

Videogame Correlates of Real Life Traits and
Characteristics.

Athanasios Vasileios Kokkinakis

PhD

University of York

Computer Science

December 2018

Abstract

This thesis attempts to link real life cognitive traits and demographics with in-game player created data. The first experiment focuses on the nicknames chosen by players and the information one can extract from them. Players with nicknames with negative valence (highly racist or vulgar) tend to report others more and receive more reports from other players themselves when compared to their peers. Additionally, many individuals tend to have their real birthdate appended to their nickname (“Jim1986”). This is the first study that has successfully shown this phenomenon. Additionally, by using the extracted dates I showed that negative interactions tend to diminish as one ages.

In my second experiment I linked age with in-game performance, as indicated by rank for the videogame League of Legends (LoL). More specifically, performance in LoL tends to peak in one’s mid to late twenties while performance in first person shooters tends to follow a different pattern with an earlier peak; both seem to have a drop after 28. Moreover, I showed that fluid intelligence, as measured by the Wechsler Abbreviated Scale of Intelligence, as well as rotational working memory, which is an overlapping construct, are positively correlated with in-game rank suggesting they are the driving force between the age-rank findings.

Finally, I examined personality through the HEXACO framework. One of the most consistent findings in personality research is the existence of Neuroticism or Emotionality which is why it was my choice of focus. Individuals scoring highly on Emotionality, which is a trait linked to anxiety and sentimentality, tend to underperform in the competitive ladder. This was replicated with two videogames of different genres: Hearthstone and LoL. This thesis suggests that we can successfully extract meaningful information at a mass level through commercial videogames.

Contents

Abstract	2
Contents	3
List of Tables	9
List of Figures	10
Acknowledgements	12
Declaration	14
1. Literature Review	15
1.1.Introduction	15
1.2.Early Research and Military Applications	16
1.3.Academic and non-military Research	24
1.4. Microworlds and Decision Making	26
1.5. Commercial Videogames: MARBLE MADNESS & Tetris	28
1.6. Animal Studies	30
1.7. Mazes	31
1.8. Psychological Tasks as videogames	35
1.9. Commercial Games as IQ tests	40
1.10. Personality and Commercial videogames	42
1.11. Future Research	46
2. Nicknames in League of Legends	48
2.1.Chapter Introduction	48

2.2.Introduction	50
2.2.1. Hypotheses	51
2.3.Methods and Materials	54
2.3.1. Data Sources	54
2.3.2. Interaction Valence	55
2.3.3. Antisocial names	56
2.3.4. Age	58
2.4. Results	61
2.4.1. Antisocial names	61
2.4.2. Age	62
2.5.Discussion	63
2.5.1. Summary	63
2.5.2. Antisocial nicknames	65
2.5.3. Age	67
2.5.4. Conclusions	71
2.6. Limitations and Disadvantages	72
3. Videogame Rank and its relationship to Age and Fluid Intelligence	74
3.1. Chapter Introduction	74
3.2.Introduction	76
3.2.1. Intelligence and Task Selection	77
3.2.2. Videogame Selection	84
3.3.Study 1: Fluid Intelligence and associated measures	90
3.3.1. Methods	90
3.3.1.1.Ethics	90
3.3.1.2.Participants	90

3.3.1.3.Rank/ MMR /Rating Extraction and Interpretation	90
3.3.1.4.Instruments	91
3.3.2. Results	93
3.3.3. Study 2: Aging and videogame performance across 4 different videogames	95
3.3.4. Methods	96
3.3.4.1. Ethics	96
3.3.4.2. Data sources	96
3.3.4.2.1. League of Legends	96
3.3.4.2.2. DOTA II	96
3.3.4.2.3. Destiny	97
3.3.4.2.4. Battlefield 3	97
3.3.4.2.5. Fluid Intelligence and Aging	97
3.3.4.2.6. Grouping and Standardisation	97
3.3.5. Results	98
3.3.5.1.Outlier Rejection and Descriptive Statistics	98
3.3.5.2. Homogeneity of variance	99
3.3.5.3. One-way ANOVA – Welch’s F	99
3.3.5.4. Contrast Coefficients	100
3.4. Discussion	102
3.4.1. Study 1	103
3.4.2. Practice as an alternative explanation	107
3.4.3. Study 2	108
3.4.4. Age and Practice as confounding variables	108
3.5. Limitations	109

3.6. Implications/Contribution	112
4. Emotionality in Videogames	116
4.1. Chapter Introduction	116
4.2. Introduction	117
4.3. Study 1	119
4.3.1. Methods and Materials	119
4.3.1.1. Ethics	119
4.3.1.2. HEXACO-60	120
4.3.1.3. Gaming Questionnaire	120
4.3.1.4. Participants	120
4.3.2. Results	121
4.3.2.1. Missing Value Analysis	121
4.3.2.2. Descriptive Statistics and Distributions	124
4.3.2.3. Rank and Emotionality Correlations	126
4.4. Study 2	126
4.4.1. Methods and Materials	126
4.4.1.1. Ethics	126
4.4.1.2. Participants	126
4.4.2. Results	126
4.5. Discussion	128
4.5.1. Emotionality, its facets and Rank	128
4.6. Limitations and Criticisms	132
4.7. Implications	134
5. Synthesis	135
5.1. Introduction	135

5.2. Conclusion	136
5.3. Limitations and problems of videogame research	137
5.3.1. The Blackbox problem and its reasoning	137
5.3.2. Different Genres of video games may tap different abilities and conclusions may not generalise even in the same game across patches and servers.	138
5.3.3. The depreciation of videogame data	140
5.4. Sampling	141
5.4.1. Survivorship Bias and Special Populations	141
5.4.2. Range Restriction	142
5.4.3. Cross-Sectional Vs Longitudinal	144
5.5. Establishing Rank Correspondence with Standardised IQ ranges based on Age	149
5.6. Future Application Areas	149
5.6.1. Theoretical Framework for Education	150
5.6.2. Theoretical Framework for Health	151
Appendix A	154
Appendix B	156
Appendix C	184

Appendix D	200
Appendix E	202
Appendix F	204
References	205

List of Tables

Table 2-1. Descriptive statistics from ANT and random non-ANT players.	58
Table 3-1. The correlation matrix of the cognitive tests used.	94
Table 3-2. Player numbers and ages (in years) for the four games in our analyses after the outlier rejection and the age limits imposed.	99
Table 3-3. The homogeneity of variance tests for each game.	99
Table 3-4. Multiple planned contrasts for each game and age group.	101
Table 4-1. Our HEXACO-60 descriptives with the alternative means of a college aged male sample in parentheses.	124

List of Figures

Figure 1-2. Two examples of microworlds.	27
Figure 1-3. A snapshot of Marble Madness.	28
Figure 1-4. An example of Sorkin's maze.	33
Figure 1-5. McPherson's Space Code (left) and Space Matrix (Right).	36
Figure 2-1. Gameplay within League of Legends.	51
Figure 2-2. Histograms of player ages estimated from usernames from five servers.	58
Figure 2-3. Joint histogram of age estimates extracted from two different sources (N=10,299) with birth years between 1985 and 1999.	59
Figure 2-4. In game metrics for the two categories.	61
Figure 2-5. The valence of interaction rates changes with age.	62
Figure 3-1. The Cattell-Horn-Carroll (CHC) Framework.	79
Figure 3-2. The Verbal Perceptual Rotation (VPR) model.	80
Figure 3-3. The Operation Span task (OSPAN).	91
Figure 3-4. The Rotation Span Task (ROTSPAN).	91
Figure 3-5. The Symmetry Span Task (SymSpan).	92
Figure 3-6. Cross correlations between variables of interest.	94
Figure 3-7. Age profiles of MMR in four different games.	101
Figure 4-1. The top 6 questions with missing values.	121
Figure 4-2. Missingness across the various questions.	122
Figure 4-3. The distribution of Emotionality in our online sample.	124
Figure 4-4. The distribution of the Maximum Rank Achieved in Hearthstone in our online sample.	124
Figure 4-5. The distribution of Emotionality and LoL Rank in our laboratory sample.	126

Figure 4-6. A correlation heatmap of the Emotionality Subtraits and League of Legends Rank.	127
Figure 5-1. A sample before and after restriction.	1XX
Figure 5-2. The difference between a longitudinal and a cross-sectional sample	1XX

Acknowledgements

I am extremely grateful to IGGI and the EPSRC. This work was supported by grant EP/L015846/1 for the Centre for Doctoral Training in Intelligent Games and Game Intelligence (IGGI - <http://www.iggi.org.uk/>) from the UK Engineering and Physical Sciences Research Council (EPSRC).

I would like to thank Dr. Wade, Dr. Cowling and Dr. Gow for their help and guidance throughout these 4 years. Alex deserves a special mention since he carried most of the burden of having me as a student (and believe me that was not easy!).

Thanks a lot man, I respect you a lot and I was really lucky to have you as a supervisor. You have literally been amazing every step of the way.

I would also like to thank my fellow (IGGI) students for being a joy to be around. Special thanks goes out to Joe Cutting, Matthew Bedder, David Gundry, Simos Gerasimou, Mihail Morosan, Daniel Berio, Nick Sephton, Pete York, Andrei Iacob, Myatt Aung and Sha Li !

I would also like to thank Dr. Drachen for bringing valuable insight and giving me some good advice.

From the York crowd, last but definitely not least I would like to thank Jo Maltby. Jo, you are invaluable, I can't even count how many times you have helped me or IGGI. Thank you.

I would like to thank the users of /r/ hearthstone for being my participants and thus helping me create chapter 4. I wanted to publish this and thank you in a proper journal article but alas there is no time.

Most importantly, I would like to thank my mother, Athena, who has encouraged me to pursue higher education and who has always been extremely supportive of me. Words cannot truly describe how much I love you, how much I respect you and how thankful I am to have you as my mother. I hope you are proud me.

I would also like to thank my father George. He is alright.

In all seriousness, I love you Dad and thanks for all you have done for me and for all the good advice you have given me. I knew it was good advice because when I ignored it other

people gave me the exact same advice. Whenever we have talked throughout these years I am reminded of this quote: *“When I was a boy of 14, my father was so ignorant I could hardly stand to have the old man around. But when I got to be 21, I was astonished at how much the old man had learned in seven years.”*

I'd like to thank my sister Helen for being a nice and funny person. I love you very much and I hope you are happy.

Finally, I would like to thank my friends Giannis and Leonardos for playing games with me which helped me keep my sanity, at least for a while.

And of course, I would like to thank you dear reader because in the end this thesis is the tree falling in the forest.

Declaration

I declare that the work presented in this thesis is my own. This work has not been submitted for any other award at this or any other institution. If information has been derived from some other source, I confirm that my thesis indicates this. Some of the material in this thesis has been previously published in the following papers :

Kokkinakis, A. V., Lin, J., Pavlas, D., & Wade, A. R. (2016). What's in a name? Ages and names predict the valence of social interactions in a massive online game. *Computers in Human Behavior*, 55, Part B, 605–613. <https://doi.org/10.1016/j.chb.2015.09.034>

Kokkinakis, A. V., Cowling, P. I., Drachen, A., & Wade, A. R. (2017). Exploring the relationship between video game expertise and fluid intelligence. *PloS one*, 12(11), e0186621.

1. Literature Review

1.1. Introduction

The concept of Intelligence being expressed in Games is not new, with a vast body of research focusing on chess since it has always been assumed to be a cerebral activity (Barret, 2002; Howard, 2005). Indeed even in non-Western countries proficiency in strategy games such as Go or Mancala¹, has been linked to intelligence and wisdom, forming the archetype of a genius both when playing but also extending to real life situations (Barret, 2002). Spitz (1974) elaborates on this idea saying that “mentally retarded individuals”, a normal term back then, have trouble performing adequately in a multitude of strategy games such as Tic-Tac-Toe and Tower of Hanoi, suggesting that Intelligence heavily mediates their performance. He goes on to add that strategy games tap general Intelligence in a number of ways through visualization of possible moves, short-term memory rehearsal (knowing how many “beans” you have or your opponent has) (Spitz, 1974). However, due to the then-developing field of Intelligence and the cost of using strategy board games in terms of time, this form of mental evaluation proved relatively unpopular when compared to its pen-and-paper alternatives which were arguably equivalent if not better (Raven, 2000).

