PsyOps: Personality Assessment Through Gaming Behavior

Shoshannah Tekofsky
Tilburg University
PO Box 90153, 5000 LE
Tilburg, The Netherlands
S.Tekofsky@uvt.nl

Pieter Spronck
Tilburg University
PO Box 90153, 5000 LE
Tilburg, The Netherlands
P.Spronck@uvt.nl

Aske Plaat
Tilburg University
PO Box 90153, 5000 LE
Tilburg, The Netherlands
A.Plaat@uvt.nl

Jaap van den Herik
Tilburg University
PO Box 90153, 5000 LE
Tilburg, The Netherlands
H.J.vdnHerik@uvt.nl

Jan Broersen
Utrecht University
PO Box 80.089, 3508 TB
Utrecht, The Netherlands
J.M.Broersen@uu.nl

ABSTRACT

Traditional personality assessment methods are based on behavioral, observational, and self-report measures [8], each of which suffers from weaknesses that stem from ambiguity (behavioral measures), cost-payoff ratio (professional observation), and reliability (self-report). Assessment through video game play offers a way of quantifying behavior, automating observations, and side-stepping self-report. To determine whether video games are a valuable addition to the arsenal of personality assessment methods, we set out to answer the question: *Does the statistically trackable play style of a player significantly correlate to his personality?* To find the answer, we conducted a survey among Battlefield 3 players. Through the use of a promotional campaign, dubbed 'PsyOps', the response to the survey ran up to 13,376 individuals. Each participant was asked to fill out a 100-item IPIP (International Personality Item Pool) Big Five personality questionnaire, and requested for permission to draw their game statistics from a public database. All in all, 173 game variables, 100 personality scores, and 5 personality dimensions were correlated for the total sample, and 11 demographic subsamples. We found that play style and personality do correlate significantly, showing three key themes. (1) Conscientiousness is negatively correlated with speed of action. (2) The game variable Unlock Score per Second correlates most often and most strongly with personality, especially with Conscientiousness and Extraversion. (3) Work ethic correlates negatively with performance in the game. Apart from these three themes, subsamples differ in correlational patterns.

An additional result was found when performing a post-hoc analysis on age. Correlations between age and play style were greater than those between play style and personality. While themes (1), (2) and (3) showed effect sizes up to the 0.2 range, age offered effect sizes in the 0.3 range for game performance and game length preference, as well as a correlation of $r = -0.42$ with Unlock Score per Second. Age and personality correlate with a similar effect size as play style and personality. Therefore, age correlates strongly to play style, while age and play style offer complimentary correlations to personality.

Categories and Subject Descriptors

J.4 [**Social and Behavioral Sciences**]: Psychology

General Terms

Human Factors

Keywords

Age, Behavior, Gaming, Personality, Personality assessment, Play style

1. INTRODUCTION

Traditional personality assessment methods fall into the categories of behavioral, observational, and self-report measures [8]. In our research we explore the potential of adding another approach to this arsenal: personality assessment through video games. Video games combine the strengths of behavioral and observational measures, while side-stepping the reliability issues inherent in self-report. Additionally, video games offer a higher ecological validity than the traditional personality assessment methods. To determine if video games may contribute to personality assessment, we set out to answer the question: *Does the statistically trackable play style of a player significantly correlate to his[1] personality?* Previous research [14, 15] has yielded interesting results with small sample sizes. In order to validate these results with greater statistical power, we have chosen to focus on gathering a large data sample. Through the effective use of promotional material, 13,376 participants were recruited for our experiment. Each one filled out an IPIP Big Five

[1]For brevity, 'he' and 'his' are used where 'he/she' and 'his/her' are intended.

personality test[2] and provided access to their game statistics for the online multiplayer shooter Battlefield 3. In order to answer the research question, correlations between play style and personality were calculated. The results are also relevant to game developers who wish to personalize a game experience to an individual player.

This paper is organized as follows. First, related work is presented (Section 2), followed by more details on the experimental setup (Section 3). Then the results from our experiment are reported (Section 4). Subsequently, the results are discussed (Section 5) and the main conclusions reviewed (Section 6).

2. RELATED WORK

Research into personality assessment in video games is evaluated on three key requirements.

1. Play style should be meaningfully quantified.
2. Personality data should be meaningfully benchmarked.
3. Sufficient participants should be recruited to supply the data of requirements (1) and (2).

Requirement 1 ensures that underlying play style constructs (i.e. speed of play) are reflected in the data. Requirement 2 ensures that personality is accurately measured. Requirement 3 ensures that the results have a strong external validity and statistical power.

The purpose of our research is to meet all three requirements. To our knowledge this has not been done before. The following three research endeavors approach our aims most closely.

