Improving performance in collaborative games through personality-based matchmaking*

Alejandro Villar[1,*,†], Carlos León[1,†]

[1]*Department of Software Engineering and Artificial Intelligence*
Computer Science Faculty
Universidad Complutense de Madrid
28040 Madrid, Spain

Abstract
Automatic player matchmaking in video games is usually focused on balancing players by taking into account their past performance. The objectives of video games in which the winning condition is clear and specifically defined by their design make this approach widely used and effective. However, this approach is not necessarily optimal when the main objective is not only winning but also obtaining other rewards like creative achievements, enjoyment, socialization or world construction. This paper describes a formal approach for automatic matchmaking that takes into account the players' personality traits. This form of matchmaking is oriented towards game genres in which competition is not necessarily the general dynamic. An implementation of the system, evaluation in a virtual escape room and results are provided. Preliminary results suggest that the approach is potentially effective.

Keywords
Videogame, Play-Style, Personality, Matchmaking, Elo Rating System, Escape Room

1. Introduction

Collaborative work is fundamental for accomplishing certain tasks. For improving collaborative performance, teams are usually created by taking into account the particularities of the task at hand. For example, they can be created based on the skills of the members creating relatively efficient teams by sacrificing a pleasant environment or based on diverse personalities so that members create good interpersonal relationships [1] and work in a more friendly way.

In online video games, team creation has typically sought to create these teams by focusing on the goal of defeating enemies with similar level. This is usually done without taking into account any form of relationship building, many times because the interaction of the group of players is limited to a single match or a short set of matches. The teams of players are created

✉ avillarrubio@ucm.es (A. Villar); cleon@ucm.es (C. León)
🌐 https://github.com/xBlacKnife (A. Villar); https://github.com/cleongh (C. León)
🆔 0000-0002-6271-2981 (A. Villar); 0000-0002-6768-1766 (C. León)

through matchmaking algorithms which use many game-specific features to group players together. For example, some games group mainly by playing skills, trying to create balanced teams where all players fulfil a specific role [2]. On the other hand, several studies are looking for ways to group players based on their personality with a social objective, either to increase entertainment [3], to improve interpersonal relationships so that players can enjoy playing with others [4] or to reduce the arguments between them [5].

The current research sets out to explore the possibility of grouping players in games in which relationships influenced by personality have an influential role in game dynamics. One game genre in which these complex team dynamics emerge is in virtual *escape room* games. Escape room games take place in rooms in which a group of people are locked in and, working creatively as a team, have to solve puzzles and achieve a final objective such as stopping a bomb or getting out of the room. This objective must be achieved within a certain time limit. If the escape room is not solved in time, it will be considered that team has not met the proposed objective [6]. Escape room dynamics are also used for educational purposes [7] and are beginning to be virtualised [8]. As such, they are a representative instance of collaborative games in which there are specific objectives beyond simply winning. Escape rooms are social games that require creativity and it is fundamental for the players to establish positive relationships with their team members. This work hypothesizes that the players' *play style* must be taken into account when creating the teams in order to maximize performance in the required creative tasks. This play style aggregates player personality and player roles when playing the game. Along this line, it is also considered important to consider the experience and skills of the team members to improve the performance. This paper introduces a formal matchmaking system for creating escape room teams. The system has been applied in an empirical experiment (section 4) and preliminary conclusions on the effectiveness on the approach have been studied (section 5).