It is tempting to engage in the debate of what does and what does not constitute a game. However, as Washburn (2003) points out the line between a computerised psychological task (where one has to respond as fast as possible for instance) and a videogame is a blurred one. Washburn (2003) concludes that one "knows" a game when they are playing it and thus due to the subjectivity of this task one should be lenient with their definitions including the definition of game-like task.

¹ Mancala is an ancient African game where one has to move beans around holes, eventually capturing all of the opponents' pieces (beans).

In this survey, I will examine the relationship of real-life traits and how they are connected to video games. These traits will be intelligence as well as similar overlapping constructs such working memory as well as personality the HEXACO space (Ashton & Lee, 2014; Lee, Ogunfowora & Ashton, 2005; Bashiri et al., 2011). I will ask whether videogame scores, the rate of learning a videogame can predict IQ and working memory. I will also consider the relationship between valid psychological tasks and their gamification. I will additionally talk about how in-game statistics have been linked to various personality traits. Finally I will ask whether we can predict one's health by tracking down their videogame statistics and what statistics should game companies and health organisations choose to track if they want to be efficient with their predictions.

1.2. Early Research and Military Applications

Arguably, the greatest driving factor behind the use of videogames in a non-entertainment setting has been the United States' Army. The U.S. Army saw that previous ability tests were time consuming and required special machinery so when ATARI released their videogames they seized the opportunity by carrying multiple experiments in the summer of 1978 (Bittner et al., 1981). Albeit with a sample of only 13 people the researchers quickly realised that the ATARI Air Combat Manoeuvring (ACM) had really high test-retest reliability after the performance of participants stabilised ($r=.93$, $p = \text{unknown}$ since it is not mentioned in neither the original paper nor in an alternate passage introducing the paper); a second discrepancy is that of the sample, although in an introductory paper it is 13 people, the author claims it is 24 in the original army report and 22 in the official journal publication (Jones, 1980, p.6; Jones, Kennedy & Bittner, 1980 p.3; Jones, Kennedy & Bittner, p. 145, 1981). Subsequent experiments showed that performance in the ACM correlated highly ($r=.81$) with the conventional test of compensatory tracking where the participant has to keep a moving circle in a specific place through the use of a control stick (Bittner et al., 1981).

This strong correlation in combination with the multitude of assessment programs in the U.S army for instance the Perspectives in Evaluation Tests for Environmental Research program (P.E.T.E.R.), mainly created because of the reliability of tests in abnormal conditions such as a naval ships came into question (Jones, 1984).

The most interesting part about ACM however, came in a much later study examining the cognitive performance of Tank Commanders (Johnson, Jones & Kennedy, 1984). In their study Johnson and his colleagues (1984) examined potential Tank Commanders in a multitude of tests: Grammatical reasoning, Mathematics, Peer-Review as to see who soldiers think will make a skilled Tank Commander, the Hidden Patterns test, a Map Planning test as well as two video-games the ACM and the "Touch-me" video-game by ATARI; this videogame has high face validity, it is most probably there because the it was used by a physician as a cheap metric for Working Memory when he wanted to examine the effects of high altitude on Human Working Memory on a group that climbed the Mazama Tirich Mir (Flanagan, 1982 as cited in Jones, 1984). The reasoning behind them using Peer Review as a variable is that, in a military context at least, has been shown to be superior to supervisor ratings (Williams & Leavitt, 1947).

Performance on the ACM and the Touch-Me videogame were highly correlated measurements along with the paper-folding task. It should be noted that out of the three groups of prospective Tank Commanders examined (each group was assigned to a specific type of tank), these correlations were consistent in two out of three groups. These correlations further persist when all three groups were merged in a subsequent analysis. This could be attributed to either the small sample they employed for the first group (N=7) or to a priori individual differences that are present when soldiers are assigned to specific types of tanks (it could well be that soldiers with higher or lower aptitude are assigned specific tank models based on their systems' complexity). In a subsequent study, Jones, Dunlap & Bilodeau (1984)

replicated Johnson's and his colleagues' (1984), findings cross-correlating five videogames with a large battery of psychometric tests taken from the Educational Testing Service (ETS) Kit of Factor Referenced Cognitive Tests as well as Reversed Printing (Ekstrom et al., 1976 as cited in Jones, Dunlap & Bilodeau, 1984); in the Reversed Printing condition the participants were given 3 minutes to write down as many letters of the alphabet backwards and upside down thus this test taps mental rotation and executive control. The five games examined were the ACM, Breakout, Race Car, Slalom and Antiaircraft. Multiple averaged correlations between the battery of tests and each videogame ranged from $r=.18$ to $r=.5$. By looking at the individual correlations we can see that these simple videogames tap a range of characteristics ranging from mental rotation, reaction times and pattern identification yet due to their "unrefined" qualities and possibly simplistic graphs not enough visual load is produced so that they yield increased correlations.

Bliss, Kennedy, Turnage, & Dunlap (1991) further build on the literature by examining the aforementioned videogames with the addition of Surround. This was a longitudinal study with testing lasted for 12 weeks (3 weeks for each videogame) with researchers wanting to see if performance on this tasks a) reaches asymptotic levels with variance in individuals remaining the same (evidence of stability and reliability) b) if these tasks can be correlated with 3 types of tracking (Critical, Dual and Compensatory). The Dual task involved two objects while the Critical one had increased difficulty with the object not staying in place and being prone to change positions over time (Petzoldt, Bellem & Krems, 2014). Slalom and Surround were removed from the analysis because performance in these tasks never reached stability (never converged/capped). The remaining tasks' final scores correlated highly with all types of tracking with correlations ranging from $r=.45$ to $r=.81$. The important thing to note is that all the tracking tasks were done one year after the original videogame scores were obtained. Therefore this research shows that videogames may show

extremely high reliability, at least in tasks that require manual dexterity, ocular control or some combination of both.

A future study that showcases the importance of these findings is that of Rosser and his colleagues (2007). They showed that surgeons who played videogames were more proficient in laparoscopic surgery. Additionally, performance in the game “Super Monkey Ball 2” (SMB2) correlated highly with laparoscopic skill and suturing ability at $r=.63$, $p<.001$. In the SMB2 the player controls a monkey that collects tokens (“bananas”) while trying to move in a highly volatile platform that moves around before time runs out. This task highly resembles to an extent the balancing-tracking-stabilising tasks by Bliss and his colleagues (1991). The researchers conclude that videogames maybe incorporated in training curricula to assess and train future surgeons.

Aside from the U.S. Army, the field of intelligence in videogames owes much of its foundations to the Defense Advanced Research Projects Agency (DARPA), which was directly responsible for the creation of one of the most influential research games, Space Fortress. DARPA funded the Learning Strategies Program, which created Space Fortress, essentially making a videogame that taps fluid intelligence, psychomotor speed, attention and memory, shown in Figure 1-1 (Donchin, et al. 1989; Mané & Donchin, 1989). Most importantly Space fortress acted as a standardized paradigm across labs helping them compare and contrast their results, and its popularity can even be seen today having undergone multiple revisions and alterations while being one of the few videogames that have been played under the fMRI scanner (Boot, 2015; Shebilkse et al., 2005; Lintern, 1989; Donchin, 1995).

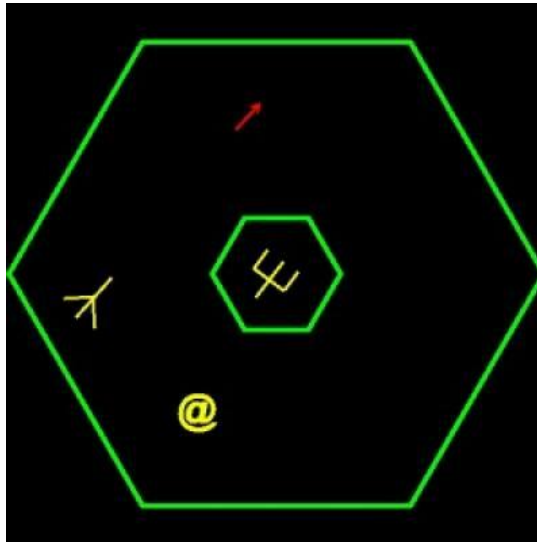


Figure 1-1. Participant has to destroy the fortress in the center by shooting missiles while avoiding projectiles shot by the fortress.

Much later in 1989 a whole issue of *Acta Psychologica* was dedicated to Space Fortress and its possible benefits in examining skill acquisition, with the exception of Rabbit, Bannerji and Szymanski. They, not only looked at skill acquisition, but also focused on whether high performance in videogames could be correlated with Intelligence. Rabbit and his colleagues (1989) also examined how Intelligence influences the rate of skill acquisition. They additionally took their experiment one step further and they examined both a normal population and a “gifted” population (top 10% of the scorers in the IQ test). Their first major finding is that indeed Space Fortress does correlate with IQ (as measured by the AH4 test (Alice Heim 4) which contains both verbal and numeric measures) up to $r=.69$, after some training took place (Heim & Batts, 1948 as cited in Rabbitt et al., 1989). They additionally found that again after enough practice the “gifted” sample’s scores correlated at $r=.3$ with IQ; the “gifted” sample was the top 10% of the population in terms of performance) . This is highly important since it is a range restricted sample which it makes these effects harder to identify (Detterman, 2014). Detterman (2014) suggests that when a sample has no low-performing individuals the true relationship of two variables may be obscured. An example

would be the lack of a strong correlation between height and points scored or salary in NBA players leading to the nonsensical point that height is not an important variable in basketball (Detterman, 2014). An important yet non-surprising finding in Rabbitt's et al., (1989) study was that Intelligence predicts the rate of acquisition of video-game skill (but see Sims & Mayer, 2002). In their regression Age remained a significant factor even when videogame experience was accounted for, hinting at Working Memory and motor deficits that come with later chronological age. This has been replicated by many studies in multiple genres such as Battlefield 3, Starcraft 2 and League of Legends (Tekofsky et al., 2015; Thomson, Blair & Henrey, 2013; Kokkinakis, 2013). Another important methodological guideline which seems to have been forgotten by modern scientists, enough to create a range of articles criticizing its lack of implementation, was the evaluation of gamers (Boot et al., 2008; Towne, Ericsson & Sumner, 2014; Latham, Patston, Tippet, 2013). Rabbitt and his colleagues (1989) did not just ask the participants about how good they were but they themselves rated them from a scale of 1 to 7 on their performance. Although it seems quite intuitive to cross-check someone's self-evaluation many modern psychologist only relied on their participants' subjective self-evaluation about how good they were or how many hours they played videogames (Green & Bavelier, 2003). This bad practice became more evident when many researchers could not replicate the now famous Nature paper by Green and Bavelier (2003) where Videogamers outperformed non-Videogamers and some short training could virtually eliminate these affects (Murphy & Spencer, 2009; Boot et al., 2008; Kokkinakis, 2011 unpublished dissertation; Unsworth et al., 2015). It should be noted that in the same issue Shapiro and Raymond (1989) trained a naïve/inefficient group to use the same ocular strategies as an "efficient" group by providing feedback on their reaction times and how they use their peripheral vision. The groups that received this improved their scores significantly. Although Shapiro and Raymond (1989) acknowledged that good players allocate their Attention better

over space, omitting irrelevant information as well as having better Reaction Times, they never explicitly theorise a link between Intelligence and Video-game score. Their results showed that the number of foveations correlated negatively with videogame score in the mine condition ($r(29)=-.49, p<.001$) while the Reaction Time to hit a mine correlated positively with the Reaction Times to hit a mine ($r(29)=.56, p<.005$).