Requirements (1) and (2) were fulfilled in the work by Van Lankveld et al. [14, 15]. The first requirement was met by creating a custom module for the role-playing game Never Winter Nights that involved an extensive and meaningful quantification of play style. The second requirement was met by measuring the Extraversion dimension of the Big Five personality inventory [3]. The third requirement was not met due to a small sample size of 24 participants. Significant correlations were found between play style and Extraversion, with an effect size of $0.40 < r < 0.50$. Their follow-up study used the same measure of play style but included all five of the Big Five personality dimensions. The sample size was increased to 44 individuals. The research yielded significant correlations between the play style variables and the Big Five dimensions with effect sizes $0.10 < r < 0.50$. Meeting requirements (1) and (2), their findings showed a clear relationship between play style and personality. However, falling short on requirement (3), the findings lack statistical power due to the relatively small sample sizes.

Requirements (1) and (3) were fulfilled in the work by Drachen et al. [7]. The first requirement was met by extracting game statistics from proprietary and public databases. The second requirement was not met as no personality data was gathered. The third requirement was met by simply extracting the data of many individuals from the game statistics databases. In this manner 260,000 gamers were included in the sample for two games: the online role-playing game Tera, and the online shooter Battlefield Bad Company 2. Such a large sample could be achieved because participants were not individually approached for permission or additional data. With the use of clustering algorithms behavioral profiles were constructed that gave a meaningful description of different play styles. Meeting requirements (1) and (3), their findings show distinct play style profiles with high statistical power. However, falling short on requirement (2), these findings are not related to personality.

Requirements (2) and (3) were fulfilled in a meta-analysis by Barrick et al. [1]. The first requirement was not met because the research was conducted in the domain of job performance, but the relevant analogue data for that domain was analyzed. The second requirement was met by reviewing data on the Big Five personality dimensions. The third requirement was met by only including large participant databases in the meta-analysis. They found significant correlations between Big Five scores and job performance for five different occupations. Effect sizes were small with most correlations having $r < 0.1$. Conscientiousness was most predictive with $0.20 < r < 0.25$. In essence, this endeavor meets all the three requirements when adjusted for the domain of job performance, resulting in a high statistical power of the correlations between job performance and personality.

The research described in this paper combines the first two approaches described above to offer a correlational analysis of the link between play style and personality with sufficient statistical power to draw meaningful results. Fulfilling all three requirements, our research endeavors to bring large-scale personality assessment to the gaming domain in a similar way as has been done in the field of job performance.

3. EXPERIMENTAL SETUP

Our experiment consisted of a data collection (Section 3.1) and a data analysis (Section 3.2) phase. In order to answer our research question, the experimental design had to fulfill the three requirements mentioned in the previous section. They can be reiterated as (1) meaningfully quantified play style data, (2) benchmark personality data, and (3) large sample size. The following is a brief explanation on how the requirements were met.

Requirement (1) was met by selecting a game that offered a publicly accessible game statistics database: the online first-person shooter Battlefield 3. The data was meaningfully descriptive as it detailed play style in terms of interesting choices ranging from player specializations to player performance on various metrics (see Section 3.2 for more details). Additionally, the game is familiar to the first author.

Requirement (2) was met by measuring the Big Five personality dimensions. The NEO-PI-R used in the research by Lankveld et al. [14, 15] demanded a high time investment of the participants. This would have negatively impacted requirement (3) as it would have limited the sample to people willing to invest 45-60 minutes in a personality test. Therefore, we decided to use the 100-item IPIP version of the Big Five which required 5-20 minutes of the participant's time. The test consisted of 100 statements that a participant was asked to grade on a 5-point Likert scale, indicating how much he felt the statement described his personality. Scores on the statements were collated into the same five personality dimensions as the NEO-PI-R, with one exception. While the NEO-PI-R measures Openness, Conscientiousness, Extraversion, Agreeableness, and *Neuroticism* (OCEAN), the

[2]http://ipip.ori.org

IPIP measures the inverse of the last dimension and labels it *Emotional Stability* (OCEA-*ES*). The IPIP version is a validated instance of the Big Five Personality Inventory [9].

Requirement (3) was met by *marketing* the research toward the participant pool in such a manner as to create an almost viral enthusiasm to contribute. Our research project was dubbed 'PsyOps', and data collection performed through a dedicated website. Here, participants could find promotional material such as game-related art work, as well as a promotional trailer explaining the basics of the research initiative. We reached out to community websites to request them to feature PsyOps on their web pages and encourage their members to participate in the research project.

3.1 Data Collection

All data was automatically collected and stored via the PsyOps website. Data collection took place over a period of six weeks. During this time, participants could visit the website to submit their data. The data form contained six fields: player name, 100-item IPIP questionnaire, age, country of residence, gaming platform, and credits. The participant was asked to give permission for anonymous use of his game statistics, which were then automatically retrieved from a public database.[3] Player name was used as the key for game statistics retrieval. It is a unique identifier of a player account in Battlefield 3. Therefore, it was used to ensure all participants were unique individuals. The credits field was a tick box where participants indicated if they wished to have their player name listed on the credits page of the final research report. After submitting all their data, participants were forwarded to a page showing their Big Five scores and an overview of what the different dimensions entail.

