

Article

# Multi-Agent System for Emulating Personality Traits Using Deep Reinforcement Learning <sup>†</sup>

Georgios Liapis <sup>‡</sup>  and Ioannis Vlahavas <sup>\*,‡</sup> 

School of Informatics, Aristotle University of Thessaloniki, 541 24 Thessaloniki, Greece; gliapisa@csd.auth.gr

\* Correspondence: vlahavas@csd.auth.gr; Tel.: +30-694-447-0170

<sup>†</sup> This paper is a revised and expanded version of a paper entitled: Liapis, G.; Vordou, A.V.I. Machine Learning Methods for Emulating Personality Traits in a Gamified Environment. In Proceedings of the 13th Conference on Artificial Intelligence (SETN 2024), Piraeus, Greece, 11–13 September 2024.<sup>‡</sup> These authors contributed equally to this work.

**Abstract:** Conventional personality assessment methods depend on subjective input, while game-based AI predictive methods offer a dynamic and objective framework. However, training these models requires large and labeled datasets, which are challenging to obtain from real players with diverse personality traits. In this paper, we propose a multi-agent system using Deep Reinforcement Learning in a game environment to generate the necessary labeled data. Each agent is trained with custom reward functions based on the HiDAC system that encourages trait-aligned behaviors to emulate specific personality traits based on the OCEAN personality trait model. The Multi-Agent Posthumous Credit Assignment (MA-POCA) algorithm facilitates continuous learning, allowing agents to emulate behaviors through self-play. The resulting gameplay data provide diverse, high-quality samples. This approach allows for robust individual and team assessments, as agent interactions reveal the impact of personality traits on team dynamics and performance. Ultimately, this methodology provides a scalable, unbiased methodology for human personality evaluation in various settings, establishing new standards for data-driven assessment methods.

**Keywords:** machine learning; serious games; personality assessment; multi-agent



**Citation:** Liapis, G.; Vlahavas, I.; Multi-Agent System for Emulating Personality Traits Using Deep Reinforcement Learning. *Appl. Sci.* **2024**, *14*, 12068. <https://doi.org/10.3390/app142412068>

Academic Editors: Rakib Abdur and Mehmet Aydin

Received: 15 November 2024

Revised: 19 December 2024

Accepted: 22 December 2024

Published: 23 December 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Personality, which can be defined under characteristic patterns of thought, emotion, and behavior that remain consistent over time and situations [1], plays a key role in shaping behavior, thoughts, and emotions, making it valuable to understand in personal and professional contexts for insights into actions, decisions, and growth. Self-report questionnaires and psychological assessments are common tools for evaluating personality, but they come with limitations, as self-perception biases may affect accuracy. These tools are best complemented by additional feedback or professional evaluations, which can be enhanced by technology [2].

Advancements, especially in gaming, have expanded into fields like education, assessment, and diagnosis. This has led to “serious games”, which use gaming for practical purposes like skill-building and evaluation. Gaming environments promote exploration, problem-solving, and skill development, enhancing cognitive and soft skills like critical thinking and adaptability, which are valuable in today’s workplace [3].

Escape Room (ER) games, a genre that emphasizes teamwork and communication, have emerged as tools in corporate settings for evaluating and fostering these skills. ER games combine physical and mental challenges in dynamic settings, allowing personality traits to manifest naturally and offering a richer understanding than static assessments.

However, traditional evaluations of team performance in ER games rely on post-game questionnaires, which can be biased and fail to capture all player actions [4]. A digital

Escape Room game could overcome these limitations, providing a more comprehensive, real-time approach to tracking and evaluating individual and team dynamics [5].

So, emulating human behavior in a gaming environment is a complex undertaking due to the complexity of personality, which makes it challenging to quantify without a large number of distinct human players with specific traits.

To develop a new personality assessment method, we can combine a digital, serious Escape Room (ER) game with AI technology. This setup allows the AI to analyze gameplay data, identifying patterns that improve assessment accuracy. The ER game triggers human player behaviors, and an AI-based regression system analyzes the data collected to assess their personality traits following the OCEAN Five model [6], which was chosen due to being the most scientifically robust and widely used model in academic psychology and research. However, training this Multi-Output Regression System (MORS) requires extensive labeled data, which is challenging to obtain from human participants with specific trait assessments.

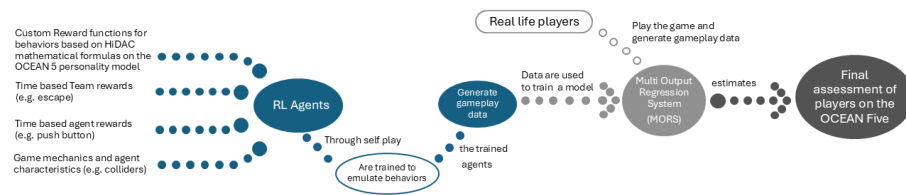
To overcome the challenge of gathering large amounts of labeled personality data, we developed a multi-agent system that can simulate gameplay and generate the necessary data through self-play. Each agent is trained on a specific personality trait, based on custom rewards functions that are based on mathematical formulas from the HiDAC model [7] that represents behaviors. The agents are rewarded not only for solving challenges within the environment but also for displaying behaviors that align with their assigned traits while cooperating.

Related simulations and their results can serve as a base and a benchmark for modeling behaviors in other environments. In this study, we used the HiDAC crowd simulation as our baseline, which is frequently utilized as a benchmark for evaluating models of human behavior.

To train the agents effectively, we used the MA-POCA (Multi-Agent Posthumous Credit Assignment) [8] reinforcement learning algorithm. This approach enables agents to continue learning even after they complete the environment—in this case, when they “escape” the game. This simulation-based training with MA-POCA provides an efficient way to gather high-quality data, refining the MORS’s ability to recognize subtle human behavioral patterns and enhancing the overall accuracy of personality assessments within the Escape Room game environment. By assigning agents a range of personality traits, from high to low, and rewarding trait-aligned behaviors, the system generates a diverse set of gameplay data that improve the MORS’s ability to assess personality accurately.

The data generated through this process are then used to train the MORS utilizing supervised algorithms. Consequently, when human players interact with the game, their gameplay data are gathered and fed to the MORS so that it can accurately assess their personality based on the behaviors exhibited during the game, as shown in Figure 1. Though in this paper, we will only delve into the multi-agent implementation, where we showcase how the final system works.

The emulation of behaviors not only for a single agent [9] or team-related traits [10] but also for all possible traits and behaviors is crucial for generating valuable and high-quality data. Moreover, our approach allows us to evaluate the effectiveness of a team of agents based on each agent’s personality attributes and gameplay style. This holistic view enables a deeper understanding of how different personality traits interact and contribute to overall team dynamics and performance.



**Figure 1.** The workflow of the MindEscape environment (the colored entities are the ones analyzed in this paper).

This paper presents a multi-agent system in a 3D digital ER environment as a simulation of individual and team effectiveness based on the behaviors and personalities of each team member. We propose a specific methodology for the reward functions so that each agent acts and emulates behaviors based on the specific OCEAN Five personality traits model. This methodology can also be used in other types of games and scenarios.

The proposed contributions are as follows:

- A reward methodology for emulating human behaviors in a dynamic gamified environment;
- A multi-agent system that measures team efficiency based on the personality traits of the team member;
- A way to generate synthetic behavioral data using deep reinforcement learning agents through self-play.

## 2. Materials and Methods

In this section, we analyze the core game mechanics, how the escape room environment works, the main components of the agents, the action and state space, the rewards, and, finally, the training method.

### 2.1. Background

#### 2.1.1. Reinforcement Learning

A Reinforcement Learning (RL) model is typically formalized as a Markov Decision Process (MDP) represented by the 5-tuple  $M = (S, A, p, \gamma, R)$ . In this tuple,  $S$  denotes the state space (all possible states that the environment can be in at any given time),  $A$  is the action space (all possible actions the agent can take in the environment),  $p$  is the environment dynamics function (rules or probabilities that define how the environment transitions between states),  $\gamma$  is the discount factor (weighing future rewards relative to immediate rewards), and  $R$  is the reward function (feedback mechanism that the environment uses to indicate the success or failure of the agent's actions) [11]. As an environment, we define the world context in which our RL agents operate.