2. Previous work

2.1. Player profile and play style

Classifying player types to offer better virtual experiences has been explored under different perspectives and goals. Many times, this classification has been carried out based on the user personality within the video game itself. Although several models were proposed before, Bartle's taxonomy is the most influential one [9]. This taxonomy describes 4 types of player: *killer*, *achiever*, *explorer* and *socializer*. This division has been criticized because players can many times be classified into more than one group. Bateman created the DGD1 model ("demographic game design") [10], which does not invalidate Bartle model but refines it, filling some gaps. The DGD1 model consists of 4 play styles:

- *Conqueror*: a independent player who will seek to win above all else. The conqueror does not exhibit excessively good communication and tries to win before time runs out and having better results than other teams. Surrender is typically not an option for this user and she will be constantly watching the time.
- *Manager*: this player is very important in any team due to his tendency to be the organizing leader. The manager controls the tasks to be performed by each user. See is a good

communicator and organizer. She tends to focus more on completing missions than on exploring, since her goal is to finish the game. The manager does not give up easily, although it is not something that affects this player considerably. She cares more about achieving everything she sets out to do in the game than finishing it quickly.

- *Wanderer*: a player who primarily seeks to explore and find as many items as possible in the game. In addition, finding hidden objects will be considered a reward for this player, regardless of whether or not it is useful to the game. Solving puzzles is not a main focus, but she enjoys enabling game options for the team through her exploration. Communication with this user is very important.
- *Participant*: the most communicative user. She likes to play with other players and keeps a good relation with the other players at all times. She does not necessarily exhibit any excellent performance in any aspect but simply enjoys the company even when the proposed objectives are not met.

Play styles are strongly influenced by existing psychological models of personality. For example, DGD1 is based on the Myers-Briggs Type Indicator [11], and associates each play style with its categories. This model, however, has been criticised due to the unclear psychological foundations, especially when compared to other models most widely accepted like the Five-Factor Model [12]. For this reason, ways to correlate the results of these two models has been studied [13]. Based on this, correlations between player personality and play style in video games have been reported [14]. These findings made it possible to use the player style gathered from the DGD1 model and create a system that matches them together according to their personalities, potentially improving performance.

2.2. Matchmaking and Elo rating system

Algorithms for grouping users in teams are usually based on variations of the *Elo* rating system. Created by Arpad Elo [15], this approach has been in use since the 1950's, mainly for chess matches. The Elo system rates players based on their achievements (number of matches won and lost), and makes it straightforward to match players with similar Elo ratings. Currently, variations of the Elo system are used in many areas like sports, dating apps and video games. The video game industry usually applies more sophisticated formulas. League of Legends (2009), Overwatch (2016) and World of Warcraft (2004) are examples of implementations of specific variations of the Elo system. However, these systems are usually focused on achieving a balanced confrontation between different users with special attention to the skill of these players.

Many studies have addressed the development of algorithms that improve the experience by taking into account several aspects related to the video game and its players beyond the wining criteria. This leads to a division between algorithms: those seeking to improve the quality of existing algorithms in order to obtain better results [16, 17], and those that are concerned with the social aspect [18], as the one presented in this paper.

3. Personality-based matchmaking system

As introduced in section 2.2, the Elo rating system assigns scores to players. In this study we propose a particular Elo-based model that takes personality into account. It has been designed with the main goal to improve performance in collaborative-creative scenarios, and meant to be tested in virtual escape rooms. The formal model still uses skill as part of the computation, but gives more relevance to the user's play style (influenced by her personality traits). In this research, the personality profiles have been created based on the DGD1 model (section 2.1).

The formula designed in this study for the Elo rating computes a value for each player. The rating will have a performance component and a personality component. The variables for the formula are described next:

- $S_{\{c,m,w,p\}}$ are the player ratings for the 4 DGD1 player styles. Each player has a particular rating for each style (*conqueror*, *manager*, *wanderer* and *participant*, respectively).
- $t_{(p)}$ is a score representing the skill in escape rooms of the player. The parameter p represents the number of times a player has played in a escape room. It was decided that once a player has some experience, the value of this component should be capped since player a few times is enough for having experienced the dynamics of a escape room. In the current experiment, the threshold has been set to 5.

$$t_{(p)} = \begin{cases} 5 & \text{if } p \geq 5 \\ p & \text{otherwise} \end{cases} \tag{1}$$

