Special Issue: New Approaches Toward Conceptualizing and Assessing Personality
Original Article

Playing With Networks

Using Video Games as a Psychological Assessment Tool

Giulio Ammannato[1] and Francesca Chiesi[2]

[1]Department of Statistics, Computer Science, Applications “G. Parenti” (DISIA), University of Florence, Italy
[2]Department of Neuroscience, Psychology, Drug, and Child’s Health (NEUROFARBA), Section of Psychology, University of Florence, Italy

Abstract: Video gaming behavior may offer information about the players and the widespread diffusion of this form of entertainment produces a staggering amount of data about gaming behaviors. The aim of the current study was to investigate the possibility to use the information about the way the player acts and reacts in a competitive video game to assess personality traits inside the HEXACO space. Deep learning was used to train deep neural networks that classified a sample of players (N = 41) with different personality traits by how they play in a Massive Online Battle Arena (MOBA) video game. Results suggested that the likelihood of correctly identifying the player’s trait level was above chance for five out of the six personality dimensions, but there is a medium to high margin of error in the classification. These findings provide interesting suggestions to set the premises for future studies to test the feasibility of this alternative assessment tool.

Keywords: personality, deep learning, video games, alternative assessment methods

Gaming, defined as “the action or practice of playing video games” (Oxford Dictionaries, 2018), is a widespread form of entertainment and there are approximately 2.3 billion gamers in the world (Statista, 2019). This large diffusion is accompanied by a few long lasting debates on the negative effects of gaming: one is the possible link between (violent) video games and aggression (e.g., Anderson et al., 2010; Furuya-Kanamori & Doi, 2016), and another one is about the possibility to develop a gaming disorder (e.g., Kuss & Griffiths, 2012; Van Rooij et al., 2018), recently added in both the ICD 11 (World Health Organization (World Health Organization [WHO], 2018) and the DSM-5 (American Psychiatric Association [APA], 2013). Nonetheless, there is also a vast literature about the positive effects of gaming (e.g., Granic, Lobel & Engels, 2014) and the use of video games as didactical, treatment and rehabilitative tools in programs aiming at improving learning as well as functional and cognitive functioning across various ages (e.g., Bul et al., 2016; Iten & Petko, 2016).

The idea behind this explorative study relies on two considerations. Firstly, the gaming behavior may offer information about the players. Accordingly, previous studies established a connection between personality characteristics and the style of play in World of Warcraft players (Bean & Groth-Marnat, 2014) and showed that gaming choices have good potential in predicting problematic and at-risk gambling behavior in adolescents (Donati et al., 2019). Secondly, the large playerbase produces a staggering amount of data that are shared with the players or used by the games’ developers for statistical purpose. For example, some games as StarCraft II (Blizzard Entertainment, 2010) and League of Legends (Riot Games, 2009) automatically register and save a player’s game, in the form of a replay. A game replay is a recording of a video game match with the player’s point of view (e.g., where he was looking, where he was clicking, what he was typing) and it can be viewed as a person’s real behavior albeit in virtual environments.

Starting from this premise, the aim of the current study was to investigate if it is possible to use the information contained in the replays (i.e., the way the player acts and reacts in a competitive video game) to assess personality traits. Indeed, it is possible to observe some psychological characteristics from the behavior in reality (e.g., a cautious person tends to stay wary and waiting for the other initiative, while a more reckless people tends to be inattentive of danger and acts impulsively, regardless the consequences). Assuming that gaming allows observation of real behaviors in a virtual space, it should be possible to assess a psychological characteristic by how a person plays (e.g., presumably a cautious player is less likely to take the initiative to engage in a fight, while a reckless player is expected to attack other players more often and/or more aggressively).