Both Inspections and Reactions times have been linked to Intelligence creating the field of Mental Chronometry (Jensen, 2006). Inspection Times refer to the recognition of a stimulus while Reaction Times also include the time it takes for an individual to respond. The stimuli used usually are highly simple and discriminant (Burns, Nettelbeck & Cooper, 1999). Although the correlations between these measures and Intelligence vary, ranging from 0.3 to 0.6, evidence points at a link (Sheppard & Vernon, 2008).

Space Fortress however was not the only game commissioned. The U.S. army also funded the creation of a now forgotten videogame which appears in only one publication and one, now unclassified, military report; the Strategic and Tactical Assessment Record (STAR) (Graham et al., 1985; Graham, 1981). Graham (1981) was investigating how soldiers' mental faculties will be influenced by sleep deprivation and emotional stress. Thus he created STAR in such a way that it can have a large number of quantifiable variables while it simultaneously is engaging enough for the participant.

In STAR the player controls a spaceship fighting against a race of Aliens called Xenoids. There is a multitude of psychometric data recorded such as Reaction Times, Motor Accuracy, Long Term Memory (the participant provides a report whenever he "docks" with his spaceship), Subjective Rating of Fatigue/Workload, Risk Assessment and Resource Allocation since the "pilot" needs to manage the ship's energy. The researchers state that performance in this test stabilises after 5 hours "which compares favourably with the

Multiple Task Performance Battery” (p.650). However, that is an empty statement since there is no display of concurrent validity (STAR correlating with an actual standardised measure). In conclusion, although there has been no continuation of this study, the vast degree of data collected from each participant could have allowed for much more precise measurements possibly raising the skill mastery required for a “good” score thus allowing for finding meaningful differences in more gifted populations and possibly even personality. In STAR some of the data logged is the amount of fuel used for outmanoeuvring, evading or attacking enemy ships thus it would be possible given a big enough sample to see some manifestation of anxiety when one is facing challenging opponents. Additionally, target accuracy was more than just binary (hit or no-hit). The researchers logged the mean absolute course error as well as the mean time it takes for the participant to calculate each trajectory.

The U.S. army experimented in different genres of videogames as evidenced by the V.I.S.T.A programme (Videodisc Interpersonal Skills Training and Assessment) which essentially was a “choose your own adventure” with both text and video presentations (Schroeder et al., 1986). Soldiers got involved in imaginary scenarios ranging from a fellow soldier being insubordinate to a wife that “is lonely because the husband is gone on training exercises frequently”. Soldiers had to write down their potential answer in order to think about the issue and then after a small period of time the possible answers to the problem appeared on-screen in a multiple choice format. Unfortunately, due to lack of funds no longitudinal assessment of leadership was taken, rather an immediate test was adopted and the program faded off.

One can see that the U.S army had realised the potential of videogames as effective cognitive tests and simulations. These attempts although innovative were not always successful as shown by the V.I.S.T.A and S.T.A.R. programmes. Of course many of these programmes have relatively recently become unclassified and it is obvious that the successful

ones are still unpublished. The evolution of the original body of research by Jones and his colleagues can be seen today, with reports that the U.S. army actively recruits highly skilled players as drone pilots in videogaming fairs, knowing that their skills will carry over (Schei, 2014).

1.3. Academic and non-military Research

One of the first academic studies that used videogames was carried out by Griffith and his colleagues (1983) showing that video-gamers had increased eye-to-hand coordination as evidenced by the purse rotor task. As the authors themselves admit this was a preliminary/elementary experiment however their study was one of the first to examine this special population, partly due to the fact that arcade games that used a stick had started to gain immense popularity. Perhaps the biggest disadvantage of this study was that it used as measurement a task which taps both motor and mental abilities without trying to differentiate between the two, however that is perfectly understandable due to the “infancy” of the field; there was no reason to assume that the experiment would work in the first place and the allocation of resources to something purely experimental without that much impact would be unwise. Finally, another weakness that seems to commonly occur in video-game research is that of correlation-causation. It could be argued that individuals with better visuomotor skills would choose to play videogames thus rendering the study null.

Another important study that highly influenced the field of videogame research was carried out by Gagnon (1985). Although previous studies focused on videogamers and non-videgamers, this study focused to what degree videogame scores correlate with visual tracking and whether pre-existing gender differences in visual perception can be eliminated through training. Indeed, both Targ and Battlezone correlated with Visual Pursuit. The study in did not include a control group to control for placebo effects, however that should not be a problem. Perhaps the biggest methodological mistake is that they did not check whether the

two games use random numbers or pseudonumbers to generate enemies. Some individuals have reported internalising the pattern that the enemies appear as well as their A.I. thus being able to predict their movements rendering certain videogames quite simple (Hoff, 2015).

Not all video games necessarily have dynamic displays. A notable alternative game genre is that of a text adventures. Marshall-Mies and her colleagues (1983) as cited in Hunt & Pellegrino (1985) created a single text-based adventure-puzzle resembling DnD where the participants (power system operators) had to use the command line to solve it. Surprisingly, the time to solve it correlated negatively albeit weakly with supervisors' rating of power systems operator ($r = -.16$). Hunt & Pellegrino (1985) report this finding as a "discouraging result" (p.250) due to the single trial and the items reliability. I would argue that although they are correct to identify that as a possible problem the study used a sample of $n=3441$ people and the variable that was used was the supervisors' ratings. It is not illogical to assume that the text based adventure was not a cautionary tale of failure but rather a promising area of research. Had they used a standardised IQ test or peer evaluation, since we know that it can be better in as mentioned above than the ratings of people in command, then that correlation would be at least moderate (Williams & Leavitt, 1947).

Psychologists, such as Pellegrino and his colleagues (1987), developed a software package of visuospatial tasks, one of which included a videogame resembling asteroids. The player/participant tries to use a "missile" to shoot down a rectangular which can move in one out of three trajectories: horizontal, sine-wave and parabolic. The player cannot move and has to wait for the perfect time to shoot effectively visualising the trajectory of the missile. "The dependent measure is the vertical distance between the missile and the target when the target crosses the missile's path" (Pellegrino et al., 1987, p.234). This of course seems quite limited since some additional measurements such as percentage of accurate shots could be included. An improvement of this videogame was later created by Jackson and his colleagues (1993).

They used the term “dynamic display test” and their game had the following alterations : a) the player could move b) there were barriers between the player and the enemy c) the barriers could move at different speeds d) scoring changed based on accuracy e) the game became progressively more difficult with each trial. They used the Multidimensional Aptitude Battery (1994) which contains both verbal and visuospatial tests. Interestingly the scores correlated with the visuospatial tasks but not the verbal ones. An important point they made was the age of participants. The Researchers stated that a body of research shows that people over 25 underperform in measures of fluid intelligence as well as videogames. Not knowing whether practice effects could influence their videogame measurements they decided to exclude participants over 25. Videogame scores were correlated negatively with in the young sample at a $r=-.26, p<.02$ however when they decided to disregard the cutoff point by adding older participants this resulted in an absolute increase of $r=-.40, p<.001$. Of course now modern research shows that old age effects ELO/MMR in older participant regardless of training effects (Tekofsky, 2015). These results state that even simple videogames can detect Aging effects.

1.4. Microworlds and Decision Making

Microworlds are dynamic simulations where psychologists use to examine decision making. Even though they are not games they might have gamelike features, similar to the videogame SimCity. A notable example is the Moros microworld where the participant has to help a small, Saharan semi-nomadic tribe survive for 20 simulated years through a changing environment as shown in Figure 1-2 (Doerner, Staeudel & Strohschneider, 1986 as cited by Strohschneider & Güss, 1999). Participants had to allocate resources such as money or assign cattle to specific fields in order to avoid overgrazing, buying fertilizer and building hospitals among other things. Although, these worlds may sound rather simplistic, this passage over common mistakes can perhaps bring light to how multiple variables interact : “These

catastrophes stem from the fact that the participants tend to remove natural checks on cattle and population by fighting the tsetse flies and diseases without introducing sufficient artificial controls in the form of birth controls and slaughter of the cattle, and to increase the area of grazing land at a sufficient rate. As the number of cattle increases exponentially, the participants soon experience overgrazing. When the participants try to cope with this problem by further increasing the grazing land by boring more wells, they deplete the groundwater level with severe water shortages as a consequence...” (Juslin & Montgomery, 1999).

This game was sensitive to cultural differences with Germans showing different patterns of in-game problem solving when compared to Indians with less strategic mistakes and using more controlled approaches; although these groups were specifically chosen for their differential thinking styles, making this an extreme-group comparison, with Germans having a less formulaic/more Socratic way of reaching conclusions rather than memorising answers according to the researchers. In contrast the Indian schooling system has children learn solutions by heart. Additionally, managers seem to outperform students in running the Moros tribe due to their skills (Juslin & Montgomery, 1999).

A later study by Gonzalez, Thomas and Vanyukov (2005) would show that dynamic decision making is highly correlated with Visuospatial Working Memory as measured by VSPAN as well as fluid Intelligence as measured by Raven’s Matrixes, especially if timing demands are increased as shown in Figure 1-2.

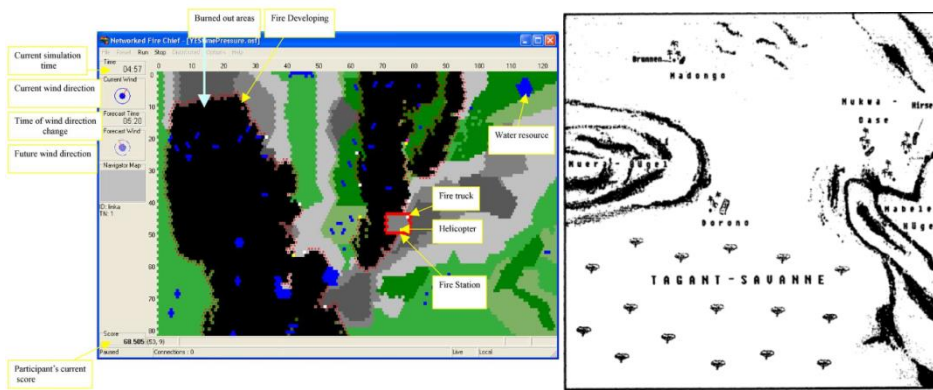


Figure 1-2. Two examples of microworlds. On the left (“Firechief”) the player has to distribute resources such as water or vehicles in order to put out a fire. On the left one can see the Moros game where the player has to micromanage a tribes resources for 20 years. Their key difference is time, with “Firechief” being real-time.

1.5.Commercial Videogames: MARBLE MADNESS & Tetris

The game Marble Madness was used by Subrahmanyam and Greenfield (1994) to examine the established spatial gender differences and the potential videogames had to eliminate them. They examined 10 year old children because according to the then-current literature their visuospatial abilities had some stabilised and they were reliably measurable. Although they focused on gender differences and training and did not acknowledge gIQ some conclusions can be drawn indirectly from their data. Initial scores in the spatial tests did not correlate with Marble Madness scores, they did correlate with the final level of mastery after training similar to the Rabbitt et al. (1989) study with $(r(29)= -.33, p<.056)$; they create a latent factor which combined three tests, the higher the score the worse their performance. The three tests they used were the following: an extrapolation test where the participant has to judge where a line will intersect with a second line, an intercept test where the participant has to judge the speed of an object and shoot it and the Memory Lane test which they later

deducted from the analysis due to poor performance by all the subjects (Pelegriño et al., 1987). This modest correlation shows that a videogame designed for commercial purposes could be used to identify visuospatial abilities, Figure 1-3.



Figure 1-3. A snapshot of Marble Madness. Players have to move the ball across obstacles in to the finish line within a time limit.

