An Exploration on Automating Player Personality Identification in Role Playing Games*

Alejandro Villar[1], Leonor Cuesta[2], Rodrigo M. Pérez[3], and Carlos León[4]

Computer Science Faculty and Institute of Knowledge Technology, Universidad Complutense de Madrid, Spain
{[1]alvill04,[2]leonorcu,[3]rodrip01,[4]cleon}@ucm.es

Abstract. When creating videogames, designers adjust the game characteristics to optimize the experience of the target players. This design process is usually manual, and the amount of insight that videogame designers have over the player is limited. Identifying the psychological profile of a videogame user can lead to specific adaptation of videogame aspects like narrative, length or challenge. Additionally, performing this identification automatically can both leverage the game experience and reduce the amount of work needed for customizing the players' experience. This research describes an emergent methodology to automatically create the personality profile of a player, and a prototype implementation. Additionally, a pilot study has been run, with preliminary but promising results.

Keywords: Videogame · Personality · Profiling · Non-intrusive · Interactivity

1 Introduction

Nowadays, videogames include several adaptive features that react to the player actions and strategies, and this produces more adapted experiences in which users can craft unique and comfortable experiences. These adaptations are usually based on ad-hoc approaches focusing on the task at hand, and react to the player performance and in-game decisions. Designers are constrained by the accessible data that a specific game can provide, usually restricted to variables related to gaming performance instead of what the player expects from the experience. This limits the designer's options when aiming at enabling a closer player-focused design.

Player-specific game adaptation has been tested in several games, but the implementations mostly rely on difficulty changes. This usually imply translating adaptability into adding or reducing features like the number of enemies, speed

* This work has been supported by the CANTOR project (PID2019-108927RB-I00) funded by the Spanish Ministry of Science and Innovation; and by the project FEI INVITAR-IA (FEI-EU-17-23) of the University Complutense of Madrid.

of events or time allowed until the user manages to complete the level. This approach can lead to successful gaming experiences. However, more complex games like roleplaying games, which have more complex dynamics, cannot rely on this simplification for a rich customized experience.

In particular, one of the most important aspects in roleplaying games is the player personality profile. Roleplaying games have a very important component of decision making, and the different players can take different options according to both their role and their own personality traits. Adapting roleplaying videogames to the particularities of these different types of personality can leverage the players' experience.

Against this background, this paper sets out as an exploratory effort to provide a method to automatically identify specific aspects of the players personality. The research is based on the hypothesis that it is possible to carry out automatic personality profiling in a non-intrusive way in a roleplaying videogame. In order to deploy the methodology and test the validity of the hypothesis, a study where the player profile was automatically acquired was run. The research used the Big Five personality model, and the overall identification of the personality traits was inferred from the player interaction with non-player characters in a virtual environment. The results were compared against the results of a running a validated Big Five questionnaire with the subjects.

The results (detailed in section 4) show that, while the exploratory character of this work does not yet permit to draw strong conclusions, the initial hypothesis of being able to identify user personality in an automatic non-intrusive way through videogames is plausible.

2 Previous Work

While not exclusively related to psychological models, the idea of automated user profiling has already been explored in several videogames. For instance, *Wii Fit* [16] seeks to profile the user physical condition split on speed, balance and stamina through daily tests, tracking the exercises realized in-game; *Dr. Kawashima's Brain Training* [14] tries to keep the brain healthy and young based on memory or concentration exercises, and adapts the difficulty to the predicted mental age as well as the current performance in the selected exercise, changing dynamically [17]. *Pokémon Mystery Dungeon* [15] assigns a Pokémon to the player according to a small personality test where answers add to a set of 14 personality types, assigning the Pokemon linked to the highest trait at the end. In all these cases, the system tries to be non-intrusive. In most cases, however, the level of profiling is relatively shallow and does not provide valuable data beyond that specific for the particular game.