Specifically, we referred to the HEXACO model of personality proposed by Lee and Ashton (2004), which is a modified and extended version of the more classic five-factor model of personality, or the Big Five (Costa & McCrae, 2009). HEXACO is the acronym for six broad personality dimensions: Honesty-Humility, Emotionality, feXtraversion, Agreeableness, Conscientiousness, and

https://doi.org/10.1027/1015-5759/a000608

Openness to Experience. Between the Big Five and HEXACO models, there are some strict correspondences; that is, extraversion, conscientiousness, and openness to experience are basically equal, rather nuanced differences; that is, emotionality and agreeableness are variants of Big Five neuroticism and agreeableness (Ashton & Lee, 2007), and a striking extension; that is, a sixth basic personality factor, named honesty-humility (Lee & Ashton, 2004). Since gaming includes the pleasure of engaging in a challenge with other people and the thrill of taking part in situations with uncertain outcomes and/or danger, we were particularly interested in assessing the Honesty-Humility and Emotionality traits. The honesty-humility dimension has been defined as "the tendency to be fair and genuine in dealing with others" (Ashton & Lee, 2007, p. 156). As such, it is conceptualized as the basic personality factor that drives sincere, cooperative, non-exploitative behaviors. Emotionality includes fearfulness, anxiety, sensitivity to risks and harms. Therefore, the gaming behaviors should be the expression of the honesty-humility trait, which might affect the way people engage in a challenge with other people and the risk they take for attaining gains, and of the emotionality trait because fearfulness, sensitivity, and reaction to the risks and harms might influence the game action.

As stated before, gaming behaviors' information is enclosed in the replays that the video game recording device provides automatically. As such, much game data are tracked, logged, and stored, but not fully analyzed because the assessment process can be extremely complex and time demanding (Drachen et al., 2013). Indeed, when datasets become very large and complex, methods designed for large datasets must be employed. These methods are collectively referred to as data mining, which can be defined as the process to extract new aspects and patterns from a large data set (Feyyad, 1996). The application of data mining techniques, which include statistical models, mathematical algorithms, and machine learning methods, can be found in several domains. For example, educational data mining is an emerging interdisciplinary research area that explores data originating in an educational context to resolve educational research issues (Baker, 2014; Romero & Ventura, 2010), and it can be carried out on game data of individual players (Drachen et al., 2013).

Thus, we hypothesized that machine learning can be used to extract novel information hidden in game data. A machine learning system is trained and learns from data. In a supervised learning algorithm, both input data and a target output are fed to the system, which try to transform the input data into a meaningful representation, with the aid of a target and a feedback signal (Bishop, 2006; Chollet, 2017). After a successive approximation process, a model is developed and tested, and finally, if adequate, it is used to categorize new data. Deep learning does the same thing but it learns from successive layers of increasingly meaningful representations, stacking different layers of neural networks (Bengio et al., 2015; Schmidhuber, 2015). Therefore, the aim of the current study was to explore an alternative psychological assessment method based on video game replay analysis done with deep learning. Specifically, we wanted to explore whether it would it be possible to discern between players characterized by different personality traits inside the HEXACO space starting from their in-game actions.

Methods

Participants

A total of 41 participants took part in the study (95% male) with a mean age of 29.64 years (*SD* = 6.53). All players have played at "Heroes of the Storm" (Blizzard Entertainment, 2015a) in the last 6 months. Participants were from 20 different countries in the world: Italy (n = 8), United States (n = 8), Germany (n = 4), Canada (n = 3), Brazil (n = 2), Algeria (n = 2), Germany (n = 2), England (n = 1), Russia (n = 1), Romania (n = 1), Mexico (n = 1), Portugal (n = 1), Malaysia (n = 1), Estonia (n = 1), Finland (n = 1), China (n = 1), Czech Republic (n = 1), France (n = 1), Greece (n = 1).

Thirty-five participants were recruited online and six were recruited offline. The online recruitment was done via the official game forum and on Reddit, in the specific game subreddit. The offline participants were recruited through word of mouth among acquaintances.

Participants received an informative on the study and they were informed that participation was both voluntary and anonymous; assuring that their data privacy would be guaranteed. Before starting, a written consent was completed. This study has been approved by the local research ethics board.

Materials and Procedure

Among the wide array of game typologies, a game from the Massive Online Battle Arena (MOBA) type was chosen. MOBAs are particularly interesting in a psychological prospective because both individual and group dynamics are represented since players have to cooperate as a team and compete with another team. Among MOBAs, Heroes of the Storm (Blizzard Entertainment, 2015a) was selected because, besides the below described characteristics, it provides a tool, called "heroprotocol," to decode replay, which is freely available from the developers (Blizzard Entertainment, 2015b). Heroes of the Storm game replays were gathered both online and offline. Online participants were asked a minimum of 4 replays to a maximum of 15 replays. Offline participants contributed with all the replays of the games they played in the last year.