- k represents the user's knowledge about escape rooms. If the player knows what escape rooms are, she will get the maximum score of 1, and 0 otherwise. This variable represents that the player *knows* what a escape rooms is, independently of having played in one.

$$k = \begin{cases} 1 & \text{if played an escape room} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

- $r = 0.8$ is a constant parameterizing the weight to role-related data. The relevance of this value comes from the fact that we mainly want to study the results in relation to playing styles. It was observed that there is no significant performance difference in the teams created with values between 0.7 and 0.9, so it was decided to use the mid point between these two values. Using a value greater than 0.9 would imply that the skill would have no importance in the computation, and using a value lower than 0.7 would mean giving too much weight to the skill, something that would change the data, creating teams that would obtain results far removed from the role definitions.

The previous variables compose the rating formula used in the experiments (formula 3). Table 1 shows example values for the variables.

$$\text{player\_score} = \sum_{i \in \{c,m,w,p\}} r \times S_i + \frac{(1 - r) \times t(p)}{5} + 0.1k \tag{3}$$

Variable	User 1	User 2	User 3	User 4	User 5	User 6
Conqueror	8	4	8	6	2	2
Manager	10	8	10	2	4	10
Wanderer	6	8	4	10	8	6
Participant	6	10	8	8	6	4
Times played	20	2	0	3	12	2
Know what an escape room is	Yes	Yes	Yes	Yes	Yes	Yes
User score	32.0	29.6	28.0	27.2	24.0	23.2

Table 1
Example values for applying the matchmaking score formula to the 6 users who participated in the final experiment.

Based on the formula, a computational system creates all possible combinations of participants. After that, the matchmaking process sorts these teams by decreasing team score (i.e. the sum of the scores of the member players) and eliminates those in which there are repetitions of users. Therefore, the output of the algorithm is a list of the teams, sorted from highest to lowest score without repetitions of users. In this way, the first team on the list is expected to have the best results and the last one is expected to have the worst results.

4. Experiment

In order to test the proposed matchmaking system, an experiment was conducted. The experiment was divided in two stages. In the first stage, a questionnaire was distributed among the participants. In this questionnaire, participants were asked to answer some questions related to their experience with escape rooms. There exists no validated questionnaire for creating a DGD1 player profile so, for our experiment, the users were asked to evaluate their own perceived player profile according the DGD1 play styles. The users had to indicate whether or not they wanted to participate to the next stage.

A total of 35 people participated in this stage, and 26 of them declared to be willing to participate in the second stage. A total of 27 males, 6 females and 2 subjects who did not want to declare their gender. Age ranged from 19 to 39 years old (mean=20.69, stdev=3, 62). 97.1% knew what escape rooms are and 68.6% had played at least once, either real or virtual.

Each team was composed by two players. Thirteen teams were created using the matchmaking system for the second phase. Three of them were selected, corresponding to the highest, the lowest and the team with the mid-range score. These teams were automatically created based on the formal model described in section 3. Therefore, the experiment was carried out with teams labeled A, B and C with a score of 61.6, 55.2 and 47.2, respectively. The remaining subjects did not participate in the test.

Once the teams had been created by using the formal method described in section 3, the participants were asked to play the commercial video game *Escape Simulator* [19], depicted in figure 1. This video game contains several escape rooms. These rooms take approximately 15 minutes to solve each. The game includes escape rooms are designed to be played individually or in teams of 2-3 players. All games were recorded, along with the players' input, face and

Figure 1: Screenshot of the video game "Escape Simulator".

voice interactions, for further analysis.

The experiment lasted 50 minutes per team. At the beginning, each participant was given 15 minutes to modify the aspect of the character they were going to play with and pass the initial tutorial where they were shown the initial controls of the video game. In order to move on to the next stage, both players from each team had to finish the tutorial and automatically enter a small escape room to practice what they had just learned. This part lasted 10 minutes.