3.2 Data Analysis

Data analysis progressed in five steps.

1. Creating and applying integrity filters to the data set.
2. Determining play style based on game statistics.
3. Determining personality based on IPIP scores.
4. Defining relevant (sub)samples.
5. Calculating correlations between play style and personality for all (sub)samples.

This section will detail the reasoning and processes underlying steps (1) to (4). The Results section will describe the outcome of executing steps (1) to (5).

(1) Filters were defined to maximize data integrity. Credits, IPIP, age, and game statistics data became the basis of four filters. First, the *credits filter* was based on the question if a participant wanted their player name to be mentioned in the credits of the research. The question was added to the data form as an integrity check. It was theorized that people who were more serious about filling in their data, would also be more likely to want their name associated with the results. Secondly, the *response set filter* was applied to remove participants who overused one response on the IPIP. This filter removed individuals with a biased response style ('response set') [4]. Thirdly, the *age filter* was applied to age, excluding individuals indicating an age below 12 or above 65. Age values could be selected from 1 to 99, and some people might enter the extreme or near-extreme values. To ensure the inclusion of the maximum number of participants, the limits were set to the onset of puberty (12) and end of working age (65). Fourthly the *player rank filter* excluded players with a player rank lower than 10. Ranks range from 0 to 145, with the last 100 ranks being honor ranks. The first 45 ranks gain the player access to additional items in the game that matter strategically. After 10 ranks, the player has unlocked a few items in his preferred class and gained a basic familiarity with the game.

(2) Play style was determined from a participant's game statistics. In order to understand the reasoning related to this process, a basic grasp of the game mechanics of Battlefield 3 is necessary. The following overview sketches the basic strategic options and objectives that players are offered in the game.

Battlefield 3 contains many strategical options. Five of the most prominent ones are briefly explained. First, a player selects one of three main game modes: Conquest, Rush, and Death Match. Each mode differs in game play, speed, and focus. However, all game modes may only be played as part of a team. Secondly, players select one of four roles to play in a match: Assault, Engineer, Support, and Recon. Thirdly, roles offer a limited and unique choice of support abilities (i.e., healing or reviving team mates, repairing vehicles, resupplying team mates, creating booby traps, or offering team mates reconnaissance services). Fourthly, roles offer a limited and unique choice of weapons. All weapons handle differently and are preferred for different play styles (i.e., close-range versus long-range). Fifthly, vehicles can be used as weapons or transport and are available to all players regardless of role.

Battlefield 3 traditionally sets players one single goal: to win the match. However, most players also strive to maximize kills, and acquire unlocks. Points are earned for reaching the goals, as well as for related subgoals such as playing objectives and providing support for the team. Self-sacrificing behavior such as giving support and staying behind to defend objectives, may help a team win, but damage someone's personal score. Additional points are awarded for kills based on team work (Savior Kills, Avenger Kills, Kill Assists, and Suppression Assists). Earning these points is conditional on two or more team members engaging one enemy. The intricacies of the game run even deeper, but this overview suffices to understand our research.

In order to determine the participant's play style, 826 game statistics were gathered. Domain knowledge was employed to combine and process the game statistics to reflect gaming behavior more accurately. The result was that 173 play style variables were defined over nine categories: Ribbon (7), Global (42), Equipment (8), Rank (1), Class (4), Score (19), Game Mode (10), Vehicle Category (7), Weapon (75)[4]. Different combinations of variables describe play style characteristics. We present three examples: tendencies toward team work (i.e., Ace Squad Ribbons, Wins per Loss), focus on kill efficiency (i.e., Kills per Death, Nemesis Kills), and preference for long versus short games (i.e., Play Time per Round, Conquest Rounds per Round).

(3) Personality was determined from the 100-item

[3] http://bf3stats.com/

[4] An overview of all the variables and their definitions are presented on the research website: http://www.psyopsresearch.com

IPIP questionnaire. The IPIP data consisted of 100 variables with a value of $[1, 5]$, where 1 denoted ”Very Inaccurate” and 5 denoted ”Very Accurate.” The values are attained through self-report. They reflect how much a participant identifies with a statement, such as ”I love children.” The Big Five scores were calculated from the IPIP data by combining the values on the statements that related to a particular dimension. The result was a value of [20,100] on each of the five personality dimensions.

(4) Subsamples were defined on the demographic variables 'gaming platform' and 'country of residence'. Gaming platform had three possible values: PC, Xbox 360, and Playstation 3. The relevance of gaming platform is threefold: (1) Platform preference might contain an inherent sample bias. (2) The interface is different between PC and the two consoles, and slightly different between the two consoles. (3) PC supports larger maps and higher server capacities than the two consoles. The relevance of country of residence was used to create a distinction between native and non-native English speakers, because the IPIP questionnaire was only administered in English. Participants were considered native English speakers if their country of residence was predominantly (>75%) English-speaking. Initially we had intended to create subsamples based on age as well. Through post-hoc analysis we found this was not meaningful, but correlations with play style and personality were interesting. For that reason, Section 4.6 was added to the Results to report on these additional findings. Thus, four partitions of the sample were made, resulting in 12 different (sub)samples: total sample (1), partition on gaming platform (3), partition on native English speakers (2), partition on gaming platform and native English speakers (6).