To assess the personality characteristics of the players we utilized the HEXACO. The HEXACO-60 (Ashton & Lee, 2009) consists of 60 items that measure six broad personality dimensions: honesty-humility, emotionality, extraversion, agreeableness, conscientiousness, and openness. This dataset of six factors was recovered in independent standard lexical studies involving different languages. Each factor is comprised of ten items and items are rated on a 5-point Likert scale (1 = *strongly disagree* to 5 = *strongly agree*).

Data were gathered online on an English survey hosted on a private domain using LimeSurvey, an open source survey software. The survey allowed collection of the HEXACO responses, the replays, and sociodemographic data.

Data Analysis

The first step in machine learning is to create representative input data. We used a custom script (see Electronic Supplementary Material, ESM 1) to organize the replays and calculate the HEXACO scores of all the participants. The raw HEXACO scores were then dichotomized in 0s and 1s because of the reduced number of participants for this pilot study that impedes accurate testing of more classes. Specifically, the scores above the normative mean (HEXACO, 2019) were recoded as 1s and the scores below to 0s. We also recoded the results in three categories: scores below −1 SD were coded as 0s, scores between −1 *SD* and +1 *SD* were coded as 1s, and scores above +1 *SD* were coded as 2s. These converted variables were then used as the targets of the deep neural network (DNN).

The replays were decoded using the “heroprotocol” (Blizzard Entertainment, 2015b) and a custom bash script to automatize the process (see ESM 2). Three files (game events, details, and message events) were obtained for each replay and were analyzed to understand the game structure. Each game of Heroes of the Storm is structured in game loops. There are approximately 1,000 game loops per minute and in each game loop there are 111 possible events/behaviors (e.g., click on the mouse, press a key, type a message, player deaths, hero/minion kills). In each game loop a player can perform zero or *n* behaviors, since it is possible to do more than one event in a single game loop. The replays were then split into 2.5 min slices (2,500 game loops) to increase the available data and to avoid cluttering of information. Then we recoded the data of all replays and initialized an all 0s matrix of [game loop × event] size (in this case: [2,500 × 111]) and proceeded to increase by 1 each event present in a single game loop for all game loops. We wrote a python script (see ESM 3) to automatize the process that outputs two files: the 3D input matrix and the associated targets used for the training of the models. We also considered checking if some events were not present in the matrix (i.e., always zero for all participants) and, if it was the case, the event columns were removed from the matrix.

To perform the analyses, we used the Keras Application Program Interface (API; Keras, 2019) with TensorFlow (TensorFlow, 2019) as a back-end and the Scikit-learn library (Pedregosa et al., 2011) in Jupyter Notebook (Jupyter, 2019).

As a preliminary step, we wanted to evaluate whether it would be possible to recognize each player by how they were playing. This is extremely important for subsequent analysis as, if the players are easily recognizable, it is necessary to have each player in a different dataset during the training of the DNN. For this step we considered using the Random Forest Classifier (Pedregosa et al., 2011), which tries to classify inputs in classes based on selected features (Scikit-learn, 2019a).

To test our research question, we implemented deep learning models. The first step was to create the datasets for the DNN training. Three datasets are necessary: A training dataset, a validation dataset, and a test dataset. The train dataset is used to train the DNN, the validation dataset is used at the end of each epoch (i.e., the number of times that the learning algorithm cycles through all the train data), to test the accuracy on data not used in the training part, and the test dataset is used at the end of the training to see how well the model performs on data never seen before. Each dataset needs to be balanced (e.g., roughly the same number of zeros and ones) and each player cannot be in more than one dataset to avoid pitfalls in the procedure. We used a custom algorithm based on Scikit Group Shuffle Split function (Scikit-learn, 2019b) to create our datasets. We opted for a 20% test and validation split. Specifically, 20% of the data were kept for the test dataset, 20% of the remaining 80% were used as the validation dataset, and the remaining data were our train dataset.