The remaining time was reserved entirely for playing the actual escape room. The escape room had a duration of 15 minutes. However, it allowed users to stay in the room for as long as they needed until they reached the objective or the team gives up. It should be noted that no team gave up, but only one team managed to finish the escape room within the maximum time.

5. Results and analysis

After running all the experiments and having collected all the videos, the content was analyzed. The analysis was carried out by tagging the recorded videos manually. During the gameplay, the users could interact with different types of elements and perform different actions with them. The analysis of the videos consisted of a direct counting of the actions carried out by each one of the users for monitoring the contributions they made in order to achieve victory. Certain items could be used in more than one situation, for example to solve two puzzles, so it was decided to count the actions given to each object according to the task it is performing. However, there are objects that do not have the same types of interactions, so it was decided to carry out an analysis of the various objects that make up an escape room. Thanks to this

analysis, it was possible to list the interactions that could be carried out with each of the objects and thus be able to look for relationships in a subsequent analysis.

Some interactions, as *look*, are complex due to the lack of an eye-tracking system makes it difficult to know whether the user is looking at an object or not. For this reason, the videos were modified by the adding dark borders leaving a clear spot in the centre of the screen. This space represents the focus of the view that users have in 3D gaming environments. This allows to understand which elements hold the player's visual attention [20] and thus avoid possible confusion (see figure 2). The main elements of the escape room are described next:

- *Pickable items*: those that could be stored in the player's inventory. These items are usually small and are the ones which most actions can be performed with. In the analysis, the actions were tagged as: look, pick, drop, observe, throw, bad idea, bad use, good idea and good use. A good or bad idea represents the player's success or failure, respectively, in identifying the proper use of the item. In the chosen escape room scenario, there were objects which the player could perform special actions with. For example, there were two books and a suitcase, which could be interacted with by opening and closing the item.
- *Movable Items*: elements present in the scenario and that could not be taken, but could be moved (pushed or dragged). The actions that could be performed are: look, move, bad idea, bad use, good idea, good use.
- *Puzzles*: indicate the progress of the team within the escape room, as the more puzzles they solve, the closer they are to victory. The puzzles are the tasks that the user must perform such as "enter the correct code" or collect the items needed to solve the puzzle, such as "a key". The actions that could be performed were more limited than with objects: bad idea, bad solution, good idea, good solution.
- *Hints*: players could get stuck. The game allowed to ask for hints to help them complete the tasks. In total there were 20 hints that are related to all the puzzles. *Picking* is the only action that could be performed with the hint.
- *Tokens*: collectible items that were hidden around the room. They were not necessary to achieve victory, but they were elements that are usually found in most video games and especially in those where you have to explore.

As it can be seen, there are actions that are repeated among the different kinds of objects, for instance *look* or *good idea*. *Look* is an action that is tagged as one type, regardless of the object. However, actions such as *good idea* depend on the object itself, as having a good idea in a pickable object is related to the use given to the object. Tagging an action as a *good idea* in a puzzle is linked to the way it can be solved. Therefore, even though they are the same actions, their definitions vary depending on the type of interaction that can be had with the particular item. This is discussed in section 6.

Once all the user data were collected, an analysis of the correlations between the players' play style and the actions performed during the game was carried out.

The large number of *pickable* items and their numerous actions mean that they were the elements of the escape room that allow the greatest extraction of data. In addition, it has been observed that the relevance of an object in solving the puzzles increases the amount of

Figure 2: Gameplay video, modified with borders representing player attention.

interactions the user has with it. For example, books are elements that were often used during the game and were the element that provides the best results (see Table 2 and figure 3).

When observing the interactions with the *books*, it stands out that the *managers* had a positive tendency to have good ideas and make good use of them, obtaining positive correlations of over 0.85 and $\rho < 0.04$. Also, it can be seen that the *wanderers* acted in the opposite way, as they had similar results but with negative correlations. Other positive results were gathered by those elements that were not useful for achieving victory. *Conquerors* tend to have a negative tendency to use them in general, obtaining correlations of less than -0.81 and $\rho < 0.05$. Results are less significant with the other item types, with *movable items* being the only ones that seem to have a bearing on player behaviour if more data were available. Puzzles, hints and tokens provided no significant conclusions.

