868592

[37] A.-S. Ulfert, E. Georganta, C. Centeio Jorge, S. Mehrotra, and M. Tielman, "Shaping a multidisciplinary understanding of team trust in human-AI teams: a theoretical framework," *European Journal of Work and Organizational Psychology*, vol. 0, no. 0, pp. 1–14, 2023, publisher: Routledge _eprint: https://doi.org/10.1080/1359432X.2023.2200172. [Online]. Available: https://doi.org/10.1080/1359432X.2023.2200172

[38] K. I. Gero, Z. Ashktorab, C. Dugan, Q. Pan, J. Johnson, W. Geyer, M. Ruiz, S. Miller, D. R. Millen, M. Campbell, S. Kumaravel, and W. Zhang, "Mental Models of AI Agents in a Cooperative Game Setting," *Conference on Human Factors in Computing Systems - Proceedings*, pp. 1–12, 2020, iSBN: 9781450367080.

[39] R. M. Ryan, C. S. Rigby, and A. Przybylski, "The Motivational Pull of Video Games: A Self-Determination Theory Approach," *Motivation and Emotion*, vol. 30, no. 4, pp. 344–360, Dec. 2006. [Online]. Available: https://doi.org/10.1007/s11031-006-9051-8

[40] Z. Ashktorab, Q. V. Liao, C. Dugan, J. Johnson, Q. Pan, W. Zhang, S. Kumaravel, and M. Campbell, "Human-AI Collaboration in a Cooperative Game Setting," *Proceedings of the ACM on Human-Computer Interaction*, vol. 4, no. CSCW2, pp. 1–20, 2020.