In our agent system, we use a centralized critic for the environment dynamics function  $p$  and incorporate a discount factor  $\gamma$  of 0.99, as they are set by the ml agents package and our configuration files. The observation space, action space, and rewards are discussed in detail in the next sections.

Training in RL occurs through iterative interactions between the agent and the environment. The agent learns to maximize cumulative rewards by selecting actions, receiving feedback, and adjusting its policy, which is the strategy it uses. Each step provides rewards, guiding the agent to refine its decision-making strategy. The process unfolds across multiple episodes, which are sequences of interactions that start from an initial state and end when a specific goal or condition is reached. Over time, the agent balances exploration (trying new actions) and exploitation (leveraging learned strategies) to optimize long-term outcomes [11].

It is important to note that multiple autonomous cooperating agents, collectively referred to as a “multi-agent system”, operate within a shared environment. These agents observe the environment using sensors and interact with it through actuators. While

pre-designed behaviors can be embedded in such agents, they often need to learn new behaviors and actions online. This continuous learning process leads to improvements in the performance of individual agents, as well as the overall multi-agent system [12].

We employed the MA-POCA (Multi-Agent Posthumous Credit Assignment) algorithm from the ML-agents package [13]. The algorithm learns a centralized value function to estimate the expected discounted returns of the group of agents and a centralized agent-centric counterfactual baseline to facilitate credit assignment [8]. Essentially, this allows for the agents that have escaped from the environment to continue training and learn quicker and more effectively.

This was the main reason we chose MA-POCA against others like Independent Q-Learning that can be limited to complicated and dynamic environments [14] or the Multi-Agent Advantage Actor-Critic (MA-A2C) that struggles in synchronization across agents and may struggle in high-dimensional environments [15].

### 2.1.2. Personality Traits

We have adopted the OCEAN Five personality characteristics model, which stands for Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. This model is one of the most widely recognized and utilized frameworks for personality assessment in the field of psychology [16]. The OCEAN model provides a comprehensive and scientifically grounded approach to understanding individual differences, making it a valuable tool for analyzing behavior in various contexts, including gaming and human-agent interaction.

While newer models offer useful insights, we have chosen to implement the OCEAN Five model due to its widespread acceptance, empirical support, and versatility. The OCEAN model remains the most commonly used framework for personality assessment, offering a reliable and consistent structure for studying a broad range of personality traits and behaviors across diverse populations. Furthermore, its applicability and recognition make it a practical choice for our work, ensuring that our findings align with established psychological research and are easily interpretable by both researchers and practitioners in the field.

## 2.2. Related Work

Previous research has investigated the development of adaptive agents within serious game environments to address various domain-specific challenges. For instance, a notable example is the multi-agent system integrated into the SIMFOR project, which is a serious game designed for crisis management training and simulation [17]. The agents in the SIMFOR system are built on the Belief-Desire-Intention (BDI) deliberation model, which enables them to simulate complex decision-making processes. These agents can be customized and configured to facilitate the construction of diverse crisis scenarios, thereby supporting training exercises in emergency response and crisis management.

In contrast, our work focuses on a fundamentally different goal. The multi-agent system we have developed is not constrained to a specific domain like crisis management but is designed to establish a generalized and standard methodology of personality emulation. Rather than contributing to the construction of specific training scenarios, our system emphasizes the emulation of personality-driven behaviors and adaptive traits that can be seamlessly applied to agents in both serious games and entertainment-oriented games.

A multi-agent system based on decision trees with data from traditional questionnaires has been developed to assess personality types [18]. It helps determine a person's dominant personality, suggest job placements, or predict reactions based on specific traits.

The main difference between this approach and ours is that it relies on self-administered questionnaires, which can introduce biases like self-reporting errors. In contrast, our system uses in-game behavior data, minimizing these biases for a more objective assessment of personality.

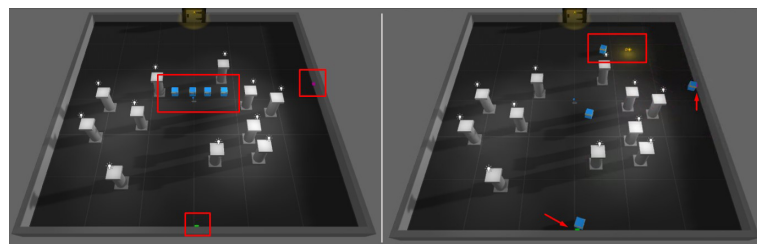
Similarly, previous studies have used game environments to simulate behavior modeling [9], such as analyzing how a single agent's movement in a simple room reflects its

openness personality trait. Similarly, [5] demonstrated how escape rooms can measure personality through gameplay metrics and puzzles. In contrast, our work develops a multi-agent system with a reward function methodology to assess team efficiency in a gamified environment while exhibiting complex behaviors driven by multiple personality traits.

### 2.3. Game Mechanics and Environment

Our multi-agent system was implemented as a 3D environment in the Unity platform. Focusing on creating an engaging yet accessible ER experience, we harnessed Unity's assets to construct a captivating setting.

Our designed environment consists of two buttons (on the walls) that, when activated, unveil a key required for unlocking the final door to escape, as can be seen in Figure 2. The team, consisting of 4 agents, must navigate through the pillars and past the columns to seek out and press buttons.



**Figure 2.** Unity Environment implementation. The agents and the buttons are illustrated with boxes in the left image, while the key and the agents pressing the buttons are shown with the arrows in the right one.

With each new game, the positions of the buttons, keys, door, and pillars are dynamically generated and positioned within the room at random intervals. Notably, the dynamism of our environment adds a layer of unpredictability to the gameplay and further enhances the immersive experience, and it tests the adaptability of the agents. This ensures that the agents are constantly faced with fresh challenges and opportunities for discovery that demand quick thinking and strategic problem-solving skills. This dynamic and ever-changing environment ensures that each gameplay session is unique and offers new opportunities for the agents to learn.

### 2.4. Action Space

In this multi-agent environment, the action space available to each agent is deliberately structured to encompass four distinct options, which are tailored to facilitate navigation and interaction within the virtual landscape. These options are split into two categories: movement and rotation. For movement, agents are equipped with the ability to advance or retreat backward, enabling them to traverse the intricate terrain of the ER with precision and agility. Similarly, rotation options afford agents the capability to pivot either left or right, affording them the flexibility to survey their surroundings and strategize accordingly.

Each of these action choices is encoded as a Boolean variable, ensuring clarity and efficiency in decision-making processes. This streamlined approach not only simplifies the agents' decision-making process but also enhances the overall responsiveness and adaptability of the system. By empowering agents with a diverse array of movement and rotation options, we aim to provide them with the requisite tools to navigate and interact with the virtual environment in a manner that closely mirrors real-world behaviors and capabilities.

In addition to the overall architecture of the multi-agent environment, our monitoring system records and analyzes the unique features of each agent's gaming style. This holistic method enables us to gain deeper insights into the complexities of agent behavior, shining light on crucial areas such as movement dynamics, interpersonal relationships, and involve-

ment with the environment, which can range from picking up keys and pressing buttons to attempting to access the exit door before it is unlocked.

One key aspect is the agents' movement patterns, providing valuable data on their navigation strategies and spatial awareness within the virtual realm. By scrutinizing their locomotion, we can discern tendencies and preferences that inform their decision-making processes, offering valuable insights into their cognitive mapping abilities and strategic maneuvering.

### 2.5. State Space

Before proceeding with any decision-making process, the agents observe and diligently gather specific information about their surroundings. This includes information like the location of buttons, keys, and doors, as well as their current state, such as whether they are pressed, found, or unlocked. What's more, when one agent picks up a key or pushes a button, this information is immediately communicated to all the other agents through their observations.

By leveraging these observations, agents can effectively assess their environment, allowing them to make informed decisions and navigate through the virtual world with precision and efficiency.

In our Unity implementation, agents utilize a sophisticated observation mechanism centered around ray-cast observations. This advanced technique utilizes physics functions to project a ray into the environment scene, providing agents with crucial insights upon a successful intersection with a target object. This method returns a Boolean value for each one of some predefined tags, creating a final vector.

So, we use ray-cast arrays that project a total of 15 rays, each of which checks for specific tags, which include the key, buttons, doors, other agents, and pillars. Also, all the agents share 3 Boolean variables as common knowledge, regarding the buttons (pressed or not), the key (picked or not), and the door (unlocked or not). This results in a final state space size of 78, which captures all the relevant details about the environment that the agents need to be aware of.