DNNs rely on functions, called layers. Multiple layers are stacked on each other until there is an exit layer that tries to predict the desired target. The layers also contain the network’s knowledge, in the form of weights. Different layers work better for different matrix formats and different types of data processing. In our case we used a recurrent layer, specifically Long Short-Term Memory (LSTM; Hochreiter & Schmidhuber, 1997) to keep track of time, and a variable number of dense layers, to reduce the data to a binary output. In a binary classification the network should end with a dense layer with one unit and a sigmoid activation.

Additionally, following Chollet (2017) and Tieleman and Hinton (2012), Root Mean Square Propagation (RMSProp; Tieleman & Hinton, 2012) was used as optimizer. An optimizer uses the loss gradient obtained by the loss function and utilizes it to update the parameter of the model. It implements a specific variant of stochastic gradient descent. The recommended loss functions for binary classification, the tasks we wanted to perform, is binary cross-entropy (Goodfellow et al., 2016). A loss function calculates a distance score between the true target (i.e., the expected output) and the prediction made by the network, showing numerically how well did the network perform for a specific set. The network tries to find the best data representation through a successive approximation process that best tunes the weights of the layers' connections of the neural networks. Specifically, after each cycle, four metrics are outputted, namely train accuracy, train loss, validation accuracy, and validation loss. Accuracy tests the precision of the models, the number of correct predictions over the total number of cases, both on the train data and on the validation data. Loss is a measure of how much information is lost, how many targets are not identified, and how far are the predictions from the real values, and it is calculated for both the train and validation datasets. We decided to end the training process after 50 epochs and the model with the lowest validation loss obtained during the training was saved as the best performing model. The chosen model was then verified on the test dataset, consisting of data never seen before during the training process, and again accuracy and loss metrics were obtained.

Because the model is trained on the same data for numerous cycles, one of the biggest problems in deep learning is over-fitting. Over-fitting happens when the model fits really well on training data and validation data but fails to generalize the acquired knowledge on the test data. There are various techniques to try to reduce over-fitting: to have a shallow network (i.e., not many layers), to gather more data, to use dropout layers (Hinton et al., 2016), and to add weight regularization to the model (Chollet, 2017). Applying these techniques, three models were identified to explore our research question (see ESM 4, ESM 5, ESM 6, and ESM 7 for more details).

Results

Through splitting each replay in 2.5 min sequences, we obtained 2,230 batches and there are approximately 1,000 game loops per minute in the matrix. In each game loop, there are 111 possible events, the resulting data matrix had the following dimensions [2,230 × 2,500 × 111]. Then we checked if any of the events were not present in the data and we found only 29 events (e.g., "pings" on the map or on other players, chat messages, number of mouse clicks, number of keys pressed) out of the possible 111. We proceeded to delete the unnecessary columns and we obtained the following matrix: [2,230 × 2,500 × 29].

We used these 29 features of online players with 10 or more replays as the input of the Random Forest Classifier and the algorithm was able to recognize the players 92% of the time (Figure 1). Because the players were easily identified by their playstyles, to avoid misleading results due to confounding variables it was necessary to have each player in a different dataset for the training of the DNN. Looking at the distribution of categorical data for each HEXACO dimension, we selected the most balanced groups (see Table 1). Subsequently, using the replays three balanced datasets were created for each HEXACO dimension (Table 2).

We trained three models for every single dimension of the HEXACO (see ESM 7 for more details) and we selected the best model based on the values obtained in the test accuracy (Table 3). Specifically, for the honesty-humility dimension the highest achieved accuracy was 58.52% (loss: .76). For the emotionality dimension the best test accuracy was 69.07% (loss: .81). For extraversion the accuracy was 82.19% (loss: .42). For agreeableness, conscientiousness, and openness to experience accuracy was 67.38% (loss: .64), 62.75% (loss: .75), and 72.55% (loss: .56), respectively.

Discussion

The aim of this explorative study was to evaluate the possibilities of using video games and deep learning models to assess HEXACO personality traits. Starting from matches of an online video game, namely Heroes of the Storm, we used the player's behaviors (i.e., real behaviors executed in a virtual space, to training deep leaning models). As such, we explored the possibility to assess psychological characteristics by how a person plays (i.e., the way the player acts and reacts in the video game). Given the competitive and cooperative nature of Heroes of the Storm, we opted for the HEXACO model of personality that, compared to the more traditional Big-Five personality model (Costa & McCrae, 2009), includes an additional basic personality dimension, named honesty-humility (Lee & Ashton, 2004) and a revised trait, called emotionality. Therefore, we expected that gaming behaviors might provide indicators of the former trait through the way people engage in the game challenge for attaining gains, and indicators of the latter trait by the reactions to the game risks and harms.