A second result was that although videogame training enhanced the scores of individuals who scored badly in the beginning, they still did not outperform individuals who performed great in the visuospatial scores initially in neither the video-game group nor the control group ($t(24)=2.07, p<.05$ and $t(27)=3.81, p<.001$). A possible explanation for this is that although videogame training can enhance visuospatial abilities to an extent, it still cannot override basic individual differences with many studies supporting the idea that visuospatial abilities have a large genetic component (Park & Gooding, 2014). The authors conclude that videogames can be used not only in the improvement of visuospatial skills but also in their assessment since many professions revolve around high visuospatial skills e.g. pilot, architects and radar operators (Subrahmanyam and Greenfield, 1994).

In the same year Okagaki and Fresch (1994) examined participants with no prior Tetris videogame experience. Their main focus was to examine Gagnon's (1985) and findings on the absence of gender differences when participants were trained in videogames.

Although they looked at training and gender differences they did attempt to correlate both initial and final game performance with Tetris Scores, coded elegantly in two different ways (lines completed and Tetris in-game score). Neither initial nor final game performance predicted any of the four spatial tasks (Perceptual Speed, Card Rotation, Cube Comparison and Form Board) (all p values $> .40$). It should be noted that Sims and Mayer (2002) found similar null results however in a later conference paper a notable correlation between Tetris and the paper folding task was found ($r(64) = .244, p = .05$). Thus results are mixed and highly dependent on the metrics used. Although, the absolute scores did not correlate the difference between the initial and the post-training Tetris score was correlated with the improvement in the tasks in the second experiment at approximately .4, yet in the second experiment they replaced some shapes in the standardized tasks with more tetris-looking blocks, so that correlation, although modest should be taken with a grain of salt.

1.6. Animal Studies

However, videogames cannot simply be used to measure Intellect and Working Memory just in humans. Rumbaugh, Richardson, Washburn, Savage-Rumbaugh, & Hopkins (1989), defying a 50-year-old literature that suggested otherwise, had monkeys (*rhesus macaque*) play tic-tac-toe successfully. Washburn (2003) further extolls videogames due to the standardised measurements and procedures they provide, for instance tic-tac-toe is perfectly understood by researchers with measurable responses, learning curves and strategies. Two monkeys for instance always defeated the computer when its responses were random while failing to do so when the computer used a simple strategy; they did not block the Os. The researchers concluded that monkeys did not exhibit even the simplest of strategies that children already display. However, that is not always the case. Martin and his colleagues (2014) showed that chimpanzees (*Pan troglodytes*) outperformed humans in an experimental economic game. Washburn (2003) concludes that standardized games show the

generalizability and standardization of Psychometrics not only across different labs but even across species.

1.7. Mazes

In a study with similar intentions of examining gender differences Moffat, Hampson and Hatzipantelis (1998) used a virtual maze with the dependent variable being the time one takes to find the exit. Although, finding one's way out of a virtual maze is not a videogame per se, labyrinths have been extensively used in multiple videogames thus making their inclusion mandatory (Gazzard, 2009). Males outperformed females even when videogame experience was taken into account (but see Levy, Astur & Frick, 2005). More significantly the maze score significantly correlated with a number of visuospatial tests such as the Vanderberg Mental Rotation Test (VMRT), the Money Road Map Test of Direction Test (MRMT), the Guilford-Zimmerman Spatial Orientation Test (GZSO), the Advanced Vocabulary Test I (VOCAB), the Controlled Oral Word Association Test (COWAT).

A really interesting finding is that of the distinctive correlations between the maze scores and the verbal correlations. Although, some researchers have supported the distinction between these two abilities, others support that they are both similar components. This becomes even more complicated when we see differential correlations between males and females and when one takes into account the fact that university students may behave mentally different than a normal sample; not only in terms of peak performance but also in actual basic processes since verbal Working Memory in students behaves differently when compared to a more normal sample (Kane et al., 2004). The dissociation between the female and male scores is also interesting with researchers supporting the notion that female participants may be using some form of verbal strategy to aid in their navigation. Of course this explanation is rather superficial since they do not provide any specifics. The more likely

explanation is that the COWAT task, in which participants generate words from a single letter in a minute, taps some form of speed and executive control, paralleling the concept of neural efficiency and mental chronometry, as mentioned in a previous paragraph. If one looks at the VOCAB test which is a measurement of crystallized intelligence where the participant has to choose the synonym of a word out of 5 options, a significant correlation is absent. Thus COWAT is driving the relationship. Indeed in a more recent study examining a Japanese sample Kawamura, Kobayashi & Morioka (2012) divided participants into three groups (high, average and low) based on their performance on a Working Memory task. Participants had to generate words based on four cue words. The a priori division of participants mirrored their scores with participants in the high-span group generating more words followed by the average-span group and the low-span group.

Building upon these maze findings, Ku and his/her colleagues (2003) created a virtual environment to see if it could be used in the diagnosis of schizophrenia since its expression is linked Working Memory deficits. Participants had to escape a Pyramid by choosing one of three doors in addition to evading evil Mummies. Each door had a marking and an underlying rule similar to the Wisconsin Card Sorting Test. In the Wisconsin Card Sorting Test the participant is given the choice between a number of cards with each card having a specific payoff as well as risk expressed in points lost (Nyhus & Barcello, 2009). The Wisconsin Card Sorting Test (WCST) measures risk-reward understanding as well exploration and is linked with frontal lobe activation. "Normal" participants would experiment and shift their choice from a bad paying card to a better one showing disengagement/strategy (Nyhus & Barcello, 2009). A wrong choice did not result in obvious loss of points but in a vibration indicating the mistake with the door opening anyway. Although, highly interesting their results could have been improved if they had used matched sampling focusing on variables such as WM and IQ.

Sorkin (2006) used similar methodology of virtual “Mazes” to diagnose Schizophrenia. Although Sorkin (2006) calls it a “Maze” it is more of a virtual environment since participants cannot get lost in them; a mistake in choosing the wrong door/path simply results in waiting time. She had participants choose one of three doors out of a total of 50, with each door either leading to a normal room or a room that would delay them for 20 seconds. If a normal maze was used, one could assume due to the Moffat et al. (1998) findings that a schizophrenia diagnosis could be reached since normal mazes still tap WM as well as other mental rotation processes. Indeed, many newer studies that did not use in-built WM tests still managed to identify schizophrenia patients even when using simple virtual worlds where the participant had to find his/her way back to a previously shown point (Zawadzki et al., 2013; Weniger & Irle; 2008). Sorkin still had issues with the exact classification of symptoms due to the complexity of schizophrenia symptoms and the comorbidity many of them have.

Sorkin additionally inserted surreal images in a virtual world (Incoherencies Detection Task) since reality monitoring is an issue with many patients. She created 3 conditions a) an Audio-visual Incoherency where the object would emit a different sound than normal (for instance an ice-cream truck with an ambulance siren or vice versa) b) an Incoherent Colour for instance “blue potatoes” on top of a pile of normal ones c) an Incoherent Location for instance a giraffe inside a shop as shown in Figure 1-4. Participants had to click the mismatch in the game world as well as provide a verbal report of it. The sound condition was the best in terms of disease prediction. If one inspects the histograms one could detect a major problem with the sound condition which is that a large percentage of patients was non-responsive possibly “dragging” the scores down. A second problem recognised by the author is that of the small sample and the multitude of symptoms subclassification which they admitted is should be the scope of future work.



Figure 1-4. An example of Sorkin's maze. In the first column one can see the surrealistic incongruent condition while in the second one one can see the modified WCST.

Although this body of work seems relatively impractical, why download a virtual maze instead of using a short questionnaire, it has a lot of merit (Mason, Linney & Clarridge, 2005). Mazes can be incorporated in already existing games since a large body of games uses them even today (Gazzard, 2013). Moreover, one can use this subtle manipulation of the WCST to refine sensitive population selection. Additionally, the Incoherencies Detection Test can be used non-invasively in terms of “bug” reporting to monitor at risk populations. Fake audiovisual “bugs” could be introduced in a game played by an individual with high schizotypy, and thus at risk of schizophrenia, with the report rate of those “bugs” becoming the new dependent variable (detection rate). Finally, some really important conclusions by Sorkin are that the nature of the Incoherencies should be a) auditory, since that was the case with the most mistakes b) the object-sound mismatch is more easily detectable if the two objects share some similarities (eg a cat barking would be more easily identified as a Incoherency/bug than a car barking) c) it easier to move a sound to a moving object than a static one.

1.8. Psychological Tasks as videogames

Ryan (1994) published “Memory for Goblins” which measured both short-term and working memory, for the Apple IIe and II+. Ryan (1994) had previously published the abstract for this project in “The Gerontologist” Journal stating that “Memory for Goblins” is a variation of the Counting Span Task where goblins replace the red dots which they have to report later from memory (Ryan, 1986 abstract; Case, Kurland & Goldberg, 1982 as cited in Ryan 1986). Similar to STAR, this game did not have any concurrent validity and has not been built upon since then. However this is one of the first mentions of using videogames for memory training for a geriatric population.

Berger, Jones, Rothbart & Posner (2000) made a notable, yet methodologically questionable and impractical, attempt to create videogames that are reskinned versions of the Alert Task (the child needs to tap on a touch-screen on an animal to put it back on the farm it escaped from), the Stroop Task (in this audiovisual variation the child needs to tap on the picture of a dog when the sound of a cat is heard and vice-versa), the Orienting Task (the child needs to feed the fish which can appear in one out of two directions) and the Spatial Conflict task (a house with a picture can appear either on the left or the right and then after a short time the picture appears again in either the left or the right and the child needs to match them).

Their results “made sense” replicating past findings regarding congruent and incongruent trials, for instance when one hears a dog barking and then presses the dog picture one’s reaction time is faster than pressing the picture of a cat. However, as the researchers acknowledge, the use of a touchpad lead to “misleadingly long” RTs in 3% of the trials which were corrected by the use of cameras present in the lab (p.300). However, the use of touchscreens in RT research for the following reasons a) the games are not highly portable across institutions b) the measurement of RTs is highly dependable on the instrument use, as

evidenced by the controversial study claiming that Victorians had better RTs and possibly IQ than the current population (Woodley et al., 2013 but see Parker 2014 and Dodonova & Dodonov, 2013). In a similar note Berger and her colleagues (2000) fails to cross-correlate the games with the standardised tasks they claim to replace. Finally, she fails to show the *raison d'être* behind the creation of these videogames which is that children like to play them because they are fun and captivating. Although the tasks last for 5,7 and 10 minutes they still have limited skins and the whole process is highly repetitive. Berger's only disappointing statement is that "The reaction of all our 5-year-old subjects to the games has been positive. Moreover, the children seem very fond of the games and enjoy the lab session." (p.302). In conclusion this experiment is highly flawed and perhaps even pointless as recent data show that simple, monotonous, reskinned versions of psychological tasks may not improve data collection even though they improve the experience of the participant (Hawkins et al., 2013).

McPherson created the game Space Code as part of his thesis to create tests that will be popular with children (McPherson & Burns, 2007; McPherson, 2008). In his first article McPherson used an older speeded reaction clicking task, Digit Symbol shown in Figure 1-5, and overlaid it with better graphics and sounds in order to create Space Code (McPherson & Burns, 2005). The reasoning behind tasks that tap Gs (speed) is that they are relatively simple given enough time and can be completed with zero mistakes (eg tap a button when a light appears) but when time limits are introduced then they tap some form of "neural speed" which correlates with IQ (Burns, Nettlebeck & Cooper, 1999).

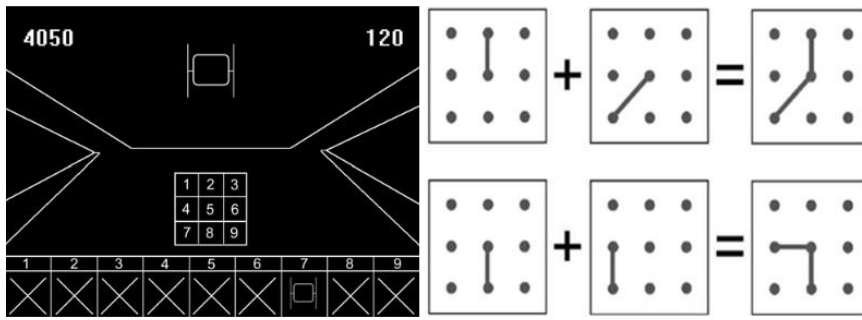


Figure 1-5. McPherson's Space Code (left) and Space Matrix (Right). In Space Matrix they need to memorise the lines and answer True or False when the solution appears.