### 2.6. Rewards

The optimal design of MDPs is often challenged by the issue of sparse (infrequently or only upon achieving specific goals) reward [11], which poses a significant obstacle to the agent's learning process. This challenge is also encountered in the context of an ER setting, where the most significant rewarding events occur only when all agents manage to escape, making it difficult for individual agents to identify the chain of events that led to the successful outcome [11].

To address this issue, we propose a two-fold reward system. The first part of the reward system is related to the ER game and the team's ability to solve it. The second part is composed of custom reward functions that are designed to promote specific agent behaviors. These two types of rewards are monitored separately, with the former reflecting the team's success in escaping the room and the latter aimed at making individual agents to exhibit specific behaviors while assessing the whole team's performance.

Regarding the ER game, we have implemented appropriate reward functions for each agent that reward them when they reach specific checkpoints in the room, such as picking up a key. Furthermore, a final team reward is given when all agents successfully escape. The rewards are meant to encourage the agents to strive toward their ultimate objective, which is to escape the room, as opposed to just locating the key or opening the door as soon as possible. The proposed reward system includes a time-based reward for each agent when they find a key, unlock a door, or successfully escape. Additionally, a team reward of multiplied time-based reward is awarded when all agents successfully escape.

As previously mentioned, agents are rewarded based on their behaviors, which are modeled using advanced mathematical approaches inspired by the HiDAC crowd simulation framework. HiDAC introduces mathematical formulas to represent specific behaviors

linked to personality traits. For example, panic behavior is associated with traits such as conscientiousness and neuroticism. In HiDAC, personality traits are modeled as Gaussian distributions, where each behavior is a function of the corresponding trait calculated using a Gaussian value (ranging from 0 to 1).

While HiDAC formulas focus primarily on the Gaussian distribution itself, our implementation goes a step further by correlating these Gaussian values with gameplay actions. Using HiDAC's mathematical models as a foundation, we developed custom formulas (Table 1) that map personality-based behaviors to specific in-game actions (e.g., running, pushing other agents) or mechanics (e.g., collision detection).

We must note that in our implementation, the Gaussian values ranged from  $-1$  to  $1$ , depending on if we want to train the agent with a low ( $-1$  to  $0$ ) or high ( $0$  to  $1$ ) trait. This range allowed us to nullify the rewards of specific behaviors (when set the Gaussian to  $0$ ) so that we could focus the train on specific traits.

**Table 1.** Traits to rewards based on actions and characteristics

Personality Trait	Behaviors (Original)	Reward (Custom)
Openness	Train Explore	(Agent Characteristic – added new state) num of correct actions $\times 10$
Conscientiousness	Panic Impatience Right Preference	$0.3 \times -2 \times \Psi^C + 2$ if run & push $0.3 \times (1 - \Psi^C)$ If $\Psi^C > 0$ then $\Psi^C \times (\text{times right/time}) \times 0.3$
Extraversion	Leadership Communication Impatience Pushing Personal Space Walk speed Gesture	$0.3 \times \text{mean speed} \times \Psi^E$ 1 if num of communication actions used $\geq \Psi^E \geq 0.5$ $0.3 \times 2 \times \Psi^E - 1$ if $\Psi^E > 0$ 1 if num of push actions used $\geq 0.3 \times \Psi^E \geq 0.5$ (Agent Characteristic – collider) Max walk speed + 1 Num of correct gestures $\times 10$
Agreeableness	Impatience Pushing Right Preference Wait Radius Wait Timer	$0.3 \times (1 - \Psi^A)$ if run each step 1 if num of push actions used $\geq 0.3 \times (1 - \Psi^A) \geq 0.5$ $0.3 \times (\text{Times right/time}) \times \Psi^A$ (Agent Characteristic – collider) (Agent Characteristic – wait timer)
Neuroticism	Leadership Panic	Mean speed $\times (1 - \Psi^N) \times 0.5$ $\Psi^N \times 0.5$ if run and push

So, the agents are placed in the game environment and, before their training, their personalities are chosen, and the corresponding Gaussian is set. At the end of each episode, the agent is rewarded based on the metrics of each action. So the process is the following:

- We select the trait we want to train the agent to emulate (e.g. conscientiousness);
- We set the Gaussian to define the selected trait ( $G_o = 0$ ,  $G_c = 1$ ,  $G_e = 0$ ,  $G_a = 0$ ,  $G_n = 0$ );
- During training at the end of each episode:
  1. We collect behavior action metrics (panic behavior = number of push actions on other agents);
  2. We reward the agent by multiplying the behavior metric  $\Psi$  with the corresponding Gaussian ( $R = \text{Panic behavior} \times (G_c + G_n)$ ).

It is important to note that a single behavior can be associated with multiple personality dimensions, as it may have both positive and negative effects. The relationships between different types of behaviors and personality traits have been further explored and are detailed in Table 2. For example, Leadership can be found only as a positive trait if a person is positively conscientious. It can also be found as a negative attribute if he shows negative

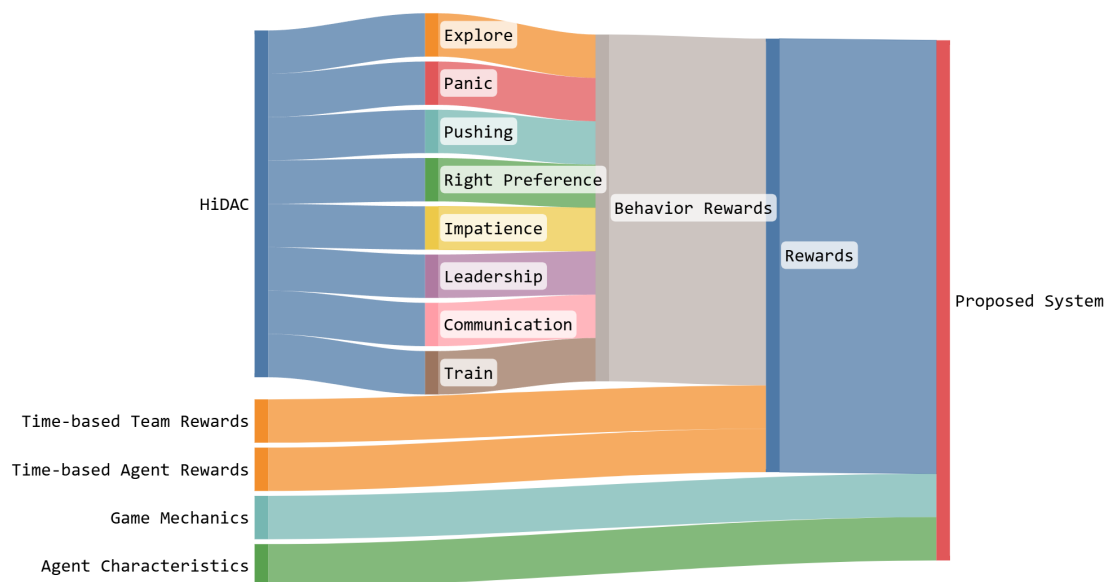
agreeableness traits. Otherwise, if he has neurotic tendencies, the leadership behavior will have both positive and negative influences.

Last but not least, the rewards were designed to balance task-oriented objectives (e.g., solving puzzles, escaping the room) with personality emulation and behaviors, since there are three layers of rewards: the team reward (escaping), the game-oriented reward (pressing a button or finding the key), and the behavior rewards that we analyzed.

**Table 2.** Behaviors, personality traits, and bibliography review.

Behaviors	Traits	Reference
Exploration	Openness (+)	[19]
Personal Space	Openness (+)	[20]
Panic	Neuroticism (+), Conscientiousness (−)	[21]
Leadership	Extraversion (+), Conscientiousness (+), Neuroticism (−)	[22]
Impatience	Agreeableness (−), Conscientiousness (−)	[23]
Training	Openness (+)	[24]

In conclusion, the rewards for MindEscape’s agents are based on the HiDAC framework, with specific adaptations shown in Table 1. The agents, their actions, and custom reward functions are built upon the core mechanics of HiDAC while introducing novel methods to emulate the traits of the OCEAN Five personality model, as shown on Figure 3.