Regarding psychological aspects, the particularities, types and features of personality have been studied from several perspectives. For instance, Raymond Cattell proposed a theory in which personality was divided into 16 factors [4]. Briggs and Myers [13] studied the innate characteristics of a person and how they establish basic personality characteristics. In particular, this project has applied the model proposed by Lewis Goldberg [7], the *Big Five* model. The Big Five model is broadly accepted and it is one of the most commonly used models of personality. The model divides human personality in five categories, namely *openness to experience* (O), *conscientiousness* (C), *extraversion* (E), *agreeableness* (A) and *neuroticism* (N). It is also known as the *OCEAN* model [8]. These five factors are briefly explained below:

Openness to experience describes how creative, adventurous and curious the person is. Those with high scores are more curious and open to new experiences. *Conscientiousness* refers to the efficiency and organization of a person. A low score indicates greater spontaneity. *Extraversion* is connected to how outgoing and energetic a person is. The higher the score, the more enthusiastic and outgoing a person is. *Agreeableness* describes how friendly and empathetic a person is. *Neuroticism* describes how sensitive and nervous people are. A low score indicates that the person is emotionally stable.

Based on the Big Five model, there exist several *Big Five Inventories* (BFI) with different amount of elements, but always with the same approach and objective. One of the most widely used and currently accepted questionnaire is the 44-item BFI created by Oliver P. John and Sanjay Srivastava [9]. The English version of the questionnaire and the necessary calculations can be found in the work cited above. A Spanish version is provided in another study where Oliver P. John himself was one of the researchers [3].

These models have been previously used in videogames. Chris Bateman and Richard Boon profiled four types of videogame players [1]. The results were used for a related research in which player performance in Fallout New Vegas was used to identify the category of the player after the gameplay [11]. More recent research has studied how to create an automatic player profile from a purely conversational adventure [10], using the dispositional approach psychological model [19]. Additionally, there are studies on the relation between play style in World of Warcraft and the results from the Big Five questionnaire [2]. To our best knowledge, no research has applied the Big Five model to a general RPG setting without any intermediate assumption on the game type.

3 A RPG-based Tutorial for Profile Acquisition

In order to test the hypothesis presented in section 1, a simple prototype of a roleplaying game has been implemented. The objective is to have subjects playing freely and create a user personality profile based on the initial stages of the gameplay and the decisions they make.

The videogame has been created with Minecraft [12].The use of Minecraft for research on videogames has been previously explored [6]. It offers a complete and finished work environment [20], so it is possible to bypass a relatively large proportion of the development time that would have been required for building an experience from scratch.It gives the possibility to play in different game modes, each with different functionalities, and to be able to choose the difficulty level. This allows to adapt the experience in a very simple way to the research's needs. Choosing Minecraft as development environment came with several upsides, with plenty of customization from the game itself, since it allows for game modes to just add content to a map (*creative mode*) or to experience a finished one without changing it (*adventure mode*), combined with difficulty settings that do not spawn enemies (*pacific*) creates the perfect medium to both create and conduct experiments where even newcomers can interact with the world freely. In the prototype, a simple town with several characters and a few simple quests were implemented. A screenshot of the prototype can be seen in Figure 1.

Fig. 1. Screenshot of the game prototype.

Among others, the prototype used two *mods*: *JourneyMap*, a customizable mini-map so that any tester can find their way through the level, and *Custom NPCs*, a tool designed to create RPGs based on answering dialogues and completing quests.

3.1 First Pilot Study

A complex, multiple-branch story was written for deploying the system. The story was composed as a set of dialogs. The Non-Player Characters (NPCs) were assigned dialogue trees that follow a common narrative (composed by 4 short stories). The different branches were written to cover the Big Five categories, adding or subtracting points from them as the user chooses an option, their personality mirroring their answer to specific conversational stimulus. These points have been assigned to each of the dialog interventions based on the research team criteria.

This first prototype sought to observe the initial response to the interactive experience. For the same reason the experiment was conducted in person and the researcher had to observe the player while he or she played without interfering with his decisions, and write down the options he chose for later analysis. 6

samples were obtained. Once the user had finished playing, a questionnaire with the original Spanish version of the Big Five Inventory was handed over [3]. The answers were recorded and the results were processed to calculate the values of each of the Big Five factors. At the end of the experiment, player dialogue decisions were processed to calculate the result of each factor according to the values assigned for each interaction. Since the values of each factor of the Big Five were obtained as a value between 0 and 100 according to the answers of the questionnaire, it was necessary to normalize the results of the gameplay to the same range. The maximum and minimum values that could be obtained by the user along the virtual experience can be seen in the Table 1. These values correspond to the points assigned to the dialog options. It must be noted that the neuroticism category has not been included since at this point in the experiment it was not intended to obtain its value through user interactions.