Overall, 173 game variables, 100 personality statements and 5 personality dimensions were correlated for the 12 (sub)-samples. Additionally, age was correlated with both play style and personality for the total sample. Correlations were determined by means of the Pearson's Correlation Coefficients (r). Correlations were considered significant at $\alpha < 0.05$.

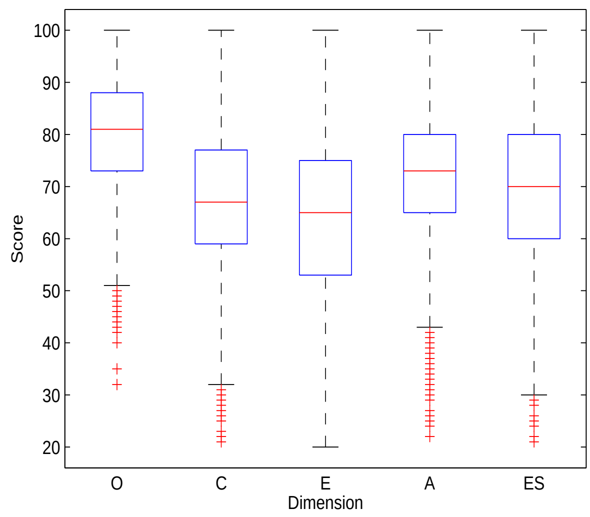
4. RESULTS

This section will detail the outcome of executing the five steps described in the Data Analysis section. These can be identified by the key words 'Filters' (Section 4.1), 'Play Style' (Section 4.2), 'Personality' (Section 4.3), 'Sample Partitioning' (Section 4.4), and 'Correlations' (Section 4.5). Section 4.6 will discuss the results of a post-hoc analysis on age, and Section 4.7 will briefly summarize our findings.

4.1 Filters

In order to maximize the integrity of our results, the data set was filtered. The final data set contained data from 13,376 participants. During the data collection phase, the third-party game statistics database was restructured to accommodate an upcoming expansion of the game. The restructuring process shifted the format of the collected data so only the first 9368 submissions were usable. The result of applying the four filters mentioned in the previous section were as follows. The credits filter excluded 2584 participants; the response set filter excluded 501 participants; the age filter excluded 31 participants; the player rank filter excluded 85 players. In total, 2995 entries were excluded, leaving 6373 participants in the sample (206 participants were excluded by more than one filter).

Figure 1: IPIP Big 5 Distribution for Total Sample



4.2 Play Style

Core game performance statistics were reviewed to gain an idea of what the dominant play style in the sample was. Overall it was found that mean scores on core performance metrics were well above the norm of the populace, while high standard deviations indicate that there was a wide range in performance within our sample. Such findings are not due to the filtering outlined in the previous section, as results were similar for the unfiltered sample. High performance metrics indicate that our sample consisted in large part of expert players.

4.3 Personality

The distribution of Big Five scores are shown in Figure 1. All dimensions correlate significantly with each other in a positive direction, with an effect size ranging from 0.17 between Openness and Emotional Stability, to 0.44 for Agreeableness and Extraversion. Overall, players in our sample report themselves to score high on Openness, and roughly middling on Conscientiousness, Extraversion, Agreeableness, and Emotional Stability. Additionally, someone scoring high on one dimension is likely to score high on all other dimensions. The pattern is not due to a biased response style as the 'response set' filter was applied to shield against this.

4.4 Sample Partitioning

The total sample was partitioned on native English speakers and gaming platform (individually and combined), resulting in 11 subsamples. The distribution of the partitioning variables was as follows. The native English speaker distribution is such that about 69% of participants were classified as native English speakers, while 31% were classified as non-native English speakers. The platform distribution is about 40% on PC, 28% on Xbox 360, and 32% on Playstation 3. The sample sizes that resulted from the partitioning are shown in Table 1 (denoted N). The table is further explained in the next section.

4.5 Correlations

Correlations between play style and personality were calculated for the total sample and each of the subsamples. For the total sample the result was 311 significant correlations between the 173 play style variables and the Big Five dimensions, and 4442 significant correlations between the 173 play style variables and the 100 IPIP scores. Each of the 11 subsamples offered a correlational frequency in the same order of magnitude. Therefore, there are too many correlations to report on individually. Instead, we present an overview in three parts: (1) correlational frequencies across the (sub)samples, (2) details on the correlations in the total sample, and (3) correlational themes over the (sub)samples.