**Figure 3.** From HiDAC to Mindescape environment.

### 2.7. Training Methodology

Each agent in the team is trained on specific personality traits and behaviors, using the prior reward model. The team’s objective is to ensure the successful escape of all agents while striving to achieve the same tasks as before.

For the agent’s training, we tried different hyperparameter values, as seen in the following Table 3. The best hyperparameter values were different for the simple agents and the ones emulating personality traits, something that was expected since the second agents learned to show more complex behaviors.



**Table 3.** Hyperparameters and best values.

Hyperparameters	Values
batch size	64, 128, 256, <u>516</u>
buffer size	64,000, 128,000, 256,000, <u>516,000</u>
learning rate	0.005
hidden units	512
number layers	2

Italics are for the best default agents and the underlined for the personality agents.

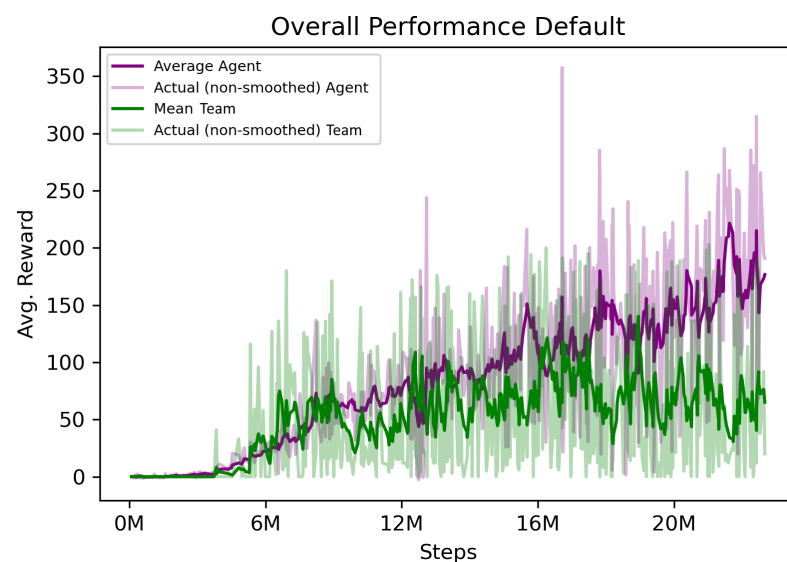
### 3. Training Results and Evaluation

To train the agents that imitate human behavior, we first trained a simple multi-agent system to solve the ER. In all the following figures of this section (Figures 4–8), as well as in the Appendix A, the X axis represents the training steps, and the Y axis shows the rewards (or the time) for each step.

All the agent teams were trained for 25 million steps, which took around 6 h for each team. For assessment purposes, we meticulously scrutinized group rewards, episode length, and, in certain instances, behavior metrics to gauge their proficiency and effectiveness.

#### 3.1. Default Team Rewards

The results of the best default multi-agent team (without imitating human behavior) are shown in Figure 4. As we can observe, agents learned almost from the start to press buttons and had some positive rewards. In contrast, the team learned to escape and solve the room much later, almost at 3M steps, as can be seen by the first peaks of the green line. This was due to the dynamically changing environment, and this performance was set as a standard for the next agents, which was displayed in every figure in a light and shaded way. We did not intend to compare the agents' to the team rewards, but this is an indication of when training is effective for the team.

**Figure 4.** Default (no personality) agents and agent teams training results.

#### Uniform Team Rewards

Following that, we implemented the reward functions as stated before, regarding the behaviors, and trained agent teams with different kinds of traits. The teams consisted of four agents, each with the same personality at first, and then combinations of different kinds of traits. In the first four teams that were trained, all their members emulated behaviors from only one of the available personalities: Openness, Non-openness, Conscientiousness, Non-conscientiousness, Extrovert, Introvert, Agreeable, and Non-Agreeable, Neurotic, and

Non-neurotic. This way, we could set a baseline of how agents with the same personality traits are performing.

In all the diagrams of this section, we observe the smoothed and actual rewards of each team at an agent level and at a group level, as well as the default agent reward, to be able to understand how the behaviors change the rewards and effectiveness.

In Figure 5, we observe the performance of agents with varying levels of Openness. The team with high Openness scored lower rewards compared to the default agents. This difference in performance is primarily due to the high Openness agents' propensity to explore more extensively. Their curiosity and desire for new experiences lead them to spend more time investigating their environment, which, while potentially valuable for gathering information, results in less efficient task completion.

The exploratory behavior of high-Openness agents can divert attention from immediate goals, causing delays and reducing overall effectiveness. This tendency to prioritize exploration over direct action can lead to lower rewards, as these agents may overlook simpler, more immediate solutions in favor of investigating less obvious possibilities.

Teams of agents with all the personalities were trained, and the diagrams with complementary analysis can be found in the Appendix A. Table 4 presents a comparison of the teams, summarizing their mean reward, meantime, and success rate to provide an overview of their overall effectiveness.

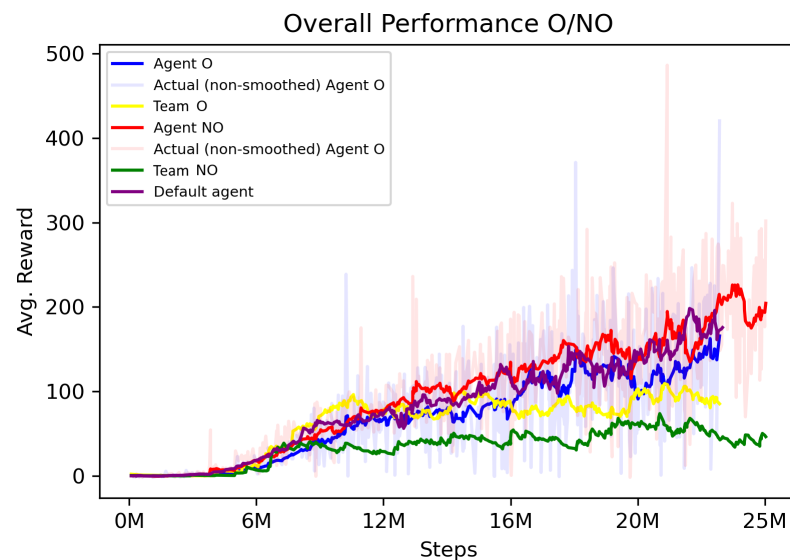


Figure 5. Openness (O) and Non-Openness (NO) agents and team training results.

We present the Openness agents diagrams as an indicative showcase, and we include the diagrams of all the agents in the Appendix A for clarity.

Table 4. Uniform personality teams and effectiveness.

Personality (+/−)	Mean Reward <sup>1</sup>	Mean Escape Time <sup>2</sup>	Success Rate (All Agents Escaped)
Default	1970	410	27%
Openness	1695/2154	505/411	37/20 %
Conscientiousness	1890/2005	446/430	29/21%
Extraversion	1776/1891	428/409	29/30%
Agreeableness	1819/2003	440/430	37/27%
Neuroticism	2077/1999	415/432	20/38%

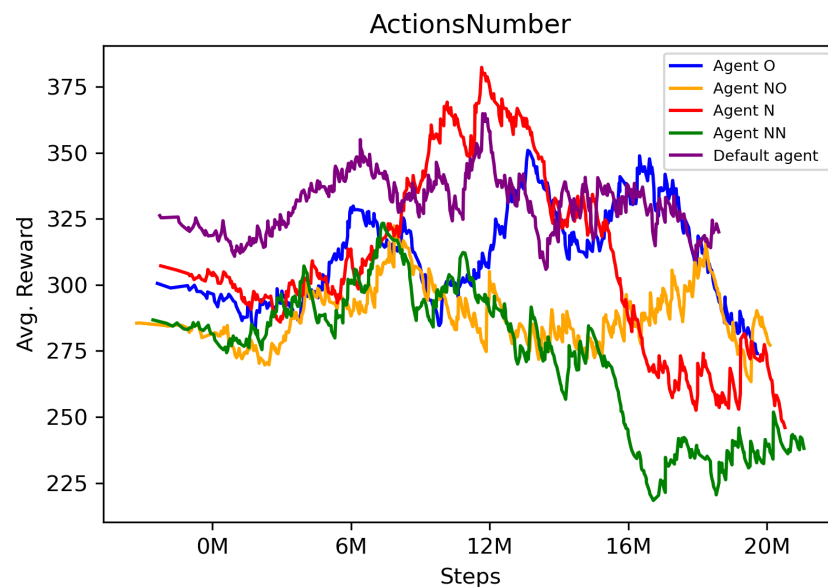
<sup>1</sup> The final reward number based on agent actions. <sup>2</sup> The mean play time in seconds.