	Openess	**Conscientiousness**	**Extraversion**	**Agreeablenes**	**Neuroticism**
Max	14	7	21	12	-
Min	-10	-10	-5	-12	-

Table 1. Gameplay's maximum and minimum of each factor that could be obtained (First Prototype)

Once the results from the questionnaire and the calculation of the game interactions are obtained, a comparison was made to relate both values and get the error rates for each factor. The results were obtained by applying the arithmetic mean of the absolute values of each subject's error. The results can be observed in the Table 2.

	Openness	**Conscientiousness**	**Extraversion**	**Agreeableness**	**Neuroticism**
Mean	16.94	21.79	23.06	14.08	-
Std. Dev.	8.73	21.3	12.25	13.25	-

Table 2. Average differences and standard deviation between BFI Questionnaire and Gameplay.

It can be seen that the category with the highest error is the *Extraversion*. *Agreeableness* is the category where the results are most related and the data is concentrated around the average. Following these first results, it was hypothesized that the source of error for each category is due to the value ranges of the user's decision range are not equidistant. This is based on the fact that the best results obtained correspond to *Agreeableness*, this being the only factor that fulfills this equidistance. This means that the user start with 0 points and add or subtract with the chosen dialog option, where 12 points in the gameplay is equal to 100 in the questionnaire and -12 is 0. Section 3.2 describes the changes applied in the second prototype of the research.

3.2 Second Pilot Study

Based on the results and conclusions of the first iteration, a number of changes were applied to address the limitations. For this purpose a modification of the gameplay was made.

The dialogues interventions were revalued to offset category's maximums and minimums, as shown in Table 3. In addition, it was decided to include values to *Neuroticism* category. Additionally, only those branches of dialogues that the player accessed were taken into account for the total sum. Otherwise, if if all the possible lines were taken into account, the results would be normalized against non-valid data points.

	Openness	**Conscientiousness**	**Extraversion**	**Agreeableness**	**Neuroticism**
Max	47	18	45	54	34
Min	-44	-17	-45	-53	-36

Table 3. Maximum and minimum obtainable values in the gameplay (Second Prototype).

In addition, and in order to increase the access to potential players, for this second iteration, a telemetry system for automatic capture and storage of user data was implemented. Screen sharing systems were used for this purpose. Steam Remote [5] and Parsec [18] were used. In the sessions, the user had full control of the researcher's computer.

4 Overall Results and Analysis

In the second pilot study, a total of 42 experiments were carried out. Figure 2 shows the values obtained in this second pilot by each player for each one of the categories in the questionnaire and gameplay.

As it can be seen in Figure 2, the results yield a very high dispersion along the Y-axis in some factors such as *Conscientiousness*. However, other factors concentrate their results around smaller aspects such as *Openness*. For *Conscientiousness*, the dispersion cover almost all the possible range. While this is in line with the usual dispersion of the personality traits in the Big Five model, it contrast with the less dispersed factors. It is assumed that this is because of the characteristics of the videogame: most of the evaluation points have strong

Conscientiousness differences, and they define the type of game that the player chooses. As such, the stories are better suited for a diverse range. On the contrary, aspects less involved in the narrative decisions (like *Openness*) tend to show a more condensed set of answers.

In parallel with the analysis of the responses, the accuracy of the model (i.e. the similarity between the results of the gameplay and the results of the Big Five questionnaire) is shown in Figure 3, which shows the mean squared error (MSE) of the Big Five dimensions. In all cases except for *Conscientiousness* in story 4, the MSE value is below 20%. This leads to support the idea that, while the accuracy of the player profiling process is not perfect, the gameplay helps to approximate the Big Five values for players. The higher error that can be observed for *Conscientiousness* is assumed to be caused by the same characteristic discussed earlier, namely the wider range of captured values for *Conscientiousness* in the gameplay due to the narrative particularities of the environment.

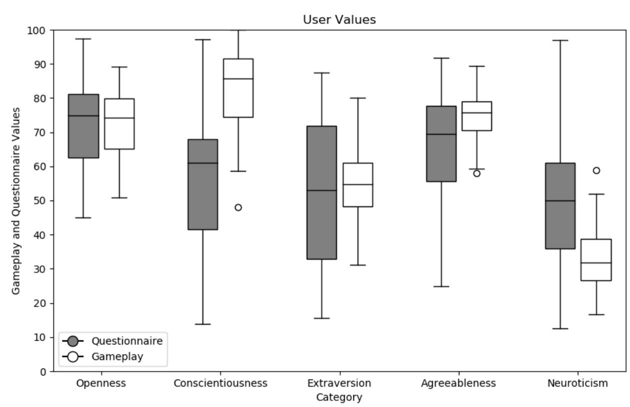

Fig. 2. Players' results in the Questionnaire and the Gameplay.