The obtained results look promising, because adopting a strict procedure to create the datasets and testing the model accuracy through several steps, we were able to train models that identify people with different levels of extraversion

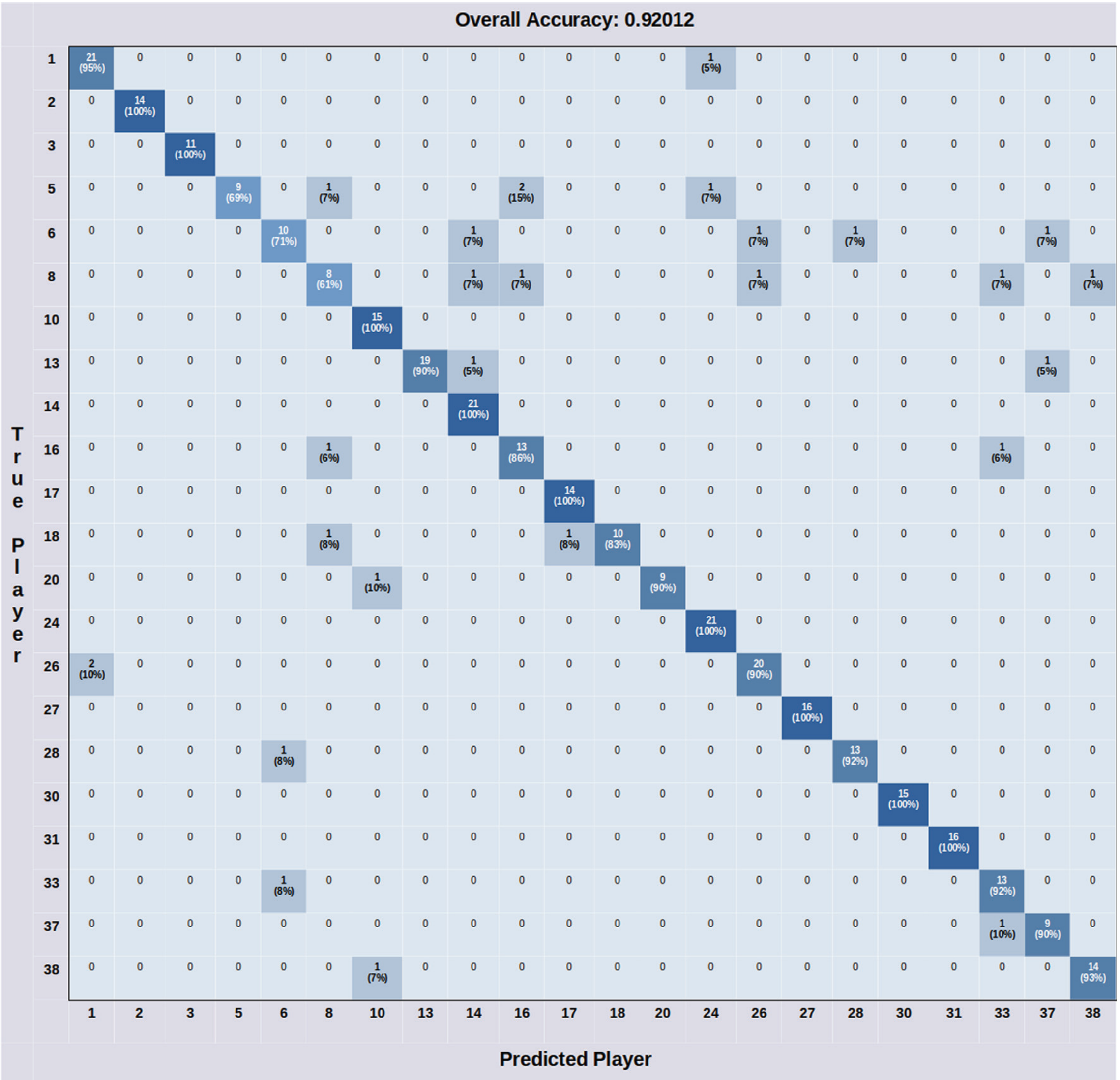

Figure 1. Prediction accuracy in identifying players with 10 or more replays (N = 38) based on their play-styles.