In Space Code the participant controls a spaceship overlooking space through a cockpit. In the centre of the screen a number appears and the participant has to click the same number in a dial pad in order to “destroy” it as shown in Figure 1-5. He subsequently ran two experiments cross-correlating Space Code scores with a number of tests general Speed and visuospatial tests. In the first experiment he used a simpler graph version of Space Code while in the second one he included a timer, an aversive sound when points were lost, a score box, a cooldown on the missiles to counter the strategy of “spamming” them as well as audiovisual enhancements (explosions, lasers etc). Two tests yielded the strongest and most notable correlations. The first is the Digit-Symbol Test (WAIS-III) where the participant has two minutes to fill out blank cells based on a key on top of the page with the correlation of $r=.4, p<.001$ and $r=.54, p<.001$ in the first and second experiment respectively. The second one is the Visual Match Task the Woodcock–Johnson III Tests of Cognitive Abilities where the participant has to circle as many similar pairs of numbers as they can in 3 minutes with $r=.49, p<.001$ and $r=.60, p<.001$. The scores in the second version significantly correlated with Raven's Advanced Progressive Matrixes short-form, a highly used IQ test ($r=.49, p<.001$). It would be tempting to link this game with Intelligence or Visuospatial Working Memory, however one should not forget the “third variable problem” in correlational

analyses; it could well be that mental speed, the main variable measured here correlates with the aforementioned concepts without necessarily tapping the same mechanisms, even though theories about mental/neural efficiency and chronometry purport otherwise (Linden, 2007).

Indeed, in a following paper Space Code failed to correlate with grades in a school sample of 11 to 14 year olds, while an alternative creation Space Matrix that taps visuospatial ability/WM did. Space Matrix was a variation of Dot Matrix a task developed by Law, Morrin & Pellegrino (1995) and later modified by Miyaki and his colleagues (2001). Participants had to view a link between two dots in a 5x5 matrix with each link appearing briefly for 1.5 seconds. They subsequently were presented with a potential answer with participants decided whether it is true or false, as shown in Figure 1-5. Although, many researchers recognise the superiority of WM over reaction and inspection times and it is tempting to classify Space Code as a “failed” videogame I believe that the researcher had made many methodological mistakes which should be pointed out as they have been previously reoccurred in the literature.

The first obvious “hint” is that Digit-Code the non-game version of Space-Code are almost identical in terms of “mechanics” yet differs only in graphics and scoring, is correlated with preparatory school grades albeit weakly $\tau = .25, p < .01$ while $\tau = .19, p = n.s.$ The second obvious discrepancy is that in their new study Space Code correlates at an $r = .29, p < .001$ with the school sample instead of an $r = .45, p < .001$, with the university sample. At this moment it should be mentioned that the new Space Code was more simplified with removed visual components and background music which could in theory act as a distractors thus raising the ceiling cap of the task. In the previous paper McPherson & Burns (2007) had a harder version of the task which correlated at $r = .54, p < .001$. Thus by making these small alterations they lowered the difficulty of the task which endangered its concurrent validity.

The second major problem with the study is that of the highly biased sample. In his article Henrich and his colleagues (2010) question the generalisation of many conclusions because they are demonstrated and conceived based on predominantly WEIRD societies. WEIRD stands for Western, Educated, Industrialized, Rich, and Democratic. Although, I partially disagree, his argument becomes highly apparent/strong/evident with McPherson & Burns (2007) article. The first sample they used was exclusively University Psychology students which albeit questionable is still understandable since the majority of institutions work this way. Unfortunately their second sample was highly problematic because it consisted solely of private school students and on top of that, all of them had As, Bs and Cs with no marks below that (p.976). This is a highly biased sample (“Educated” and “Rich”) sample which again translates to the restricted sample Detterman (2014) spoke which makes it harder to find meaningful relationships between two variables. Moreover, if one looks at the distribution of grades with the majority of them being As (85), followed by Bs (70) and finally only 55 Cs. Thus with no low-performing participants and with a highly biased samples no meaningful conclusions can be made; similar to the problem of range restriction reported above (Detterman, 2014). Indeed in a different study examining 184 children in mostly Russian public schools Gs (Reaction Times and Inspection Time) was highly predictive of school grades (Dodonova & Dodonov, 2012). Dodonova & Dodonov (2012) found that the discrimination task (highly similar to Space Code), where the participant has to discriminate between triangles and other geometric shapes that briefly flashed on-screen, and the Recognition of Meaningless Figures Test (RMFT) another Inspection Test both correlated with Math ($r=-.21, p<.001$ & $r=-.261, p<.001$ respectively). They also correlated with Language ability ($r=-.253, p<.001$ and $r=-.238, p<.001$ respectively). One should remember that the correlations are negative because slower reaction and inspection times are linked to decreased intellectual abilities. It should be noted that in this sample they did not have failing

students but the grades were centrally distributed instead of towards “A”. Finally, as Deary, Penke & Johnson (2010) claim in their article cognitive variables are more predictive with lower performing individuals.

Another minor problem is the lack of practice. Although some studies show that initial scores are predictive of final scores, others claim that some practice is required (list of studies). Especially, since there is a lot of variation in terms of game experience with the sample being predominantly female (52 over 42 males) and with gender being correlated at an $r=.25, p<.05$ (higher scores mean males played more). Although to the authors credit they did a split analysis separating the two genders which did not yield any significant differences. This analysis however is partially misleading. In their correlation tables gender did correlate weakly with performance with every task albeit it was not significant at the $p<.005$; it could well be that it was significant at the $p<.1$ and with the adult male sample outperforming the adult female sample in all these tasks it should not be ruled out that there was some gender bias, especially if one takes into account the large samples required for Pearson’s correlation to detect weak relationships (Moinester & Gottfried, 2014). Thus a one-off practice is not enough for a participant to reach their “true score” nor is the sample high enough to detect subtle relationships. In conclusion even though Inspection Times and Reaction Times correlations and their interpretations are relatively controversial they should not be dismissed when designing videogames (Wagenmaker & van der Maas, 2006).

1.9. Commercial Games as IQ tests

One of the most important and rounded studies was done by Baniqued and her colleagues (2013). She investigated what abilities are tapped and by casual videogames the likes of which can be found in websites such as newsgrounds.com or miniclip.com; simplistic flash videogames. She cross-correlated them with an wide battery of psychological tests : Raven’s Progressive Matrixes, Shipley Abstraction, Letter Sets, Spatial Relations, Paper

Folding Test, Form Boards, Digit Symbol, Letter and Pattern Comparison, Logical Memory, Free Recall, Paired Associates, WAIS Vocabulary, Picture Vocabulary & Synonym-Antonym. Additionally, they included task switching, trail making, working and short-term memory, visual short-term memory, n-back, spatial working memory, forward and backward digit span, the Attention network test, Stroop, Attentional Blink.

The 20 games they selected can be found in Appendix A, adapted from (Baniqued et al., 2013). It should be noted that the researchers were highly diligent in their selection of participants excluding participants who were videogamers thus effectively getting more objective measurements since videogames have been linked to improvements in either all the aforementioned tasks or some underlying processing that mediates them (McDermott, Bavelier & Green, 2014). Baniqued and her colleagues (2013) concluded that many of these games are positively correlated with actual psychometric tasks ranging from fluid intelligence, working memory as well as speed.

Martinovic and her colleagues (2015) used the NEPSY-II, an Intelligence Test designed for children, with 15 games that were selected ("critic proofed") in another study; Martinovic and her colleagues (2014) created the method of "critic proofing", which is a qualitative methodology that is used to define what abilities are tapped by a commercial videogame. Martinovic and her colleagues (2015) examined children between the ages of 7-12. She found moderate correlations between the games and the NEPSY-II subcomponents. The study however had many methodological mistakes which their authors readily admit such as the lack of counterbalancing with the tasks having a fixed order. Although, this not always an issue the authors admit that the testing sessions were extremely long lasting from 3 to 4,5 hours. Thus fatigue effects may have been present. Similar to McPherson's experiments, the researchers recognised that although they did try to get a diverse sample they mostly got academically decent individuals. Another confounding variable is that the

authors did not control for videogame experience which may have been a factor as stated before, especially in a non-uniform sample (various ages and gender). Additionally, the experimenters admit that when they run partial correlations controlling for Age some of their correlations disappear perhaps reflecting Working Memory maturational processes. In her article "Neuroimaging: Taking "developing" seriously" Karmiloff-Smith (2010) warns against making judgements between young participants stating that due to the ever developing brain different age groups provide a snap-shot of scores without necessarily providing information about what happens in between. Thus researchers should not assume a linear relationship in many sensitive constructs between 7 year olds and 12 year olds.

Quiroga (2015) published a similar study a month afterwards. She used 12 videogames in the Wii and one on PC and cross-correlated them with a number of gIQ measures such as the Raven's Progressive Matrixes test, verbal tests as well as speeded reaction tests. The model yielded extremely good results with videogames being correlated highly with Intelligence scores ($r=.93$). These results are highly promising in using videogames as assessments with younger populations. It should be noted however that 10 out of 12 of these videogames were from Big Brain Academy thus they were carefully designed by Psychologists to measure Intelligence rather than being commercially made and readily available to the public. Furthermore they are operating on Wii rather than a computer thus their mass distribution and applications seems more improbable.

1.10. Personality and commercial videogames

The manifestation of Personality in videogames has been a topic with long history, starting as early as 1989 (Bartle,1996). In his seminal work, Bartle (1996) summarised his notes on what experienced players found fun or wanted out of his MultiUser Dungeon (MUD). MUD effectively was a multiplayer real-time text based setting, akin but not identical to a videogame. When looking at his notes Bartle found considerable overlap in the

wants of the players which he further classified into 4 distinct player types: the Achiever who is interested in completing tasks or goals, the Explorer who is interested in experimenting with the virtual world, the Socialiser who wants to roleplay and converse or interact in some way with others and finally the Killers who wants to impose his/her will on others essentially killing them in order to cause distress.

However, few studies have used actual psychometric questionnaires such as the Big-Five or the HEXACO-100. A key landmark study was conducted by Yee and his colleagues (2011) where they extracted a large number of individual data from an online tracking website for the famous Massively Multiplayer Online Role-Playing Game (MMORPG) World of Warcraft (WoW). The data he collected ranged from amount of time played to how many in-game gestures the individuals performed; an individual can perform gestures such as waving or hugging. They subsequently administered a big five questionnaire and cross-correlated the aggregated data. Interesting yet weak correlations emerged, for instance a negative correlation between Extraversion and the amount of in-game characters one has with $r = -.12, p < .05$. Additionally, many player versus player variables were negatively correlated with Agreeableness ($r = -.08, p < .05$) suggesting that individuals who liked that aspect of the game may have been less friendly. This study is extremely important for many reasons.

Firstly, it did not rely on self-report but made use of in-game data. Secondly, the author appear to take into account the game's complexity and tried to normalise their data based on the number of characters one had in WoW. Thirdly, it was highly exploratory in nature bringing attention to other researchers about the multitude of variables one can extract from a single game. A major problem with this study however was the lack of any corrections for the multitude of comparisons (46), for instance Bonferroni, which seems problematic since there were no major a priori hypotheses. However, that is easily understandable and excusable due to the study's highly exploratory nature in combination with the noise of in-game metrics

which could lead to Type 2 error; with Yee and his colleagues calling for additional exploration of these variables which would help against Type 1 error. Additional research by Yee and his colleagues (2012) looked at how demographics for instance gender and age influence in-game behaviour in WoW. They found that younger males tend to spend a large number of times on competitive and rewarding activities (for instance player vs player matches), while females and older individuals were more time constrained and focused more on exploration.