Figure 3 shows relatively positive results. However, it is concluded that a larger volume of samples of each of the factors is needed for more statistical validation. It can be seen that *Conscientiousness* is the factor with the worst results due to its scarce presence in the experience as opposed to *Openness* as indicated in the previous table.

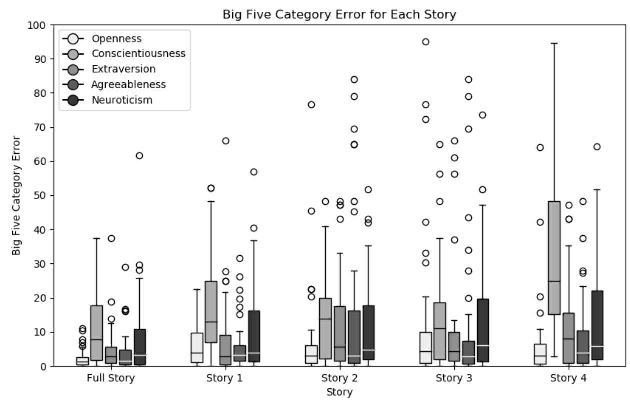

Fig. 3. Big Five category average error for each story.

5 Discussion

The current state of the project is still in its early stages and the general applicability of the results is yet to be fully determined. One of the most noteworthy points is the decision of a prospective assignment of personality values for the dialogue options, that is, why it was decided to establish some previous values for their subsequent balancing. In the presented prototype, the values were set by the authors, but this implies a certain level of noise in the results. Even with this limitation, the results are positive, but further refinements of the system require an additional methodology of expert-based assignment of the personality values.

Based on the hypothesis that personality can be analyzed in a non-intrusive way through videogames, an underlying hypothesis can be extracted, that there is a relationship between a player personality outside the videogame and the way they play within it. To validate it, it is possible to refer to the data extracted during the second iteration where a clear relationship between the results can be seen, with an error attributable to this lack of psychological basis in the assignment.

One of the most direct uses of the proposed method is, in parallel with the assignment of question values by experts, the use of machine learning algorithms to, based on the results of the Big Five Inventory questionnaire, assign automatically computed values. The relevance of this approach was tried in this research. In particular, the consistency of the results were tested by carrying out a Pearson correlation study between the results from the gameplay and the results from the Big Five Inventory questionnaire. The results were $\rho \approx 0.17$, $p \approx 0.35$. These obtained values lack statistical significance, but it is assumed that this is due to the relatively low amount of data for the correlation to be significant. The study was completed with Support Vector Machines which was trained with the options selected by the users in the gameplay session and their results in the Big Five questionnaire. SVM ($precision = 0.15$) along with Random Forest Regressors ($R^2 < 0$), confirm that the statistical significance of the data set is still too low due to the reduced amount of data points.

In short, it has been concluded that, at least while having such a small data set, it is not yet possible to automate the identification of the personality or to study the relationship between the two sets of responses through machine learning.

6 Conclusions and Future Work

The paper has reported on a emergent study on the possibility of automated player personality profiling in RPG videogames. A plausible methodology for automatically identification user personality in a non-invasive way through an interactive experience has also been proposed The results have been reasonably positive (but yet inconclusive) and they show that there is a possibility of finding a methodology with which to identify, with a reasonable error, the player's personality where the use of dialogues and their interactions seem to be of great importance. Still, the exploratory nature of the research leaves several ways of improvement as part of the future work.

The most straightforward improvement to the proposed methodology would be to gather more data and study the effect of applying machine learning techniques to compute the best fit value, for all the dimensions in the Big Five model, to the questions in the story (as introduced in Section 5). In any case, regarding the influence of the specific questions in the story, it is clear that the assigned values in the dialogue interactions must be validated by an expert because the actual values have been assigned by the researchers and there has not been any expert involved in the process. Related to this, checking the robustness of the results and the methodology requires to increase the amount of obtained samples.