(1) Correlational frequencies across the (sub)samples are shown in Table 1. Rows represent the Big Five dimensions. Columns represent play style variables per sample. The abbreviations 'E' and 'NE' stand for "native English speakers" and 'non-native English speakers', respectively. The gaming platforms PC, Xbox 360 and Playstation 3 are abbreviated 'PC', 'XB', and 'PS'. The number of significant correlations per Big Five dimension are shown per subsample in the column labeled 'sig.' Generally, most correlations have an effect size of $r < 0.10$. The column labeled '↑' lists the number of significant correlations for which the effect size is $r > 0.10$. Table 1 shows that correlations between play style and Big Five dimensions are meaningfully different between (sub)samples. The same is true for correlations between play style and IPIP scores. However, the correlations with IPIP scores are too extensive to list in this paper.

(2) Details on the correlations in the total sample are as follows. All dimensions (O, C, E, A, ES) correlate significantly with 59, 76, 65, 60, and 51 of the 173 game variables, respectively. Effect sizes are generally small with only four correlations with $r > 0.1$ (See Table 1). Three correlations are for Conscientiousness with the variables Deaths Per Second (-0.10), Unlock Score Per Second (-0.12), and Conquest Rounds Per Round (0.11). One correlation is for Extraversion and Unlock Score Per Second (-0.11). The same calculations were done between the 100 IPIP scores and the 173 play style variables which yielded 4442 significant correlations. Almost all correlations have a small effect size of $r < 0.1$, except for the 17 correlations shown in Table 2 ($0.10 < r < 0.15$). They sketch a general pattern of higher performance in the game being positively associated with statements such as "I shirk my duties" and "I find it difficult to get down to work", and negatively correlated with the statement "I love children."

(3) Correlational themes were determined over the total sample and the 11 subsamples. The result is three themes, which were drawn from the correlations between play style and both Big Five dimensions and IPIP scores. Subsamples show different patterns in the details of the themes. The differences do not show a clear character.

1. **Conscientiousness & Speed -** Of the Big Five dimensions Conscientiousness offers the most and strongest correlations with play style. Play style variables describing certain actions per time unit have negative correlations with this dimension, pointing toward a slower play style. A slow play style is not necessarily indicative of lower performance, as Deaths per Second is also negatively correlated with Conscientiousness. Noticeable positive correlations exist with spending one's time on "slow" activities such as preferring the game mode Conquest, spending more time per round, and preferring the relatively slow tank to other vehicles.

2. **Unlock Score & Personality -** Of all the game variables, Unlock Score per Second is generally the most strongly and consistently predictive of personality. Unlock Score is earned by achieving target scores with different classes, weapons, and vehicles. The only way to maximize it is by extensively using all assets offered in the game, effectively forcing a player to vary his play style. Patterns of correlation per subsample are different, but the general theme of the total sample remains present: Unlock score correlates negatively with Conscientiousness and Extraversion.

3. **Work Ethic & Performance -** IPIP statements relating to work ethic correlate negatively with game performance. Examples of such statements are "I shirk my duties" and "I find it difficult to get down to work." Statements such as "I do things in a half-way manner" show a mixed relationship with game performance, depending on what subsample is analyzed. As such, the key factor here is the ability or willingness to attend to one's responsibilities, and not the thoroughness or dedication someone may show once they do attend to their work. Additionally, there is no significant correlation between total play time and game performance. Therefore, it is the work ethic itself that matters in this correlation, and not an increased time investment in the game.

4.6 Age

Age data was collected (Figure 2) in order to partition the data set, but it did not yield interesting results. However, through post-hoc analysis it was found that age correlates with both play style and personality, showing high effect sizes. Age correlations are not related to our research question, but will be briefly discussed for the total sample due to the strength of our findings. When discussing correlations among the three data sets, the correlational pairs will be referred to as 'age and (&) personality', 'age and (&) play style', and 'play style and (&) personality.'

Age & Personality - Age correlates significantly with all Big Five dimensions with a strength of 0.11, 0.20, 0.13, 0.08, and 0.07, respectively. Our findings are in line with previous cross-cultural research [11, 5] showing that Conscientiousness scores are higher for individuals in middle age than those younger than that. Looking at correlations between age and the IPIP statements, the greatest effect size is $0.20 < r < 0.25$ for two statements: "I know how to captivate people" and "I find it difficult to get down to work."