Another key study that partly replicates Yee's and his colleagues findings (2011) was done by Worth and Book (2014), where they cross-correlated the HEXACO-60 with many latent variables of gameplay in WoW. Some of these variables were Player-versus-Player (Fighting other Players), Working (Collecting and crafting in-game items), Helping (being friendly towards other players through emotes), Immersion, and Core Content. Significant correlations with HEXACO-60 personality traits were found for each component. Player-versus-Player activities were primarily related to low levels of Honesty-Humility and Agreeableness ($r=-.453$ $p<.001$ and $r=-.171$, $p<.05$ respectively) as well as having higher levels of psychopathic traits as measured by the Criminal Tendencies questionnaire and the Self-Report Psychopathy questionnaire. This is in accordance with Yee's et al. (2011) findings since both Honesty-Humility and HEXACOs Agreeableness overlap partially reflecting the Big Five's Agreeableness (Anglim & O'Connor, 2018). Social Player-versus-Environment activities were primarily positively correlated with Extraversion, Working activities were positively correlated with Conscientiousness, Helping and Immersion activities were positively correlated with Openness to Experience, Agreeableness and Honesty-Humility, and Core Content activities were positively correlated with Emotionality.

Another major study that focused on First Person Shooters (FPS) rather than MMOs was conducted by Tekofsky and her colleagues (2013). She focused on Battlefield 3 and used the

100-item IPIP (International Personality Item Pool) Big Five questionnaire. Tekofsky and her colleagues (2013) collected a really large sample ($n=13,376$) and attempted to correlate different personality dimensions and demographics with in-game statistics. Their key findings were that the in-game variable Unlock Score per Second, which is a measure of how well a player achieves target scores using various classes or weapons, was correlated negatively with Extraversion, Age and Conscientiousness. Moreover, Age was negatively correlated with in-game performance with older players being slower, which is echoed in one of my studies (Kokkinakis et al., 2017).

A study that focused on game preferences rather than in-game data was conducted by Zeigler-Hill & Monica (2015) and attempted to connect the HEXACO framework with the BrainHex model of player types (Nacke et al., 2015). This model assumes the following seven distinct, archetypal player types: Seeker, Survivor, Daredevil, Mastermind, Conqueror, Socialiser, Achiever. Seekers enjoy exploring new things similar to Bartle's Explorer type (Bartle, 1996; Nacke et al., 2015). Survivors enjoy fear and the excitement that is linked with evading threatening situations. Daredevils enjoy the risking in games. Masterminds enjoy solving puzzles as well as making strategical decisions. Conquerors enjoy defeating impossibly difficult opponents barely achieving victory as well as beating human players. Socializers are trusting players that enjoy helping and interacting with others (Nacke et al., 2015). Finally, Achievers enjoy completing tasks and gathering in-game valuables such as treasures or tokens via grinding (Nacke et al., 2015). Zeigler-Hill & Monica (2015) identified that HEXACO's Emotionality was negatively correlated with both the Conqueror and the Daredevil archetypes with $r=-.16$ and $r=-.23$, $p < .01$ respectively. The strongest correlation between the HEXACO dimensions and the BrainHex model was the link between the Socialiser archetype and eXtraversion correlating with $r=.34$, $p < .001$.

Overall, all the aforementioned research presented shows that real-life personality traits from valid psychometric questionnaires correlate with both in-game data as well as game genres. Most of the correlations are relatively weak however that is expected due to the noise of in-game statistics. Finally, the fact that each game might attract a specific self-selected sample could potentially range-restrict the sample which could lead to smaller correlations than normal.

1.11. Future Research

Videogames have been shown to correlate with many cognitive processes across many genres ranging from puzzle games to mazes and platform games. Videogames also have the additional advantage of being extremely popular with various heterogeneous populations. Therefore by deconstructing the link between WM, intelligence, reaction times and personality and linking them with massively played videogames we can effectively monitor at risk populations unobtrusively without using spam e-mails, letters or forcing doctor visits. Many diseases such as Alzheimer's or Schizophrenia have been shown to influence one's cognitive abilities such as WM both in the beginning of the disease as well as after its expression (Kessinger et al., 2003; Park & Gooding, 2014). Moreover, Anxiety has been linked to a number of health problems (Denollet et al., 1996; Dishman et al., 2000). Studies by Tekofsky (2015) as well as Thomson, Blair & Henrey (2013) have shown that we can detect the developmental trajectories of Aging in two distinct videogames (FPS and ARTS), even when the only thing known about the populations is their video-game score. The effectiveness of these metrics could be highly enhanced if we could use additional metrics such as short questionnaires or a priori information such as a player's medical history provided by one's parents or by the players themselves (Mason, Linney & Clarridge, 2005).

Videogame data measurements and interventions are not limited to diseased populations. Certain inferences are still highly applicable to the normal population since we can still extract meaningful data such as IQ. Deary and his colleagues (2004) followed up on the Scottish Mental Surveys which were carried out in 1932 and 1947 that had an N of 89,498. They found that mortality rates were influenced by IQ, with more intelligent individuals living longer. Deary & Der (2005) further looked at a sample of 898 people and found that Reaction Times heavily mediate that relationship. Both of these variables can be extracted with reasonable accuracy from modern videogames as evidenced previously in this Literature Review. It should also be noted that these experiments were carried out with a western sample. It is possible that in a developing society, where healthcare and disease preventions is much less prevalent, intelligence plays a much more important role at mediating mortality. In conclusion a closer relationship between the Game and Health Industry as well as Academia should develop to help the lagging field of diagnosing disease and lower IQ, especially since early interventions are highly effective against diseases such as Dementia (Robinson, Tang & Taylor, 2015).

2. Nicknames in League of Legends

2.1. Chapter Introduction

Multi-player online battle arena games (MOBAs) are large virtual environments requiring complex problem-solving and social interaction either through chat or through the form of virtual tokens (termed “Commends” or “Honor”). These tokens are generated by users at the end of each match and they reflect how other users perceived a player’s behaviour and/or actions i.e positive (Honor) or negative (Reports). Here, I ask whether these games generate psychologically interesting data about the players themselves. Specifically, I focus on whether usernames, which are chosen by players outside of the game itself and act as a near-permanent representation of the user in the in-game world, predicted in-game behaviour.

I analyzed a large anonymized dataset from a popular MOBA (‘League of Legends’) – by some measures the most popular game in the world. LoL is ideal for this experiment due to the stability of one’s nickname (if one chooses their nickname it is relatively permanent since they can only change it through paying as opposed to other MOBAs such as ‘Dota 2’ where one can change it easily via Steam multiple times per week) (Steam, n.d.). Moreover, due to the competitive nature of the game, there are many situations which are highly polarising and may lead to conflicts among the players of both teams (both within the same team as well as the enemy team). This type of conflict can elicit strong emotions and can lead to highly negative behaviours for instance name calling which would probably not arise if the game a) was not as engaging or b) it did not involve a live chat mechanism .

I additionally investigated whether individuals sometimes insert their birthdate into their username. Using Riot’s registration data , I cross-correlated the numbers from names that appear to contain birth dates with the birthdate used to register in Riot’s server. Using

robust age estimates – and based on results from other papers in the development literature, I also ask whether age also has an effect on the aforementioned Reporting and Honouring system.

Finally, I describe subtle methodological issues and limitations that are not obvious to the non-videogamer such as the effect of variety within possible Report Types. I will also take a closer examination at the Report System in general and the data that were given to us. Possible improvements on the Report System along with directions for future studies will also be discussed.

These findings suggest that players' real-world characteristics influence behavior and interpersonal interactions within online games. Moreover, one can derive actual demographics such as age from a nickname (given a large enough sample). Anonymized statistics derived from such games may therefore be a valuable tool for studying psychological traits across global populations.

2.2. Introduction

Online video games are played by hundreds of millions of people worldwide and fine-grained statistics on each game are constantly relayed to centralized servers where they can be stored and analyzed. These games often require complex team strategies and permit direct personal interactions mediated by real-time chat, as well as inter-player rating mechanisms. They therefore represent a rich potential source of data for psychological investigation.

Previous research on relating personality traits to video game characteristics have often correlated findings from personality questionnaires with game data: either statistics collected within the game environment, or statistics about the amounts or types of games played (Chory & Goodboy, 2011; King, Delfabbro, & Griffiths, 2013; Park, Song, & Teng, 2011; Teng, 2008; Worth & Book, 2014b; Yee, Ducheneaut, Nelson, & Likarish, 2011). This approach is valuable because personality questionnaires provide verified indicators about stable, real-life personality traits. However, if a subject is sensitive, respondents may respond untruthfully, not always realising it, even to questionnaires administered anonymously across the internet (Paulhus, 1998). Moreover, completing questionnaires can be extremely time-consuming for certain participants, thereby limiting the number of individuals who can be included in a study via limiting the participant who are willing to participate or via dropout effects.

An alternative approach to the aforementioned psychological testing (questionnaires) is that of 'mining' very large datasets for scientifically relevant relationships. This approach is interesting for several reasons. First, it is valuable to ask whether large datasets of this type are useful for statistical analysis at all. It may be, for example, that all players adopt a single 'optimal' or uniform strategy or behaviour that leaves little room for personal variability, rendering these datasets uninteresting from a psychological viewpoint. An obvious example

would be that of randomly-allocated players behaving in a uniformly positive manner towards their teammates because they know that it creates a good team environment that will ultimately win them the game. In the world of online gaming, this strategy is, to put it mildly, uncommon.

Secondly, if players do seem to exhibit systematic differences in behaviour, it might be that some of this variance is linked to real-world characteristics such as age, gender or personality (Worth & Book, 2014b). Understanding these relationships could provide valuable information about these characteristics at a population level, and this information could be used as a preliminary screen to identify subjects who may be suitable for further testing. Finally, from a system design point of view, if reliable metrics on player behaviour can be established, they can be used to improve the social environment within the game.

2.2.2. Hypotheses

Here I examined correlations between the valence of in-game interactions and estimates of player age and anti-social tendencies in the massive online battle arena game League of Legends. I define anti-social tendencies as being a propensity to engage in behaviour that breaches societal norms and which is likely to cause offense to a large proportion of people. Because I harvested a large, anonymized dataset this study is correlational in nature: I used two pieces of data extracted from usernames and used them to make estimates about the players' real-world attributes. I then describe how these estimated variables correlate with in-game behaviour as assessed by the game-based reporting system. I discuss methodological issues relating to the accuracy and meaning of these inferences in detail at the end of the chapter.

In LoL (Figure 2-1) players join small, competing 'teams' that proceed to challenge each other for territory and an objective (overtaking the enemy base) in a relatively short time

period (typically < 1 hour). The precise details of the game are beyond the scope of this paper but there are abundant descriptions in online sources (“League of Legends,” 2013). LoL is currently one of the most popular video games on the planet with an estimated 27 million online players every day (Gaudiosi, 2012). There are regular professional LoL tournaments with prizes worth millions of dollars and top players are eligible for US “internationally recognized athletes” visa status (Blake, 2013).