It is interesting to mention that only one team managed to finish. Looking at the data provided by the users under a qualitative perspective, it can be seen that it is the only team that has a particular combination of roles. In particular, one of the users considers herself to be an administrator and the other one considers herself a wanderer. Both roles are very good communicators, and there is one main difference between them that helps to improve the organisation. The wanderer just wants to explore, and the administrator wants to solve puzzles. Also, both users see themselves as conquerors to some extent so neither will want to give up as they care about getting good results. This kind of additional qualitative analysis will make it possible to improve the algorithm by adding new concepts for grouping users, which seems to be relevant given the current method of team building.

		User 1	User 2	User 3	User 4	User 5	User 6
User score from questionnaire	**Know what ER is**	Yes (1)	Yes (1)	Yes (1)	Yes (1)	Yes (1)	Yes (1)
	Times played	1.0	0.4	0.0	0.6	1.0	0.4
	Conqueror	8.4	4.6	7.4	6.4	3.6	3.0
	Manager	10.0	7.8	9.0	3.2	5.2	9.4
	Wanderer	6.8	7.8	4.2	9.6	8.4	6.2
	Participant	6.8	9.4	7.4	8.0	6.8	4.6
	User total score	32.0	29.6	28.0	27.2	24.0	23.2
Interactions from game	**Look**	5	8	8	8	10	8
	Pick	5	8	7	2	7	8
	Drop	4	5	4	2	4	6
	Inspect	10	8	24	8	8	29
	Throw	0	0	0	0	0	1
	Bad idea	0	0	2	0	0	1
	Good idea	3	2	3	0	2	4
	Bad use	1	0	3	0	0	1
	Good use	2	1	3	0	1	2
	Open	7	6	11	0	4	13
	Close	0	1	0	0	0	0

Table 2
Information extracted from each of the participants. Two groups of columns can be observed: (1) the data from the questionnaire and (2) the count of the interactions carried out during the experience.

6. Discussion

This study has sought to provide a user grouping system that improves the final results based mainly on playing styles. As noted in section 2.1, there exist several taxonomies, of which the DGD1 model has been selected given the limitations of Bartle's model and the use of the MBTI as a psychological basis. Although an important part of future work would be to carry out a study using psychological models of personality, applying this taxonomy has shed some light on the applicability of player profiles to matchmaking and it has been possible to define behaviours for each of the styles of play within an escape room, something necessary in the design of the video game. This allows a more exhaustive analysis of user behaviour in the absence of expert validation.

One relevant limitation of the study is the extraction of the DGD1 scores from the players. Given the used values for the parameter r (the weight of personality in the matchmaking formula 3), the play styles are the most important data for the algorithm in the presented experiment. However, this information is provided directly by the user due to the lack of questionnaires to extract this data. This can lead to errors prior to the calculation of the algorithm as a user's answer may not be completely accurate.