Table 1. HEXACO dimension means (*M*) and standard deviations (*SD*), and frequencies of the dichotomous and categorical recoding (N = 41)

	H	E	X	A	C	O
M	35.29	29.02	28.32	30.85	34.41	35.85
SD	5.93	7.40	7.46	5.99	5.90	6.85
Dichotomous recoding*						
0	14	28	34	**20**	**18**	**18**
1	27	13	7	**21**	**23**	**23**
Categorical recoding§						
0	3	**19**	**20**	6	7	5
1	**24**	**19**	**20**	29	30	29
2	**14**	3	1	6	4	7

Note. In bold the homogeneous in size subgroups selected for the analyses. *Below (0) and above (1) the *M* of the dimension. §Below −1 *SD* (0), between −1 *SD* and +1 *SD* (1), and above +1 *SD* (2) of the dimension. H = Honesty-Humility; E = Emotionality; X = eXtraversion; A = Agreeableness; C = Conscientiousness; O = Openness to Experience.

with a likelihood markedly above-chance. Also for emotionality, agreeableness, consciousness, and openness to experience the likelihood was above chance but there is a higher margin error in the classification of players. Finally, we were not able to obtain a high accuracy for honesty-humility because the probability to identify people with different characteristics on this trait is not far from chance. This is an unexpected result because our idea was that the in-game actions might highlight just that personality trait.

Overall, the current study offers an application of data mining techniques on video game data and it provides new and interesting information about individual players' psychological characteristics hidden in the game behavior. Nonetheless, given the exploratory nature of the current study, some limitations might have affected the obtained

Table 2. Raw data, balanced data, train dataset, validation dataset, and test dataset for each HEXACO dimension

	Raw data			Balanced data			Train dataset			Validation dataset			Test dataset		
	0		1	0		1	0		1	0		1	0		1
Dimension	*f*	%	*f*	*f*	%	*f*	*f*	%	*f*	*f*	%	*f*	*f*	%	*f*
H*	1,275	61	826	825	50	826	469	50	461	200	52	181	156	46	184
E§	1,110	54	956	1,110	54	956	690	54	589	197	53	174	223	54	193
X§	789	36	1,406	789	47	882	502	47	562	117	49	121	166	45	199
A#	1,111	50	1,119	1,111	50	1,119	692	50	690	214	50	211	205	48	218
C#	744	34	1,486	744	47	824	492	48	539	105	45	126	147	48	159
O#	1,156	52	1,074	1,156	52	1,074	725	52	660	243	51	234	188	51	180

Note. Each player belongs only to one group. *Between −1 *SD* and +1 *SD* (0), and above +1 *SD* (1) of the dimension. §Below −1 *SD* (0), between −1 *SD* and +1 *SD* (1) of the dimension. #Below (0) and above (1) the *M* of the dimension. H = Honesty-Humility; E = Emotionality; X = eXtraversion; A = Agreeableness; C = Conscientiousness; O = Openness to Experience.

Table 3. The three different steps in testing the model accuracy and loss of information for each HEXACO dimension

		Train		Validation		Test	
Dimension	Model	Accuracy (%)	Loss	Accuracy (%)	Loss	Accuracy (%)	Loss
H	2	69.46	.56	52.23	.64	58.52	.76
E	1	73.43	.55	72.58	.64	69.07	.63
X	1	71.72	.60	73.53	.60	82.19	.42
A	3	62.30	.63	67.53	.67	67.38	.64
C	3	79.73	.47	76.62	.53	62.75	.75
O	2	67.65	.59	58.28	.66	72.55	.56

Note. H = Honesty-Humility; E = Emotionality; X = eXtraversion; A = Agreeableness; C = Conscientiousness; O = Openness to Experience.

results and, at the same time, there are many open questions that arose from these preliminary findings. First of all, the player sample was small and thus, although we were able to obtain many replays, they all belong to these few people. Additionally, since some dimensions fitted better than others, it would be possible that the behavioral information provided by Heroes of the Storm may be more suitable for some psychological characteristics (as extraversion) but less fitted to assess other ones (as honesty-humility). Indeed, this game type offers lot of opportunities to interact with others that might be good indicators of the extraversion trait, while the game does not include the possibility to deceive others or to adopt manipulative behaviors that might be good indicators of the honesty-humility trait. Finally, there are several preliminary choices we need to do for training the models that might limit the impact and the generalizability of the current results, as the choice to dichotomize the HEXACO scores and using different categorization depending on the relative sample sizes.