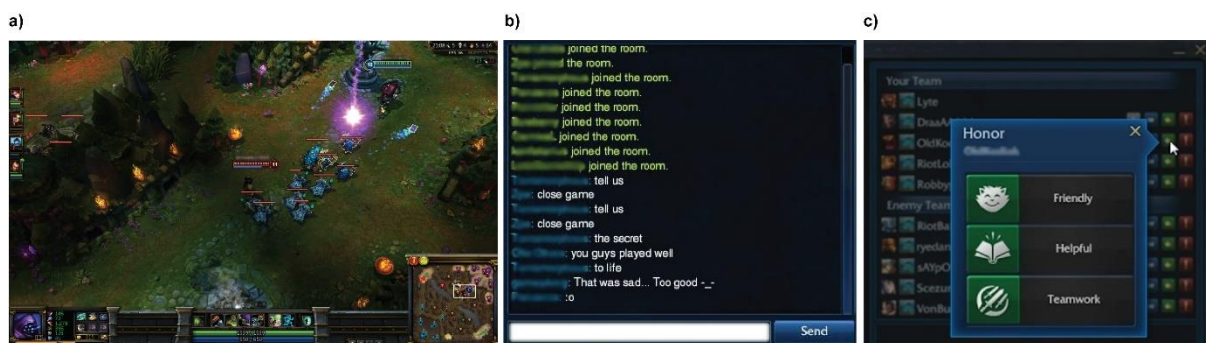


Figure 2-2. Gameplay within League of Legends. a) A screenshot of a LoL match in progress. A small portion of the playing arena is shown (illustrated in the small inset box, bottom right). Individual player ‘summoner names’ or ‘usernames’ appear above the human-controlled characters along with a health indicator. b) In-game chat. Players are able to communicate with each other both during and after a game. c) Sending negative and positive reports is possible after each game ends. Here, a player is choosing to send a positive ‘Honor’ report about a teammate.

LoL players communicate through a real-time chat facility. This facilitates coordinated game play but it also allows players to interact socially. Players are also encouraged to evaluate their teammates at the end of each game. For example, players can praise each other for their teamwork or friendliness by sending virtual tokens named ‘Honor’.

Alternatively, they can submit 'Reports' chastising other players for deliberately playing badly or sending abusive messages through the chat system. This report system allows us to gather information about the average valence of each player's interpersonal interaction within the game environment. I hypothesized that if players' real-world personality types predict their behaviour within the game, the valence of these interactions might correlate with factors that are related to real-world behaviour. Two such factors are players' ages and their tendency to use foul or offensive language in their public usernames (DeWall, Buffardi, Bonser, & Keith Campbell, 2011; Holtzman, Vazire, & Mehl, 2010) – their 'anti-social naming tendency' (ANT).

I analyzed players' self-chosen user names to estimate both age and ANT. Many of these user names contained information that informed us about these parameters. Specifically, players often embed their birth date in their user names (e.g. randomplayer1996) and in a separate analysis I show that these dates are highly-correlated with the self-reported ages of the players in the registration procedure; past research has shown that individuals have a preference for numbers that match their birthdate (Kitayama & Rarasawa, 1997). This has more recently been replicated by Jones and his colleagues (2002) where they replicated this effect and they supported that the selection of these numbers are more than just exposure effects. Thus it is possible that when they self-generate their username they append their real birthdate in their nickname. In addition, many usernames contain explicit or lightly obfuscated expletives, racial slurs and boasts that are clearly designed to attract attention (e.g. 'g0ats3x'). Players must invest some time in generating these ANT names as multi-player online games typically have simple filters in place to block straightforward examples of offensive language.

Once I had identified user names that appeared to contain either age or ANT information, I asked if there was a relationship between ANT or age and the average valence

of reports that each player sent or received within the game. I found that both age and ANT are predictive of in-game interaction valences as measured by honors and reports.

Importantly, I find this effect for both incoming and outgoing ratings (in other words, ratings generated by a player and directed towards other teammates or, alternatively, ratings generated by teammates directed to a player).

2.3. Methods and Materials

2.3.1. Data sources

Data were provided by the US-based company Riot Games (Santa Monica, CA)—the creators of League of Legends. To optimise internet connectivity, Riot Games maintains servers around the world dedicated to particular geographic regions. The data described here were obtained from servers based in North America (NA), Western Europe (EUW), North Eastern Europe (EUNE), Turkey (TK), and Brazil (BR). Riot Games supplied a representative, random sample of 450,000 data sets—one for each player. This large dataset comprised of 100,000 players on each of the NA, EUW, EUNE and BR servers and 50,000 players from the TK servers. The data represent a snapshot of the accounts on the different servers on June 13, 2013. All accounts in the dataset had been created after November 1st, 2012. The number of data sets was chosen to be as large as possible while still remaining computationally tractable.

Our analysis of anti-social user names was based on data from just the NA server (allowing us to identify English language epithets). Our age analysis was based on all available datasets. Strict controls were imposed of the type of data that were analysed. Data were collected and analysed in accordance with guidelines from both the Association of Internet Researchers (Markham & Buchanan, 2012) and the American Psychological Association (Kraut et al., 2004). It is important to note that only anonymized data sets were

analysed. I had no access to personal identifying information and no modification of players' online experience was performed as a result of this research. All players had agreed to Riot's Terms and Conditions as part of the LoL registration procedure and these explicitly allow LoL to use their data for research purposes. All procedures described in this paper were approved by the University of York Department of Psychology ethical review board.

2.3.2. Interaction valence

After each League of Legends match, players are allowed to generate feedback on the behaviour of other team members via a point and click interface (Figure 2-1c). Feedback can be positive ('honor') or negative ('reports') and can refer to a range of predefined behaviours (for example, 'Verbal abuse - Report', or 'Teamwork – Honor'). A single click on each of the feedback buttons generates a single instance of a report. Players can 'honor' or 'report' multiple team members at the end of each game but can only send a single feedback (either positive or negative) to each player. The accumulation of negative or positive reports can have consequences to a player. For example, large numbers of negative reports may lead to temporary or even permanent suspension of the player's account. Riot now implement a 'tribunal' procedure that allows other players to vote on these types of punishment and I note that the statistics of these tribunal events provide another rich dataset that may also relate to player personalities (Blackburn & Kwak, 2014).

Although both positive and negative evaluations are nominally assigned to specific categories, in reality the nature of the infraction is sometimes unclear. For example, reporting a player for "intentional feeding" implies that they are deliberately playing poorly to benefit the opposing team but some aggressive players will use this accusation indiscriminately against anyone they consider to be inferior to themselves or to vent frustration when their

team loses. There is also evidence that perceptions of toxic behaviour vary somewhat across cultures and geographic domains (Kwak, Blackburn, & Han, 2015).

Because I was interested in the overall valence of player behaviour, I used the mean of the combined ‘report’ and ‘honour’ metrics as scalar representations of negative and positive interaction. Outliers were removed using a robust outlier labelling heuristic (Banerjee & Iglewicz, 2007; Hoaglin, Iglewicz, & Tukey, 1986) which typically removed fewer than 1% of the data points. Report and Honor values were divided by the total number of games played and then log-scaled. Rates, rather than absolute levels were used to avoid conflating number of games played with average levels of anti-social or altruistic behavior. The log transform was important to ensure that data distributions were approximately normal and therefore amenable to parametric statistical analysis. The resulting datasets were found to have equal variance as assessed by Levene’s statistic.

Because the number of samples in each group was very high, these distributions were still found to be non-Gaussian by standard tests (Kolmogorov-Smirnov and Shapiro-Wilk tests; $p < .001$ in both cases) but inspection of Q-Q plots indicated relatively minor deviations. For the sake of completeness, I performed both parametric (ANOVA) and non-parametric (Kruskal-Wallis – with p-values indicated by ‘KW’) tests on our datasets and the results were found to be almost identical.

2.3.3. Antisocial names

A script containing a lexicon of common swearwords, slurs and sexual epithets as well as attention-drawing words and simple alphanumeric variations was created in MATLAB (Mathworks, MA). The list of words was derived initially from an online list (<http://www.noswearing.com/dictionary>). Additional common epithets and attention-seeking words were added by experienced game players and alphanumeric variation of the words

(e.g. “g0ats3x”) were also added algorithmically because players often use them to by-pass filters (Blashki & Nichol, 2005). Because I used databases of English language epithets, I restricted our search to data from the North-American Server (100,000 names). The full list of substrings used to identify antisocial names is provided in the supplementary material online (<https://ars.els-cdn.com/content/image/1-s2.0-S0747563215301655-mmc1.docx>). This list of target words was not exhaustive but it nevertheless identified over 2000 antisocial names from the North American server. I asked whether mean Honor and Reports sent and received (four statistics in total) were different between players with ANT and the control group. To avoid issues of multiple comparisons resulting from performing four separate t-tests, I used a standard one-way ANOVA to evaluate the statistical significance of pairs of group differences. ANT data were compared to an equal-sized random sample of players with non-antisocial names extracted from the same server. Statistical analysis was performed in SPSS (SPSS IBM, New York, U.S.A) and Matlab (Mathworks, MA). Parametric means testing is generally robust to small deviations from normality when sample sizes are large and equal—as our datasets were (see above). Because significance depends on group size, I also include measure of the raw effect size in our analysis. The descriptive statistics for ANT vs non-ANT data are shown in Table 2-1.

Table 2-1.

Descriptive statistics from ANT and random non-ANT players.

		N	Mean	Std. Dev	Std. Err	95% Conf Interval		Min	Max
						Lower	Upper		
Log(Rep Received)	ANT	2198	-1.707	.473	.010	-1.726	-1.687	-3.371	.572
	Random	2198	-1.838	.472	.010	-1.857	-1.818	-3.083	.166
Log(Reports Sent)	ANT	2198	-1.816	.537	.011	-1.838	-1.793	-3.373	.286
	Random	2198	-1.902	.555	.012	-1.925	-1.878	-3.729	.079
Log(Honor Received)	ANT	2198	-.973	.417	.009	-.990	-.955	-2.538	1.314
	Random	2198	-.910	.441	.009	-.929	-.892	-2.093	1.623
Log(Honor Sent)	ANT	2198	-1.314	.706	.015	-1.343	-1.284	-3.560	1.729
	Random	2198	-1.184	.752	.016	-1.216	-1.153	-3.604	1.975

Note. Values are log-scaled means of incoming and outgoing ‘Reports’ and ‘Honor’ feedback for each player.

2.3.4. Age

Age data were extracted from all servers. Years are conventionally indicated using either four (e.g. 1987) or two (‘87) digits. Consequently, an automated script identified dates within an appropriate 2- or 4-digit range (1985 – 2002) at either the beginning or end of the nickname (“1987Nickname”, “Nickname87”). Because I was interested primarily in developmental changes up to adulthood, and because statistical tests on very small groups are unreliable, subjects over the age of 20 were not included in our analysis.

Clearly, not all instances of two or four digits matching a ‘year’ template actually indicate players’ birth years. To examine this, I obtained an additional dataset: the years of birth reported to Riot during the game registration procedure. These represent an independent,

noisy estimate of player age. Ultimately, I obtained a total of 11,630 players who passed all the criteria with a mean age (estimated from the usernames) of 15.9 years.

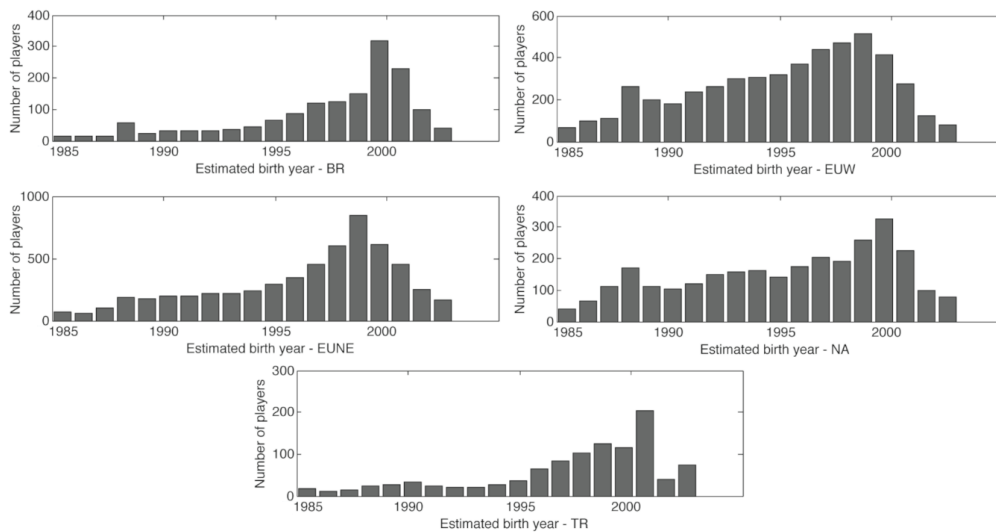


Figure 2-2. Histograms of player ages estimated from usernames from five servers. These servers correspond to the following geographical locations Brazil (BR), North America (NA), Western Europe (EUW), Nordic and Eastern Europe (EUNE), Turkey (TR), Data are shown between estimated birth years 1985 and 2003 inclusive to illustrate the full shape of the distribution.