Age & Play Style - Age correlates significantly with 151 of the 173 play style variables. Skimming off the strongest of the correlations ($r > 0.30$), age is found to correlate negatively with speed of play, correlate positively with length of play, and correlate negatively with game performance. Considering the high effect sizes, the individual game variables will be briefly mentioned with the effect sizes in parentheses: Unlock Score per Second (-0.42), Kills per Second (-0.35), Head Shots per Second (-0.35), Avenger Kills per Second

Table 1: **Big Five Correlations Total Sample**

N	TOTAL 6373		E 4373		NE 2000		PC 2515		XB 1791		PS 2067		E PC 1338		E XB 1540		E PS 1495		NE PC 1177		NE XB 251		NE PS 572	
	sig.	↑	sig.	↑	sig.	↑	sig.	↑	sig.	↑	sig.	↑	sig.	↑	sig.	↑	sig.	↑	sig.	↑	sig.	↑	sig.	↑
O	59	0	55	0	17	0	15	0	24	1	57	5	25	0	18	0	44	6	11	1	7	7	20	7
C	76	3	71	4	29	1	51	4	41	4	60	11	39	5	39	6	64	17	34	6	24	24	10	8
E	65	1	56	1	30	1	32	0	46	1	48	1	28	1	36	1	41	2	20	2	9	9	11	7
A	60	0	51	0	30	0	22	0	34	0	47	0	7	0	36	0	43	0	23	2	12	12	20	11
ES	51	0	43	0	18	0	26	1	35	0	25	0	19	1	26	0	20	0	6	0	16	16	10	3

Table 2: IPIP Correlations Total Sample

Statement	Game Variable	Power
I love children	KillsPerSecond	-0.11
	HitsPerSecond	-0.10
	HeadShotsPerSecond	-0.11
	UnlockScorePerSecond	-0.14
I am skilled in handling social situations	UnlockScorePerSecond	-0.10
I do things in a half-way manner	ConquestPerRound	-0.10
	UnlockScorePerSecond	0.11
I shirk my duties	KillsPerSecond	0.11
	HitsPerSecond	0.10
	AvengerKillsPerSecond	0.10
	ReconScorePerReconTime	0.10
I find it difficult to get down to work	HitsPerSecond	0.10
	UnlockScorePerSecond	0.12
	ReconScorePerReconTime	0.11
I am quiet around strangers	UnlockScorePerSecond	0.10
I find it difficult to approach others	UnlockScorePerSecond	0.12
I wait for others to lead the way	UnlockScorePerSecond	0.13

(-0.32), Play Time Per Round (0.31), and Conquest Round per Round (0.31).

Overall, age and play style correlate more strongly than play style and personality. Age and personality correlate with a similar effect size as play style and personality. Correlational patterns are different, making age and play style complementary in explaining variance in personality. Figure 3 schematically reflects our additional finding. Interpreting the specific correlations, we found that older players are more conscientious, less experimental and slower than younger players.

4.7 Summary

Play style correlates significantly with personality, with many correlations reaching an effect size of $r > 0.1$. Subsamples based on gaming platform and native English speakers differ meaningfully in the details of the correlational patterns. However, over all (sub)samples three correlational themes can be distinguished: (1) Conscientiousness negatively correlates with speed of action. (2) Unlock Score per Second correlates negatively with Conscientiousness and Extraversion. (3) Work ethic correlates negatively with game performance. An additional result was found during a post-hoc analysis on age. Play style and age correlate *more strongly* than play style and personality, while play style and age are *complementary* in explaining the variance in personality. Older players are more conscientious, less experimental and slower than younger players.

Figure 2: Age Distribution for Total Sample

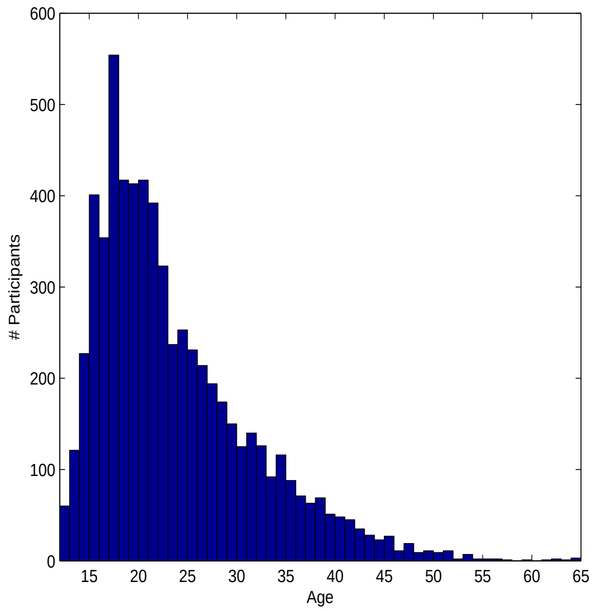
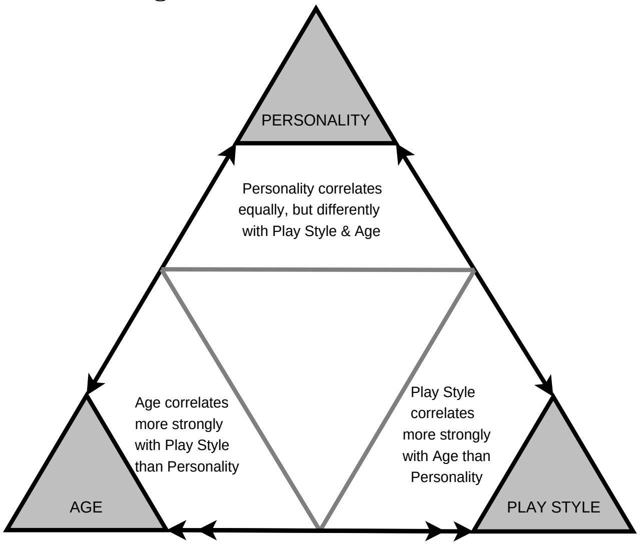