The understanding of these limitations leads to an acknowledgement of the need for further studies. For future directions, it would be desirable to gather data from bigger samples in order to obtain higher variability in the personality scores as well as in the gaming behaviors. As a consequence, it would be possible to recode the data in more categories and to obtain bigger datasets to train the deep learning models. These are definitively relevant requisites to improve and refine the quality of the proposed assessment method. At the same time, the choice of both the personality model and video game might be rethought to find better option to test our research questions.

In conclusion, whereas the results of this preliminary study are only partially satisfactory, they provide interesting suggestions and set the premises for future studies to test the feasibility of this alternative assessment tool.

Electronic Supplementary Materials

The electronic supplementary material is available with the online version of the article at https://doi.org/10.1027/1015-5759/a000608

ESM 1. Python script to calculate the HEXACO scores for each participant and convert them in two (i0, 1) and three (0, 1, 2) categories using normative scores

ESM 2. Bash script to cycle troughs all the replays and unpack them using HeroProtocol https://github.com/Blizzard/heroprotocol

ESM 3. Python script to create the 3D input matrix and matched output labels

ESM 4. Network architecture of the first model to test our research question

ESM 5. Network architecture of the second model to test our research question
ESM 6. Network architecture of the third model to test our research question
ESM 7. Three models for every single dimension of the HEXACO

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. APA.

Anderson, C. A., Shibuya, A., Ihori, N., Swing, E. L., Bushman, B. J., Sakamoto, A., Rothstein, H. R., & Saleem, M. (2010). Violent video game effects on aggression, empathy, and prosocial behavior in Eastern and Western countries: A meta-analytic review. *Psychological Bulletin, 136*(2), 151–173. https://doi.org/10.1037/a0018251

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review, 11*, 150–166. https://doi.org/10.1177/1088868306294907

Ashton, M. C., & Lee, K. (2009). The HEXACO-60: A short measure of the major dimensions of personality. *Journal of Personality Assessment, 91*, 340–345. https://doi.org/10.1080/00223890902935878

Baker, R. S. (2014). Educational data mining: An advance for intelligent systems in education. *IEEE Intelligent Systems, 29*(3), 78–82. https://doi.org/10.1109/MIS.2014.42

Bean, A., & Groth-Marnat, G. (2014). Video gamers and personality: A five-factor model to understand game playing style. *Psychology of Popular Media Culture, 5*(1), 27–38. https://doi.org/10.1037/ppm0000025

Bengio, Y., LeCun, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*, 436–444. https://doi.org/10.1038/nature14539

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Blizzard Entertainment. (2010). *StarCraft II.* Blizzard Entertainment. https://www.blizzard.com/

Blizzard Entertainment. (2015a). *Heroes of the Storm*. Blizzard Entertainment. https://www.blizzard.com/

Blizzard Entertainment. (2015b). *HeroProtocol*. https://github.com/Blizzard/heroprotocol

Bul, K. C., Kato, P. M., Van der Oord, S., Danckaerts, M., Vreeke, L. J., Willems, A., van Oers, H. J., Van Den Heuvel, R., Birnie, D., Van Amelsvoort, T. A., Franken, I. H., & Maras, A. (2016). Behavioral outcome effects of serious gaming as an adjunct to treatment for children with attention-deficit/hyperactivity disorder: A randomized controlled trial. *Journal of Medical Internet Research, 18*(2), Article e26. https://doi.org/10.2196/jmir.5173

Chollet, F. (2017). *Deep learning with Python*. Manning Publications.

Costa, P. T., & McCrae, R. R. (2009). The five-factor model and the NEO inventories. In J. N. Butcher (Ed.), *Oxford handbook of personality assessment* (pp. 299–322). Oxford University Press.