We can observe from the Table 4, the influence of positive and negative traits on the effectiveness of the teams. This showcases just how different the agents behaved in the same environment.

### 3.2. Uniform Team Behaviors and Escape Time

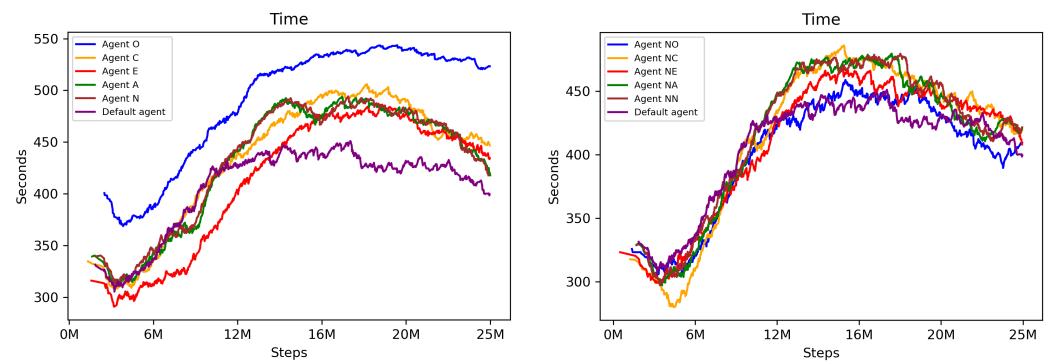
After analyzing the rewards, it is essential to examine the behaviors exhibited by the agents. In Figure 6, we observe the variations in pushing actions among agents, which were influenced by their personality traits. Neurotic agents displayed a high frequency of actions, closely mirroring the behavior of default agents. This similarity suggests that neurotic agents, driven by their heightened sensitivity to stress and urgency, are more reactive and exhibit a greater number of actions in an attempt to manage their environment and achieve their goals.

In contrast, calm, non-neurotic agents demonstrated significantly better cooperation. Their lower frequency of actions indicates a more deliberate and thoughtful approach, prioritizing coordination and strategic planning over immediate reactive behaviors. These agents tend to focus on maintaining stability and harmony within the team, resulting in more efficient and cohesive group dynamics. Their ability to remain composed under pressure allows them to execute tasks with greater precision and teamwork, enhancing overall performance.



**Figure 6.** Openness (O), Non-Openness (NO), Neurotic (N), Non-Neurotic (NN) Agents actions number.

Last but not least, one more element we need to look at is time. As seen in Figure 7, the Openness team is the one that took the longest, something expected since they tend to explore more. We can see that all the agents were generally slower than the default agents, since they showed more complete behavior. We must also note that the Non-Openness team was the quickest of all.



**Figure 7.** Agents mean play time (Openness (O), Non-Openness (NO), Conscientiousness (C), Non-Conscientiousness (NC), Extrovert (E), Non-Extrovert (NE) Neurotic (N), Non-Neurotic (NN) agents).

### 3.3. Non-Uniform Team Rewards

The next step was to train agents so that each one had a different kind of personality trait. We allocated the initial number of agents based on the 25 percent ratio of introverts to extroverts in a community [25] to also the Neurotic and Non-Neurotic traits.

Based on this hypothesis, we conducted an experiment to observe the impact of introducing one Introvert into a team of three Extroverts (red) and one Extrovert into a team of three Introverts (green), comparing their performance to the previous teams. As shown in Figure 8, the team comprising three Introverts and one Extrovert (green) outperformed all other configurations. The presence of the Extrovert agent proved beneficial, as they naturally assumed a leadership role, effectively organizing and directing the Introverts, whose preference for reflective and deliberate actions complemented the Extrovert's initiative and energy.

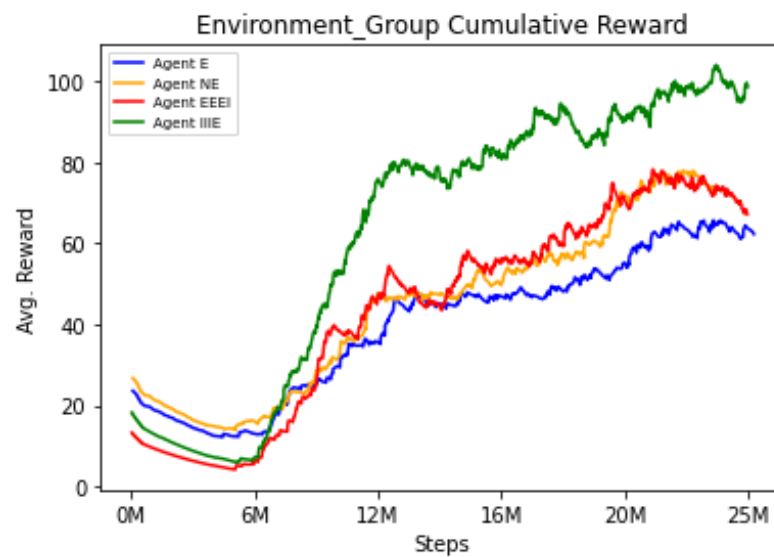
In contrast, the team with three Extroverts and one Introvert (red) performed similarly to the all-introvert team but outperformed the all-extrovert team. The Extroverts' competitive nature caused inefficiencies, but the Introverts introduced a stabilizing element, reducing competition. This allowed the team to work more cohesively and efficiently than the all-extrovert team.

Overall, these findings highlight the importance of balancing team dynamics with diverse personality traits. The Extrovert's leadership can harness the strengths of Introverts, while an Introvert's presence in an Extrovert-dominated team can introduce a calming influence, fostering a more harmonious and effective team environment.

In the same way as before, we showcase the results of other teams and with a variety of personalities, while their results, diagrams, and complementary analysis can be found in the Appendix A.

As we can see in Table 5, there was a significant change of the success rates of the teams when there was more than one personality in the agents. For example, when we included 1 Introvert, the team improved by 10% but on the other hand, when we introduced 1 Neurotic to a Non-Neurotic team, the effectiveness dropped by 5%.

Based on the results, it is safe to say that the personality traits of the agents of each team do play a significant role in how the group operates, how efficient it is, and how quickly they manage to escape. This means that the reward functions and how they are set can indeed replicate and emulate how the traits and behaviors are exhibited in real life based on logic.



**Figure 8.** Extrovert (E) and Introvert (Non-Extroverts—NE), 3 Extroverts and 1 Introvert (EEEL), and 3 Introverts and 1 Extrover (IIIIE) agents training results.

**Table 5.** Non-uniform personality teams and effectiveness.

Personality (+/−)	Mean Reward <sup>1</sup>	Mean Escape Time <sup>2</sup>	Success Rate (Difference <sup>3</sup> )
3 Extroverts and 1 Introvert	1150	380	39% (+10%)
3 Introverts + 1 Extrovert	1318	403	48% (+18%)
3 Neurotic + 1 Non Neurotic	1654	362	36% (+16%)
3 Non Neurotic + 1 Neurotic	1530	360	32% (−6%)

<sup>1</sup> The final reward number based on agent actions. <sup>2</sup> The mean play time in seconds. <sup>3</sup> The difference (number in parenthesis) with the corresponding uniform teams.

### 3.4. System Evaluation

Emulating human behavior in a game environment is a complex task that cannot be accurately measured without involving a diverse set of human players. In some cases, however, simulation results can provide valuable guidelines for predicting human behavior, especially when focusing on specific personality traits. This approach guided the evaluation of our models, with the HiDAC crowd simulation serving as the baseline for comparison [26].

The results of the training process support the initial hypothesis that the agents can emulate human behaviors to some extent. This is evidenced by two key factors: the reward values obtained during training and the visual inspection of the agents' behaviors within the game environment. Each agent exhibited a range of behaviors, shaped by the reward functions, that mimic simplified versions of human actions. While individual agents displayed different gameplay styles and actions, these variations reflect the interplay of multiple personality traits, which manifest at varying levels depending on the mental state of the player at any given moment.