Figure 3: Triad of Correlations



5. DISCUSSION

In this section four topics will be discussed. First and foremost, the relevance of the effect sizes reported in the results section are examined (Section 5.1). Secondly, limits on data quantity are discussed (Section 5.2). Thirdly, the cause and effect of our sample bias is reviewed (Section 5.3). Fourthly, directions for future work are suggested (Section 5.4).

5.1 Relevance of Effect Sizes

Based on a meta-analysis of correlational research by Meyer et al. [12], we conclude that the effect sizes of our findings are of a relevant magnitude. In terms of Cohen's classification of effect sizes [2], most of our findings would be deemed 'trivial'. However, Meyer et al. argue against this by comparing the effect sizes found in psychological research to those found in the medical field. Examples from the medical field are the correlation between aspirin and reduced risk of death by heart attack at $r = 0.02$, and the correlation between chemotherapy and surviving breast cancer at $r = 0.03$. These findings are considered relevant in the medical field. Our findings exceed these effect sizes.

Additionally, established personality assessment tools like the MMPI (Minnesota Multiphasic Personality Inventory), Beck's Hopelessness Scale and the Big Five Extraversion dimension also show small effect sizes when correlated with relevant behaviors. For example, MMPI scores and subsequent prison misconduct correlate at $r = 0.07$, Beck Hopelessness Scale scores and subsequent suicide correlate at $r = 0.07$, and Extraversion test scores and success in sales correlate at $r = 0.08$.

Our findings show effect sizes that are equal to or greater than the effect sizes found in the medical field or for other personality assessment methods. Therefore, we conclude that the effect sizes of our findings are of a relevant magnitude. Meyer et al. also point out that higher effect sizes are often found when experiments are conducted with small samples. Such experiments have lower statistical power, while our research offers a strong external validity due to the large sample size.

5.2 Limits on Data Quantity

More data could have been gathered from the participants at the risk of reducing the sample size. There is a fine line to tread between generating enthusiasm in the potential sample and the investment in the research that may be expected in return. We expect that if we had asked for more data, our sample would have been much smaller. The visitor statistics from the PsyOps website illustrate this point: The front page gained about 30,000 unique visits. It contained the promotional material to enthuse prospective participants. The questionnaire page received 20,000 unique hits. Subsequently, little over 13,000 participants completely filled out the data form and submitted their results. Of the 17,000 potential participants lost from front page to submission, it is likely some could have not been enticed into the research no matter what tweaks would have been made to the website or the data gathering process. However, it is also likely that a substantial part was discouraged by the 100-item IPIP questionnaire. It follows that even more people would have dropped out if additional questions would have been added to the data form. The current expected time investment of 5-20 minutes was considered an optimal balance between depth of information and participant retention. However, small and simple additions such as a field for gender might have yielded interesting results without adding a noticeable strain on the participants. Gender was not included because games such as a Battlefield 3 have a notoriously low female demographic. Yet, considering the size of the eventual sample, gender data would have offered significant insight into gender-related correlations.

5.3 Cause & Effect of Sample Bias

The sample bias toward expert players is due to the method of participant recruitment. The most feasible approach to reaching out to and enthusing a large group of gamers for our research, was to address those that are already deeply invested in the game. Players with lower investment in the game are by definition less likely to busy themselves with game-related actions outside of direct play, and are therefore hard to find and reach. Arguably, they would also have been less likely to invest their time in the research even if they did know of it.

The resulting sample bias impacts the external validity of our findings. Additionally, it brings to bear the fundamental question if the relationship between play style and personality might be different depending on how dedicated a gamer an individual is. If 'gaming enthusiasm' would turn out to be a confounding variable for personality assessment through video games then having a relatively homogeneous sample on this dimension might actually be considered an advantage for the research. A similar result was found in the work by Iida et al. [10] with expert players on board games.

5.4 Future Work

We consider four main directions for future work. First, the current data set will be examined using data mining techniques. We hope to discover more complex patterns governing the link between play style and personality. Secondly, insights can be gained by combining our focus on personality and play style, with Yee's focus on gaming motivation [16].

On the one hand, personality is likely to influence motivation, while on the other hand, both are likely to influence play style. Thirdly, it would be interesting to see what kind of results could be gleaned from applying a similar research method to games involving a fundamentally different game play such as strategy or role playing games. Fourthly, advances may also be made with respect to the personality measures themselves. Viewing our findings in relation to other work we see that Shaker et al. [13] found strong correlations between emotions and controllable game variables, while Drachen et al. [6] found the same between heart rate and player experience. Our research hints at that the unexplained variance may be due to personality traits that mediate emotional responses to games. In general, we expect that people express some personality traits consistently in video games while others are not expressed at all, or become fluid within the confines of a simulated environment. Therefore, we believe that a specific personality test could be constructed that reflects those traits that are indeed stable between online and offline behavior. Our research hints at such constructs like 'work ethic' and 'speed of action'.