Regarding the possibilities of data extraction, the use of a commercial video game makes it very difficult to modify the video game for data capture. This made it necessary to rely on the observation of the recorded games during the experiment sessions to count the actions of the participants. As the research line progresses, this can become a relevant limitation problem: as the observation of these videos is carried out directly by the authors, the process necessarily adds noise to the samples. Multiple annotators and expert validation is necessary before accepting these collected data as fully valid. Another limitation faced by this observational method of
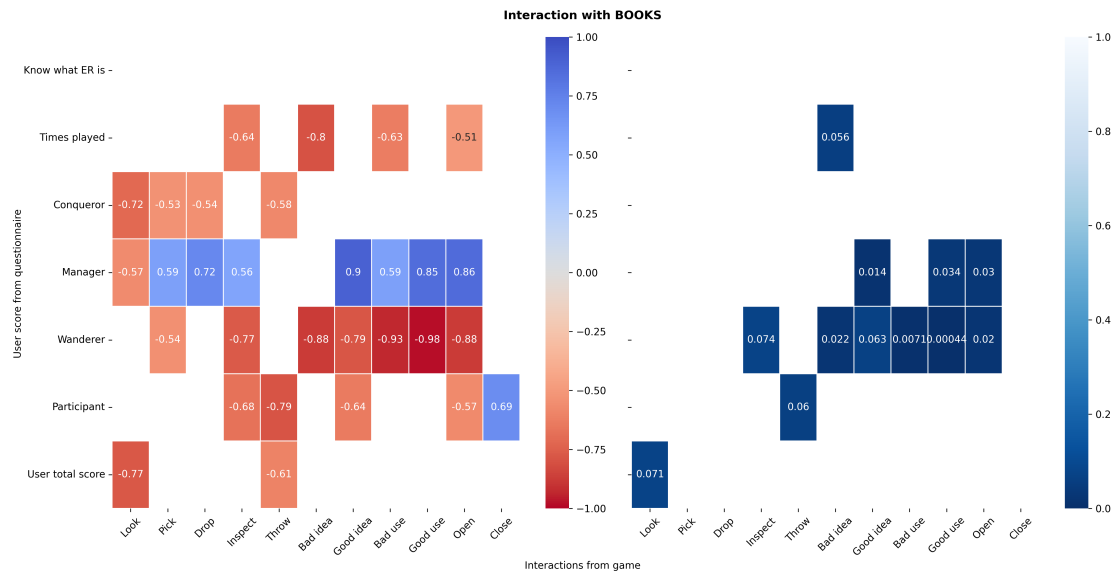
Figure 3: Analysis of results related to the *books*. The X axis corresponds to the actions that can be performed and the Y axis represents the Elo-based system results. Correlation table on the left, and p-values on the right.

analysis was to detect where on the screen the user was looking without the aid of eye-tracking systems. This was partially addressed by applying results showing the field of view of a user in 3D video games, which helped to minimise the viewing range and focus on the centre of the screen.

The approach of the proposed method seeks to apply the basic concepts of the defined roles and a small percentage of skill, thus making play styles the main components to be searched for relationships. The initial version is a first prototype that will be adapted in successive iterations based on what is analysed after the experiments. In addition, it does not currently perform any type of optimisation or automatic adaptation, something that is reserved for future versions, in the hope of improving efficiency and results. Along with this, expanding the scope of the experiment with more subjects and more scenarios is the main focus of future research.

7. Conclusions and future work

While many automatic matchmaking systems in video games rely on Elo-based approached focusing on the player's skill, several kinds of video games can be better experienced by matching players by also taking into account their personality profiles. This paper has summarized a research effort towards this goal: an original formula based on the DGD1 player profile has been designed, implemented and tested in an empirical experiment using a virtual escape room. The experiment was recorded and each match was annotated by hand. According to the obtained results, parts of the performance is affected by this process and, in several cases, it is relatively improved. The results are partially positive, but more data is required in order to reach a more

robust set of conclusions.

The current experiment was designed with quantitative data in mind, but the analysis of the videos has suggested that non-countable information can be of great value for identifying the potential benefits of the approach. Subsequent versions of the formal model will make use of what has been learned in this work. For instance, it has been observed that the interactions that players have with the elements of the environment are qualitatively different among player types. An expert analysis of the recorded data will be applied in order to obtain an informed qualitative analysis. In addition, the sampling volume with which these conclusions have been reached is limited, so more tests will have to be carried out to obtain a larger volume of data to consolidate the analysis. Different groups and different combinations of the players will be tested. The current results will be taken into account in future experiments that use this information in a more complex matchmaking system, possibly empirically identifying the best combination of player types.