Although personality is inherently complex, we were able to distill each trait into simplified behavioral patterns, allowing us to create agents that simulate core aspects of a specific trait. This simplification, however, may be considered a limitation of the work, as it reduces the full depth of human personality into more basic representations.

For an effective evaluation of the agents' behavior, it is essential to establish clear behavioral ground truths.

As an initial benchmark, experimental results from Durupinar provide a useful point of comparison. These results, summarized in Table 6, serve as a reference for assessing the agents' performance and offer insights into how well the modeled behaviors align with expected personality traits. This shows that the agents indeed have similarities with the human behaviors that they emulate.

**Table 6.** Tested scenarios used for result comparison and evaluation.

Traits	HiDAC Behavior	Our System Behavior
Openness	As Openness increases, the number of places they explore increases, and thus, they leave the building later.	The mean escape time of the Openness agents team is increased (+25%).
Extroverts and Introverts	The Extroverts approach the attraction point in a shorter time. In addition, when there are other agents blocking their way, they tend to push them to reach their goal.	The Extroverts show better times of escaping (+5%), as well as higher push action metrics (+35%)
Conscientiousness and Agreeableness	The shortest time is achieved when Conscientiousness and Agreeableness are the highest. The result is expected as agreeable and conscientious individuals are more patient, they do not push each other, and are always predictable, as they prefer the right side to move on. Also, the longest time is obtained when both values are minimal.	The agents that have high Conscientiousness and Agreeableness are quicker than the ones with lower (+5% and +4%, respectively) and lower push actions (>−25%)
Neurotic	Agents that are Neurotic with less Conscientiousness tend to panic more, pushing other agents forcing their way through the crowd, and rushing to the door.	The Neurotic agents are quicker to finish the room (+5%) though less successful (−55%).

#### 4. Discussion

Based on the findings, it can be confidently asserted that the individual personality traits exhibited by the agents within each team have a substantial influence on the overall dynamics of the group, including its operational efficiency and the speed with which they achieve their objectives, such as escaping.

This underscores the critical importance of the reward functions employed and their configuration, as they have the capacity to accurately mirror and replicate the diverse range of traits and behaviors observed in real-life scenarios. In essence, the design and implementation of these reward structures have the potential to faithfully emulate the complexities of human traits and behaviors through logical frameworks.

In our analysis, we have observed intriguing trends regarding the performance of agents based on their personality traits. Particularly noteworthy is the efficiency and effectiveness exhibited by agreeable agents, who demonstrate a remarkable propensity for cooperation and swift problem-solving. These individuals, characterized by their affable and cooperative nature, seamlessly navigate through challenges, leveraging their strong interpersonal skills to foster collaboration within the team dynamic.

On the other side, introverted agents, while equally effective in their problem-solving capabilities, tend to approach tasks at a more deliberate pace. Their cautious and introspective nature often translates into a methodical approach to problem-solving, resulting in slower but effective progress through the ER challenges.

Moreover, our findings reveal intriguing dynamics when teams are composed of agents representing a spectrum of personality traits. In such scenarios, we have observed challenges in collaboration, particularly during the initial stages of engagement. The diverse array of personality traits within the team dynamic can lead to discrepancies in communication styles, decision-making processes, and conflict-resolution strategies, posing initial obstacles to seamless cooperation.

These insights underscore the intricate interplay between individual personality traits and collective team dynamics within the multi-agent environment. By understanding and leveraging these nuances, we can tailor strategies and interventions to optimize team performance, fostering a cohesive and synergistic approach to problem-solving within the ER setting.

There is a foundational premise supporting the assertion that team efficiency within a gamified environment, where agents must collaborate to achieve objectives, is intricately linked to the individual behaviors and traits exhibited by each team member. The findings from our study reinforce this notion, highlighting the significant impact of specific metrics and rewards established within the system.

Furthermore, our findings offer valuable insights into the broader applicability of this reward methodology beyond the confines of our specific multi-agent environment. The success of our approach in shaping agent behaviors and fostering efficient collaboration serves as a promising indication of its potential utility across a diverse array of gaming scenarios.

In essence, our study not only showcases the intricate relationship between individual traits and team efficiency but also underscores the transformative potential of tailored reward methodologies in shaping agent behaviors and optimizing performance across various gaming contexts.

In addition, these results open the door for further investigation into how external factors—such as task difficulty, time pressure, and environmental complexity—may interact with agent personality traits to influence team dynamics. For instance, under time-critical conditions, teams with dominant Agreeable or Conscientious agents may perform better due to their cooperative and focused problem-solving tendencies, whereas under exploratory tasks, more Open or Introvert agents might excel due to their propensity for careful analysis and creativity.

Finally, the results not only showcase the relationship between individual traits and team efficiency but also underscore the potential of custom made reward functions in shaping agent behaviors and optimizing performance across various gaming contexts. These findings show the way for further advancements in designing agents with personality for interactive simulations, virtual training environments, and commercial games, where emulating realistic team behaviors and personality can enhance immersion, engagement, and problem-solving capabilities.

## 5. Conclusions

The present study introduced a multi-agent environment in the form of an ER, where agents emulate the OCEAN Five Personality Traits characteristics based on custom reward functions. The environment was designed in a way that permitted us to gather data regarding the play style of the agents, as well as their interactions with the room and between them.

The RL agents emulate human-like decision-making processes through several technical elements. The game environment is modeled as a high-dimensional space, capturing items, player positions, and interactions. The action space defines how agents can move, pick up objects, solve puzzles, and interact, with each action impacting the game state based on the agent's personality. We implemented a custom reward function designed to balance task-oriented objectives, such as solving puzzles and escaping the room while looking at personality emulation and behaviors. During the training process, the agents are trained using advanced Deep RL algorithms like MA-POCA (Multi-Agent Posthumous Credit Assignment), involving extensive game sessions to improve strategies through trial and error. These elements together ensure that the RL agents provide a comprehensive and dynamic understanding of personality traits within the game.

In summary, the multi-agent teams developed in this study successfully emulated personality traits and exhibited distinct behaviors while solving tasks within an Escape Room (ER) environment. During the training process, we collected sufficient data to analyze and compare the efficiency and speed of each team based on the personality traits of their

respective members. The findings highlight that personality traits significantly influence team dynamics, task-solving efficiency, and the speed at which specific objectives are achieved. These results suggest that our approach can serve as a foundation for developing agents capable of learning to solve tasks in other game environments emulating behaviors, using these or similar reward functions. Furthermore, such agents could also be deployed as Non-Playable Characters (NPCs) capable of exhibiting realistic, personality-driven behaviors, enhancing immersion and player experience in interactive digital environments.

Moreover, the data generated can be used to train a Multi-Output Regression System (MORS) using supervised machine learning algorithms as future work. This process will use a MORS to assess personality traits by analyzing data and behavioral patterns exhibited during gameplay by human players. Future work will investigate the implementation of a MORS in live, interactive gaming environments, where the system assesses players' behaviors and provides immediate personality insights.

**Author Contributions:** Conceptualization, G.L. and I.V.; Methodology, G.L. and I.V.; Software, G.L.; Validation, G.L.; Resources, G.L.; Data curation, G.L.; Writing—original draft, G.L.; Writing—review & editing, G.L. and I.V.; Visualization, G.L.; Supervision, I.V. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

**Acknowledgments:** This paper is a revised and expanded version of a paper entitled Machine Learning Methods for Emulating Personality Traits in a Gamified Environment, which was presented at Pireus, Greece, 11–13 September 2024.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