Donati, M. A., Frosini, A., Izzo, V. A., & Primi, C. (2019). The effectiveness of the game of dice task in predicting at-risk and problem gambling among adolescents: The contribution of the neural networks. *Journal of Gambling Studies, 35*(1), 1–14. https://doi.org/10.1007/s10899-018-9796-5

Drachen, A., Thurau, C., Togelius, J., Yannakakis, G. N., & Bauckhage, C. (2013). Game data mining. In M. S. El-Nasr, A. Drachen, & A. Canossa (Eds.), *Game analytics* (pp. 205–253). Springer.

Feyyad, U. M. (1996). Data mining and knowledge discovery: Making sense out of data. *IEEE Expert, 11*(5), 20–25. https://doi.org/10.1109/64.539013

Furuya-Kanamori, L., & Doi, S. A. (2016). Angry birds, angry children, and angry meta-analysts: A reanalysis. *Perspectives on Psychological Science, 11*(3), 408–414. https://doi.org/10.1177/1745691616635599

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning* (Vol. 1). MIT Press.

Granic, I., Lobel, A., & Engels, R. C. (2014). The benefits of playing video games. *American Psychologist, 69*(1), 66–78. https://doi.org/10.1037/a0034857

HEXACO. (2019). *Descriptives*. http://hexaco.org/downloads/descriptives_60.pdf

Hinton, G. E., Krizhevsky, A., Sutskever, I., & Srivastva, N. (2016). *US Patent No. 9,406,017*. US Patent and Trademark Office.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Iten, N., & Petko, D. (2016). Learning with serious games: Is fun playing the game a predictor of learning success? *British Journal of Educational Technology, 47*(1), 151–163. https://doi.org/10.1111/bjet.12226

Jupyter. (2019). *Jupyter notebook*. https://jupyter.org/

Keras. (2019). *Keras*. https://keras.io/

Kuss, D. J., & Griffiths, M. D. (2012). Internet gaming addiction: A systematic review of empirical research. *International Journal of Mental Health and Addiction, 10*(2), 278–296. https://doi.org/10.3390/brainsci2030347

Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research, 39*(2), 329–358. https://doi.org/10.1207/s15327906mbr3902_8

LimeSurvey. (2019). *LimeSurvey*. https://www.limesurvey.org/

Oxford Dictionaries. (2018). *Gaming*. https://en.oxforddictionaries.com

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

Riot Games. (2009). *League of legends*. Hachette Book Group. https://euw.leagueoflegends.com/

Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40*(6), 601–618. https://doi.org/10.1109/TSMCC.2010.2053532

Statista. (2019). *Number of active video gamers worldwide from 2014 to 2021 (in millions)*. https://www.statista.com/statistics/748044/number-video-gamers-world/

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks, 61*, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003

Scikit-learn. (2019a). *Random forest classifier*. https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

Scikit-learn. (2019b). *Group Shuffle Split*. https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GroupShuffleSplit.html

TensorFlow. (2019). https://www.tensorflow.org/

Tieleman, T., & Hinton, G. (2012). Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural Networks for Machine Learning, 4*(2), 26–31.

Van Rooij, A. J., Ferguson, C. J., Colder Carras, M., Kardefelt-Winther, D., Shi, J., Aardataseth, E., & Deleuze, J. (2018). A weak scientific basis for gaming disorder: Let us err on the side

of caution. *Journal of Behavioral Addictions 7*(1), 1–9. https://doi.org/10.1556/2006.7.2018.19
World Health Organization. (2018). *Gaming disorder*. http://www.who.int/features/qa/gaming-disorder/en/

History
Received April 15, 2019
Revision received May 16, 2020
Accepted May 26, 2020
Published online January 19, 2021
EJPA Section / Category Methodological topics in assessment

Acknowledgement
The authors would like to thank Chloe Lau for her useful comments and suggestions, and for her help in editing the paper.

ORCID
Giulio Ammannato
https://orcid.org/0000-0002-1193-2626

Giulio Ammannato
University of Florence
Department of Statistics, Computer Science, Applications "G. Parenti" (DISIA)
Room 115
50134 Florence
Italy
giulio.ammannato@unifi.it