Applying a virtual or augmented reality application in future experiments might help to break down the barrier between the real and virtual worlds. With this approach, the immersion of the player can be increased by making actions more realistic, even physically interacting with existing objects through haptic devices. This will provide a richer experience and a more detailed data acquisition system.

The third aspect to be improved in future experiments is the identification of player personality. The current approach is based on questionnaires for the DGD1 player profile. This has limitations regarding the application of the proposed solution in real video game experiences (as no one expects to have to fill a personality questionnaire before playing a video game). Automatic identification of personality traits, while more limited than validate psychological techniques, can potentially be a valid approach to this problem.

References

[1] R. D. Mann, A review of the relationships between personality and performance in small groups., Psychological bulletin 56 (1959) 241.

[2] M. Myślak, D. Deja, Developing game-structure sensitive matchmaking system for massive-multiplayer online games, in: International Conference on Social Informatics, Springer, 2014, pp. 200–208.

[3] G. Chan, A. Arya, A. Whitehead, Keeping players engaged in exergames: A personality matchmaking approach, in: Extended abstracts of the 2018 CHI conference on human factors in computing systems, 2018, pp. 1–6.

[4] J. Riegelsberger, S. Counts, S. D. Farnham, B. C. Philips, Personality matters: Incorporating detailed user attributes and preferences into the matchmaking process, in: 2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07), IEEE, 2007, pp. 87–87.

[5] S. Schuh, Psychological matchmaking for teams in games (2020).

[6] M. Wiemker, E. Elumir, A. Clare, Escape room games, Game based learning 55 (2015) 55–75.

[7] M. J. Vergne, J. D. Simmons, R. S. Bowen, Escape the lab: An interactive escape-room game as a laboratory experiment, Journal of Chemical Education 96 (2019) 985–991.

[8] J. M. Kutzin, J. E. Sanders, C. G. Strother, Transitioning escape rooms to a virtual environment, Simulation & Gaming 52 (2021) 796–806.

[9] R. Bartle, Hearts, clubs, diamonds, spades: Players who suit muds, Journal of MUD research 1 (1996) 19.

[10] C. Bateman, R. Boon, 21st century game design (game development series), 2005.

[11] I. B. Myers, The Myers-Briggs type indicator (1962).

[12] R. R. McCrae, P. T. Costa Jr, Empirical and theoretical status of the five-factor model of personality traits. (2008).

[13] R. J. Harvey, W. D. Murry, S. E. Markham, A "big five" scoring system for the myers-briggs type indicator, in: annual conference of the Society for Industrial and Organizational Psychology, Orlando, Citeseer, 1995.

[14] N. McMahon, P. Wyeth, D. Johnson, Personality and player types in fallout new vegas, in: Proceedings of the 4th International Conference on Fun and Games, 2012, pp. 113–116.

[15] A. E. Elo, The rating of chessplayers, past and present, BT Batsford Limited, 1978.

[16] M. Véron, O. Marin, S. Monnet, Matchmaking in multi-player on-line games: studying user traces to improve the user experience, in: Proceedings of Network and Operating System Support on Digital Audio and Video Workshop, 2014, pp. 7–12.

[17] O. Delalleau, E. Contal, E. Thibodeau-Laufer, R. C. Ferrari, Y. Bengio, F. Zhang, Beyond skill rating: Advanced matchmaking in ghost recon online, IEEE Transactions on Computational Intelligence and AI in Games 4 (2012) 167–177.

[18] E. Horton, D. Johnson, J. Mitchell, Finding and building connections: moving beyond skill-based matchmaking in videogames, in: Proceedings of the 28th Australian Conference on Computer-Human Interaction, 2016, pp. 656–658.

[19] Pine Studio, Escape Simulator, 2021. URL: https://escapesimulator.com.

[20] M. S. El-Nasr, S. Yan, Visual attention in 3d video games, in: Proceedings of the 2006 ACM SIGCHI international conference on Advances in computer entertainment technology, 2006, pp. 22–es.