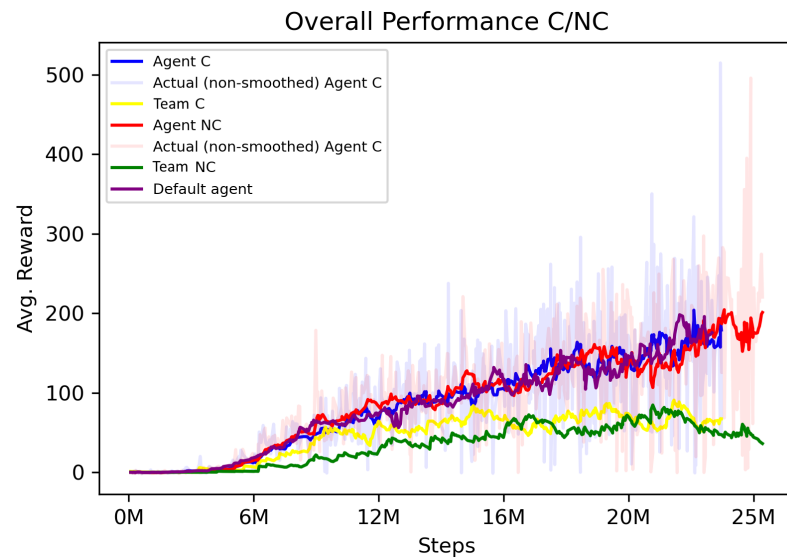
The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
ER	Escape Room
RL	Reinforcement Learning
MDP	Markovian Decision Process
MA-POCA	Multi-Agent Posthumous Credit Assignment
OCEAN	Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism
MORS	Multi-Output Regression System
MA-A2C	Multi-Agent Advantage Actor-Critic

## Appendix A. Training Results and Analysis

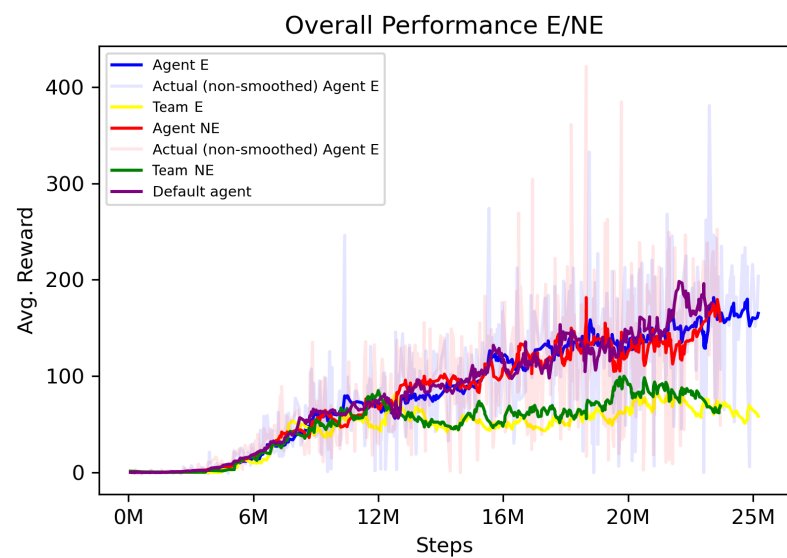
The Figure A1 showcases agents with varying levels of Conscientiousness. We observe that both teams, those with high and low Conscientiousness, performed similarly to the default agents. However, the Non-Conscientious team exhibited some significant performance drops.





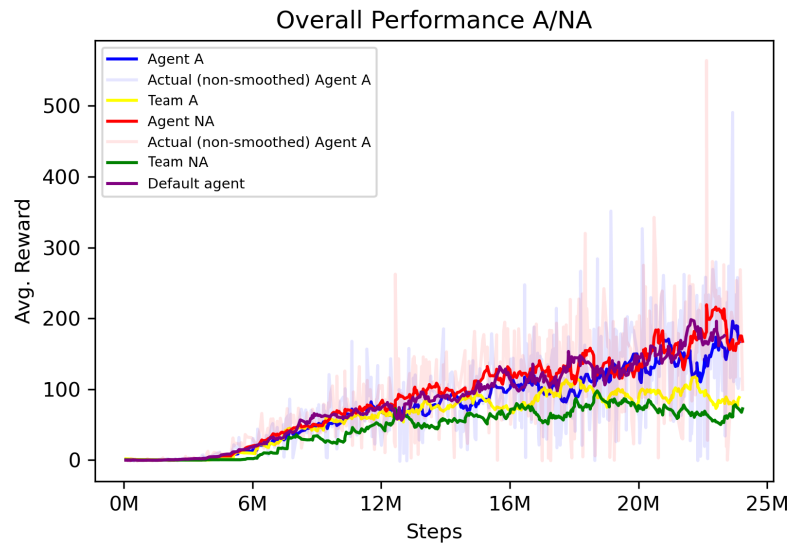
**Figure A1.** Conscientiousness (C) and Non-Conscientiousness (C) agents and team rewards.

In Figure A2, we can observe the performance of the Extrovert and Non-Extrovert (Introvert) agents. Both teams outperformed the default agents, suggesting that the default agents' behavior lacks stability and efficiency. Extrovert agents often act independently, focusing on individual tasks rather than cooperating with others. This tendency to work alone can lead to a lack of coordination and synergy within the team. On the other hand, Introvert agents tend to be less proactive and may hesitate to take initiative, which can result in slower progress and less overall activity.



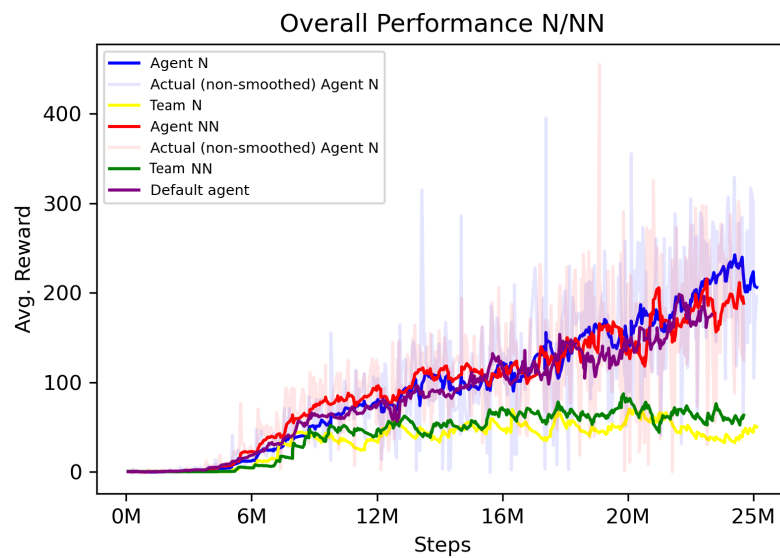
**Figure A2.** Extraversion (E) and Non Extraversion/Introversion (NE) agents and team rewards.

A similar behavior is shown in Figure A3 from the agents that have high or low agreeableness. Non-agreeable agents tend to perform better since they proactively take the initiative to achieve the goal. On the other hand, agreeable agents focus more on organizing and coordinating with each other. This tendency to prioritize cooperation and harmony can sometimes slow down their progress, as they spend more time on communication and consensus-building rather than taking decisive action. As a result, non-agreeable agents, who are more willing to act independently and assertively, often reach their objectives more efficiently.



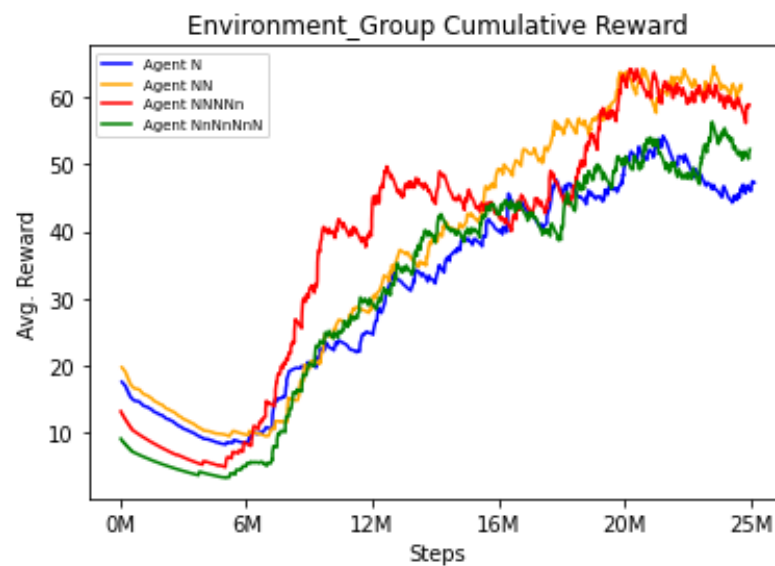
**Figure A3.** Agreeable (A) and Non-Agreeable (NA) agents and team rewards.