6. CONCLUSIONS

Our aim was to answer the question: *Does the statistically trackable play style of a player significantly correlate to his personality?* The answer we found is *yes*. Our findings have a high statistical power due to the large sample size we acquired through an elaborate promotional campaign (PsyOps). Effect sizes are in line with those seen for professional, medical, and psychological applications of the MMPI, Big Five personality inventory, and Beck's Hopelessness Scale. Our research specifically focused on the online tactical shooter Battlefield 3. Here we discerned three themes among the correlations between play style and personality. (1) Conscientiousness is the most predictive personality dimension, correlating negatively with speed of action. (2) Unlock Score per Second is the most predictive game play variable, correlating negatively with Conscientious and Extraversion. (3) Work ethic correlates negatively with game performance. Apart from these three themes, subsamples defined on gaming platform and native English speakers show different correlational patterns. Overall we may conclude that people do show their personalities in their play style.

An additional result was acquired during post-hoc analysis on age. Play style and age correlate *more strongly* than play style and personality, while age and play style are *complementary* in explaining variance in personality. Overall, older players are more conscientious, less experimental and slower than younger players.

7. ACKNOWLEDGMENTS

The authors would like to thank Gwen Ferdinandus, Armon Toubman, Jay Abdeljalil, Finn Görges, and Andreas Lundgren for their technical and artistic contributions to PsyOps.

8. REFERENCES

[1] M. Barrick and M. Mount. The big five personality dimensions and job performance: a meta-analysis. *Personnel psychology*, 44(1):1–26, 2006.

[2] J. Cohen. A power primer. *Psychological bulletin*, 112(1):155, 1992.

[3] P. T. Costa and M. R. R. *The Revised NEO Personality Inventory : Clinical and Research Applications*. Springer, 1998.

[4] L. Cronbach. Response sets and test validity. *Educational and psychological measurement*, 6(4):475–494, 1946.

[5] M. Donnellan and R. Lucas. Age differences in the big five across the life span: evidence from two national samples. *Psychology and Aging; Psychology and Aging*, 23(3):558, 2008.

[6] A. Drachen, L. Nacke, G. Yannakakis, and A. Pedersen. Correlation between heart rate, electrodermal activity and player experience in first-person shooter games. In *Proceedings of the 5th ACM SIGGRAPH Symposium on Video Games*, pages 49–54. ACM, 2010.

[7] A. Drachen, R. Sifa, C. Bauckhage, and C. Thurau. Guns, swords and data: Clustering of player behavior in computer games in the wild.

[8] R. Ganellen. Assessing normal and abnormal personality functioning: Strengths and weaknesses of self-report, observer, and performance-based methods. *Journal of Personality Assessment*, 89(1):30–40, 2007.

[9] A. Gow, M. Whiteman, A. Pattie, and I. Deary. Goldberg's 'ipip' big-five factor markers: Internal consistency and concurrent validation in scotland. *Personality and Individual Differences*, 39(2):317–329, 2005.

[10] H. Iida, N. Takeshita, and J. Yoshimura. A metric for entertainment of boardgames: its implication for evolution of chess variants. In *IWEC2002 Proceedings*, pages 65–72. Kluwer, 2003.

[11] R. McCrae, P. Costa, M. de Lima, A. Simões, F. Ostendorf, A. Angleitner, I. Marušić, D. Bratko, G. Caprara, C. Barbaranelli, et al. Age differences in personality across the adult life span: parallels in five cultures. *Developmental psychology*, 35(2):466, 1999.

[12] G. Meyer, S. Finn, L. Eyde, G. Kay, K. Moreland, R. Dies, E. Eisman, T. Kubiszyn, and G. Read. Psychological testing and psychological assessment: A review of evidence and issues. *American psychologist*, 56(2):128, 2001.

[13] N. Shaker, G. Yannakakis, and J. Togelius. Towards automatic personalized content generation for platform games. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE). AAAI Press*, 2010.

[14] G. van Lankveld, S. Schreurs, and P. Spronck. Psychologically verified player modelling. In *10th International Conference on Intelligent Games and Simulation GAME-ON 2009 (ed. Linda Breitlauch)*, pages 12–19.

[15] G. van Lankveld, P. Spronck, J. van den Herik, and A. Arntz. Games as personality profiling tools. In *2011 IEEE Conference on Computational Intelligence and Games (CIG'11) (ed. Mike Preuss)*, pages 197–202, 2011.

[16] N. Yee. Motivations for play in online games. *CyberPsychology & Behavior*, 9(6):772–775, 2006.