Further research includes the need to test this methodology against other videogames and thus be able to verify that the hypothesis applies to different genres. It is believed that this kind of approach is not of application in general, since it would be hard to apply dialogue-based interaction in a car-racing game, for instance. However, we believe that the approach is not necessarily restricted to classic, adventure based roleplaying games. For instance, decisions based dynamics not specifically rendered as dialogues could be a source of information as well.

References

1. Bateman, C., Boon, R.: 21st Century Game Design (2006)
2. Bean, A., Groth-Marnat, G.: Video gamers and personality: A five-factor model to understand game playing style. Psychology of Popular Media Culture **5**(1), 27–38 (2016). https://doi.org/10.1037/ppm0000025, place: US Publisher: Educational Publishing Foundation

3. Benet-Martínez, V., John, O.P.: Los Cinco Grandes Across Cultures and Ethnic Groups: Multitrait Multimethod Analyses of the Big Five in Spanish and English (1998). https://doi.org/10.1037/0022-3514.75.3.729, https://psycnet.apa.org/record/1998-12057-011
4. Cattell, R.: Confirmation and clarification of primary personality factors (Sep 1947). https://doi.org/10.1007/BF02289253, https://link.springer.com/article/10.1007/BF02289253
5. Corporation, V.: Steam. https://store.steampowered.com/ (2003)
6. Frazier, S., Riedl, M.: Improving Deep Reinforcement Learning in Minecraft with Action Advice. arXiv:1908.01007 [cs, stat] (Aug 2019), http://arxiv.org/abs/1908.01007, arXiv: 1908.01007
7. Goldberg, L.R.: The Structure of Phenotypic Personality Traits (1993). https://doi.org/10.1037/0003-066X.48.1.26, https://psycnet.apa.org/record/1993-17546-001
8. Hodent, C.: The Gamer's Brain (2017). https://doi.org/10.1201/9781315154725
9. John, O.P., Srivastava, S.: Big Five Inventory (Bfi) (1999). https://doi.org/10.1525/fq.1998.51.4.04a00260
10. Karami, A.B., Encelle, B., Sehaba, K.: A Pilot Study on Estimating Players Dispositional Profiles from Game Traces Analysis, pp. 1101–1120 (jan 2020). https://doi.org/10.1007/978-3-030-29516-5_82
11. Mcmahon, N., Wyeth, P., Johnson, D.: Personality and player types in fallout new vegas (09 2012). https://doi.org/10.1145/2367616.2367632
12. Mojang: Minecraft. https://www.minecraft.net/en-us/ (2009)
13. Myers, I.B.: The Myers-Briggs Type Indicator: Manual (1962) (Apr 2014). https://doi.org/10.1037/14404-000, https://psycnet.apa.org/record/2013-29682-000
14. Nintendo: Brain Training. https://www.nintendo.co.uk/Games/Nintendo-Switch/Dr-Kawashima-s-Brain-Training-for-Nintendo-Switch-1656777.html (2005)
15. Nintendo: Pokémon Mystery Dungeon. https://mysterydungeon.pokemon.com/en-us/ (2005)
16. Nintendo: Wii Fit. https://www.nintendo.com/wiifit/launch/ (2007)
17. Nouchi, R., Taki, Y., Takeuchi, H., Hashizume, H., Akitsuki, Y., Shigemune, Y., Sekiguchi, A., Kotozaki, Y., Tsukiura, T., Yomogida, Y., Kawashima, R.: Brain Training Game Improves Executive Functions and Processing Speed in the Elderly: A Randomized Controlled Trial. PLOS ONE **7**(1), e29676 (Jan 2012). https://doi.org/10.1371/journal.pone.0029676, https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0029676, publisher: Public Library of Science
18. Parsec: Parsec
19. Staw, B.M., Bell, N.E., Clausen, J.A.: The Dispositional Approach To Job Attitudes: A Lifetime Longitudinal Test. Administrative Science Quarterly **31**(1), 56–77 (1986). https://doi.org/10.2307/2392766, https://www.jstor.org/stable/2392766, publisher: [Sage Publications, Inc., Johnson Graduate School of Management, Cornell University]
20. Steve Nebel, S.S., Rey, G.D.: Mining learning and crafting scientific experiments: A literature review on the use of Minecraft in education and research (apr 2016), https://psycnet.apa.org/record/2016-19735-025