Last but not least, the Neurotic and Non-Neurotic agents, as shown in Figure A4, tended to score better than the default agents. This difference in performance can be attributed to their inherent behavioral tendencies in such scenarios. Neurotic agents, often characterized by high emotional reactivity and sensitivity to stress, may excel due to their heightened vigilance and urgency in completing tasks. Their propensity to anticipate potential problems can lead to faster and more meticulous execution of their goals.



**Figure A4.** Neurotic (N) and Non-Neurotic(NN) agents and team rewards.

Conversely, the teams with one Non-Neurotic agent alongside three Neurotic agents (red) outperformed all other configurations, as shown in Figure A5. This superior performance can be attributed to the balancing influence of the Non-Neurotic agent, who brings a sense of calmness and stability to the team. The Non-Neurotic agent’s emotional resilience and steady demeanor help to mitigate the high reactivity and stress sensitivity of the Neurotic agents. This calming effect reduces the overall anxiety within the team, allowing the Neurotic agents to function more effectively and focus on their tasks without becoming overwhelmed.



**Figure A5.** Only Neurotic agents (N) team, only Non-Neurotic agents (Nn) team, team with 3 Neurotic agents (NNNNn) and team with 3 Non-Neurotic agents (NnNnNnN) team rewards.

## References

- Bergner, R. What is personality? Two myths and a definition. *New Ideas Psychol.* **2020**, *57*, 100759. [CrossRef]
- Martinez, K.; Menéndez-Menéndez, M.I.; Bustillo, A. A New Measure for Serious Games Evaluation: Gaming Educational Balanced (GEB) Model. *Appl. Sci.* **2022**, *12*, 11757. [CrossRef]
- Reisenzein, R.; Hildebrandt, A.; Weber, H. Personality and Emotion. In *The Cambridge Handbook of Personality Psychology*, 2nd ed.; Matthews, G., Corr, P.J., Eds.; Cambridge Handbooks in Psychology; Cambridge University Press: Cambridge, UK, 2020; pp. 81–100. [CrossRef]
- Fotaris, P.; Mastoras, T. Escape Rooms for Learning: A Systematic Review. In Proceedings of the 13th European Conference on Games Based Learning (ECGBL 2019), Odense, Denmark, 3–4 October 2019.
- Liapis, G.; Zacharia, K.; Rrasa, K.; Liapi, A.; Vlahavas, I. Modelling Core Personality Traits Behaviours in a Gamified Escape Room Environment. *Eur. Conf. Games Based Learn.* **2022**, *16*, 723–731. [CrossRef]
- Angelini, G. Big five model personality traits and job burnout: A systematic literature review. *BMC Psychol.* **2023**, *11*, 49. [CrossRef] [PubMed]
- Durupinar, F.; Pelechano, N.; Allbeck, J.; Gudukbay, U.; Badler, N. How the Ocean Personality Model Affects the Perception of Crowds. *IEEE Comput. Graph. Appl.* **2011**, *31*, 22–31. [CrossRef] [PubMed]
- Cohen, A.; Teng, E.; Berges, V.P.; Dong, R.P.; Henry, H.; Mattar, M.; Zook, A.; Ganguly, S. On the Use and Misuse of Absorbing States in Multi-agent Reinforcement Learning. *arXiv* **2022**, arXiv:2111.05992.
- Liapis, G.; Lazaridis, A.V.I. Escape Room Experience for Team Building Through Gamification Using Deep Reinforcement Learning. In Proceedings of the 15th European Conference of Games Based Learning, Virtual, 23–24 September 2021.
- Liapis, G.; Vordou, A.V.I. Machine Learning Methods for Emulating Personality Traits in a Gamified Environment. In Proceedings of the 13th Conference on Artificial Intelligence (SETN 2024), Piraeus, Greece, 11–13 September 2024.
- Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*, 2nd ed.; Adaptive Computation and Machine Learning Series; The MIT Press: Cambridge, MA, USA, 2018.
- Silva, J.; Dutta, A. Multi-Agent Deep Reinforcement Learning for Multi-Robot Applications: A Survey. *Sensors* **2023**, *23*, 3625. [CrossRef] [PubMed]
- Unity ML-Agents Toolkit. 2021. Available online: <https://github.com/Unity-Technologies/ml-agents> (accessed on 21 December 2024).
- Foerster, J.; Nardelli, N.; Farquhar, G.; Afouras, T.; Torr, P.H.S.; Kohli, P.; Whiteson, S. Stabilising Experience Replay for Deep Multi-Agent Reinforcement Learning. *arXiv* **2018**, arXiv:1702.08887.
- He, K.; Doshi, P.; Banerjee, B. Latent Interactive A2C for Improved RL in Open Many-Agent Systems. *arXiv* **2023**, arXiv:2305.05159. [CrossRef]
- Jang, K.L.; Livesley, W.J.; Vernon, P.A. Heritability of the big five personality dimensions and their facets: A twin study. *J. Personal.* **1996**, *64*, 577–591. [CrossRef] [PubMed]
- Oulhaci, M.; Tranvouez, E.; Fournier, S.; Espinasse, B. A MultiAgent Architecture for Collaborative Serious Game applied to Crisis Management Training: Improving Adaptability of Non Played Characters. In Proceedings of the 7th European Conference on Games Based Learning (ECGBL 2013), Porto, Portugal, 3–4 October 2013.

18. Ramírez, M.R.; Moreno, H.B.R.; Rojas, E.M.; Hurtado, C.; Núñez, S.O.V. Multi-Agent System Model for Diagnosis of Personality Types. In Proceedings of the Agents and Multi-Agent Systems: Technologies and Applications 2018, Gold Coast, QLD, Australia, 20–22 June 2018; Jezic, G., Chen-Burger, Y.H.J., Howlett, R.J., Jain, L.C., Vlacic, L., Šperka, R., Eds.; Springer: Cham, Switzerland, 2019; pp. 209–214.
19. Abu Raya, M.; Ogunyemi, A.O.; Broder, J.; Carstensen, V.R.; Illanes-Manrique, M.; Rankin, K.P. The neurobiology of openness as a personality trait. *Front. Neurol.* **2023**, *14*, 1235345. [[CrossRef](#)] [[PubMed](#)]
20. Nam, N.; Hang Nga, N. Influence of personality traits on creativity and innovative work behavior of employees. *Probl. Perspect. Manag.* **2024**, *22*, 389–398. [[CrossRef](#)]
21. Javaras, K.N.; Schaefer, S.M.; van Reekum, C.M.; Lapate, R.C.; Greischar, L.L.; Bachhuber, D.R.; Love, G.D.; Ryff, C.D.; Davidson, R.J. Conscientiousness predicts greater recovery from negative emotion. *Emotion* **2012**, *12*, 875–881. [[CrossRef](#)] [[PubMed](#)]
22. Li, W.; Zhang, H.; Zheng, Y. Personality and Leadership: A Critical Review and Future Research Agenda from a Dynamic Perspective. In *Oxford Research Encyclopedia of Business and Management*; Oxford University Press: Oxford, UK, 2024.
23. Jiang, N.; Shi, M.; Xiao, Y.; Shi, K.; Watson, B. Factors Affecting Pedestrian Crossing Behaviors at Signalized Crosswalks in Urban Areas in Beijing and Singapore. In Proceedings of the ICTIS 2011: Multimodal Approach to Sustained Transportation System Development: Information, Technology, Implementation, Wuhan, China, 30 June–2 July 2011. [[CrossRef](#)]
24. Bergold, S.; Steinmayr, R. Personality and Intelligence Interact in the Prediction of Academic Achievement. *J. Intell.* **2018**, *6*, 27. [[CrossRef](#)] [[PubMed](#)]
25. Maureen, I.; Imah, E.; Savira, S.; Anam, S.; Mael, M.; Hartanti, L. *Innovation on Education and Social Sciences: Proceedings of the International Joint Conference on Arts and Humanities (IJCAH 2021) October 2, 2021, Surabaya, Indonesia*, 1st ed.; Routledge: London, UK, 2022. [[CrossRef](#)]
26. Xu, Z.; Bai, Y.; Zhang, B.; Li, D.; Fan, G. HAVEN: Hierarchical Cooperative Multi-Agent Reinforcement Learning with Dual Coordination Mechanism. *Proc. AAAI Conf. Artif. Intell.* **2023**, *37*, 11735–11743. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.