I performed two separate analyses based on player ages. For the correlational analysis in Figure 2-3 (where I compare reported vs extracted dates of birth to assess the reliability of name-derived age estimates) I deliberately excluded ages less than 14 (year of birth 2000) from the final analysis. I did this to ensure that our results were unlikely to have been skewed by players lying about their age deliberately to pass the registration stage (Riot imposes a nominal minimum age of 13).

Age estimates from the two independent sources (usernames and registration) were correlated. Figure 2-3 shows a joint histogram of ‘name derived’ vs ‘reported’ ages for a total

of 10,299 players whose birth years lay between 1985 and 1999. The areas of the circles indicate the relative number of players that fall into each year.

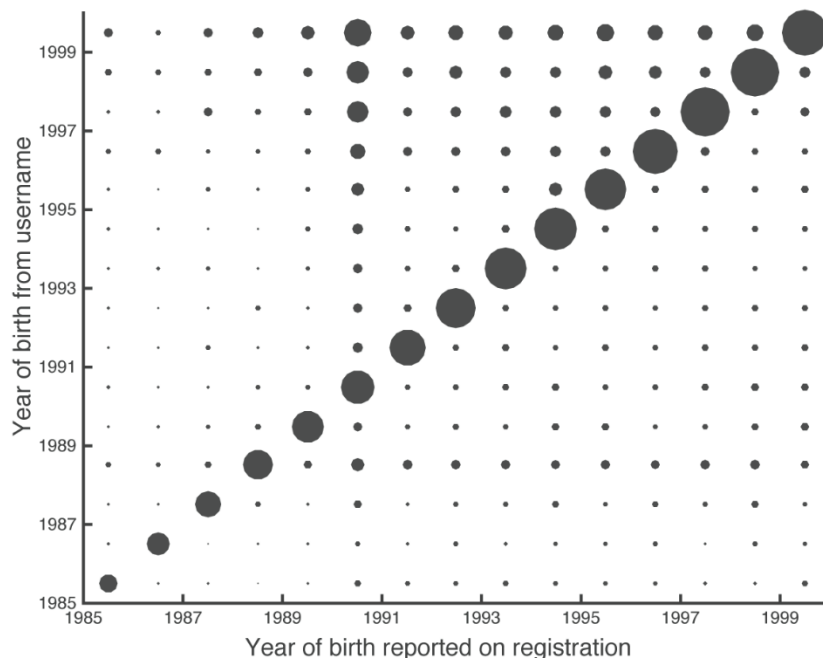


Figure 2-3. Adapted from Kokkinakis et al., (2015). Joint histogram of age estimates extracted from two different sources (N=10,299) with birth years between 1985 and 1999. The area of each circle represents the number of players in each group. Age extracted from alphanumeric usernames correlates strongly (Pearson's $r=0.60$) with age entered in the registration procedure.

The data clearly show that many players use the same age in both their user names and during registration and I find a statistically significant correlation ($p<.001$) between the two measures with a medium to strong effect size (Pearson's $r=.53$, Spearman's $\rho=.51$). My colleagues and I noted some interesting phenomena: The first is that the number '88' is an outlier in terms of its frequency of appearance in usernames. I believe this is likely due to its cultural significance (see Discussion). Players also tend to over-report the birth year 1990 during registration and there is a particularly strong correlation between players who report a

birth year of 1990 and use the digits '2000' or '00' in their username. The reason for this is unclear but when this report year is omitted from the analysis the effect size for the correlation between reported and extracted date of birth increases to $r=.6$, $p=.58$ ('strong').

In the main analysis examining the relationship between age and interaction valence, I included the full set of 11,630 players with estimated ages as young as 11 (birth year 2002) because ages were estimated solely from username information which is not vetted. Players therefore have no reason to 'lie' about their date of birth in their usernames.

2.4. Results

2.4.1. Antisocial names

Out of the 3,229 hits in the North American Server, 1031 nicknames were rejected as false positives after visual inspection by an expert English speaker who was blinded to the statistics associated with each name. For example, there would be nothing deliberately anti-social about the name "ThePenIsMightier" despite it generating a hit in the swear word dictionary lookup. After false positive rejection, I obtained a sample of 2,198 users whose names were unequivocally designed to be anti-social – generally containing blatant racial, sexual or scatological epithets. An equal random sample of 'control' (non-antisocial) names from the North American server was selected for comparison so that the ANOVA was operating on groups of equal sizes. Human inspection of the control group identified no 'false negatives' or missed incidences of antisocial names.

I found that players with antisocial names had significantly higher *sent* ($F(1,4394) = 27.31$, $p<.001$, $\epsilon^2=.0064$, $r=.08$, $KW p<.001$) and *received* ($F(1,4394) = 84.2$, $p<.001$, $\epsilon^2=.02$, $r=.14$, $KW p<.001$) Report rates compared to the control group, reflecting an increase in anti-social behavior. They also had significantly lower *sent* ($F(1,4394) = 34.517$, $p<.001$, $\epsilon^2=.0081$, $r=.09$, $KW p<.001$) and *received* ($F(1,4394) = 23.11$, $p<.001$, $\epsilon^2=.0049$, $r=.07$,

$KW p < .001$). Honor rates, indicating a reduction in altruistic or prosocial behavior. These differences are illustrated in Figure 2-4.

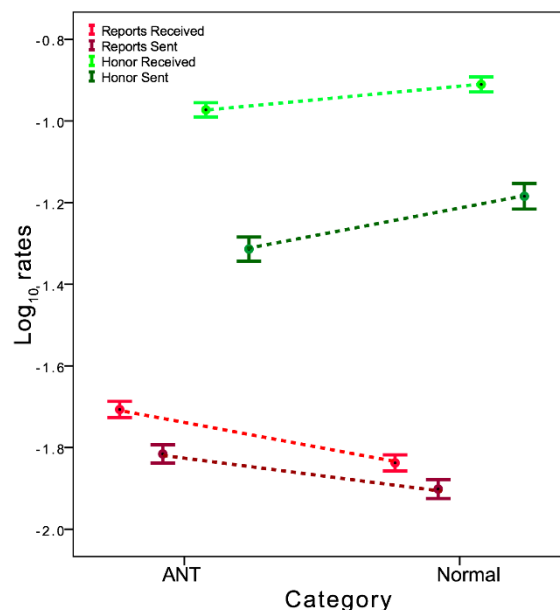


Figure 2-4. In game metrics for the two categories. This information is identical to Table 2-1 and is presented here again for visualization purposes. There are significant differences between players with 'normal' and 'ANT' nicknames in all the traits I examined. Players with antisocial nicknames tend to have higher levels of negative incoming and outgoing interactions ('reports') and lower levels of positive interactions ('honor'). All differences are significant at $p < .001$. Error bars are ± 1 SEM.

Overall, our control group sent and received positive 'Honor' at a rate that was 25% higher than that of their antisocially-named peers. Similarly, antisocial-named players sent and received negative 'Reports' at a 25% higher rate than controls.

2.4.2. Age

I found a significant relationship between age and online interaction rates. First of all, I note that overall, all games generate around 8 times as many positive interactions as negative ones despite the fact that there are slightly more categories for negative compared to positive

interactions. It is important to note that there was no difference between the average number of games played at different ages.

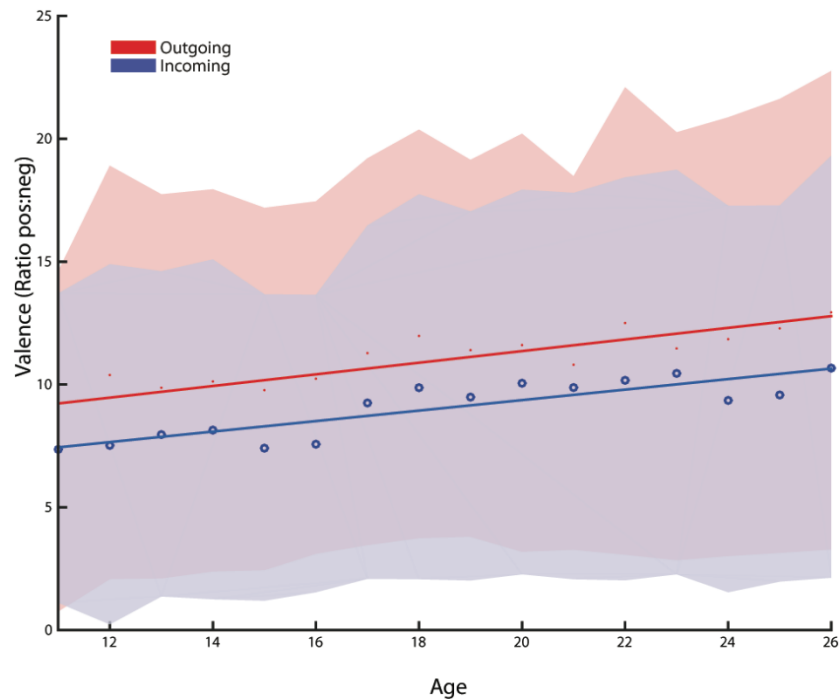


Figure 2-5. Adapted from Kokkinakis et al., (2015). The valence of interaction rates changes with age. The ratio of positive to negative interactions increases approximately linearly as a function of age. The shaded areas represent 95 % confidence intervals.

Our ability to predict changes in the overall behaviour of a particular age group is quite strong. A linear regression model fitted to the average ratio of positive to negative interactions (the ‘valence’ of overall interactions) gives an excellent fit ($R^2 = .8, p < .001$) – as shown in Figure 2- 5. On average therefore, player behavior within LoL games experiences a slow, significant and linear increase between the ages of 11 and 26. This effect is seen equally strongly in the valence of both incoming and outgoing interactions.

2.5. Discussion

2.5.1. Summary

Although there is evidence from questionnaire-based studies that personality types are reflected to some extent in online game interactions (Worth & Book, 2014b) and even in email addresses (Back, Schmukle, & Egloff, 2008), I ask here whether psychologically interesting information could be obtained purely from a large, anonymized gaming dataset. I chose to examine two game-independent attributes associated with individual players (age and antisocial tendencies) because information relating to both of these can be estimated from a single, publicly displayed data string chosen by the players themselves.

Naturally, these data are not perfect reflections of real-life player attributes. Numbers, for example, may reflect culturally significant digits rather than years. For example, '88' is a culturally laden number with Chinese speakers where it represents good luck. In addition, older players may attempt to appear younger to mislead other players with regard to their expertise and younger players may attempt to appear older to gain status. Nevertheless, our comparison of two independent estimates year of birth age (Figure 2-3) suggests a strong correlation between ages extracted from usernames and those provided as part of the registration procedure.

It is possible that players choose user names that reflect a personality that they choose to adopt within the game rather than one that matches their own real-world personality. I find this plausible to some extent (video gamers are, after all, playful) but the extreme nature of some of the obscene usernames makes it unlikely that they are chosen by pro-social individuals even as a form of escapism. Moreover, the studies on Implicit Egotism and the name letter effect in combination with the age results are particularly encouraging in this respect as they correlate well with registration data showing that players are less likely to systematically choose alternative numerical data codes to propagate an alternative online personality although I am aware that certain numbers can be used to advertise an affiliation

with extreme political beliefs (for example the number 88 also has significance within the culture of far-right Nazi sympathizers).

2.5.2. Antisocial nicknames

Although the actual usernames cannot be reported here for reasons of privacy, they lie well outside the adult societal norms and there can be little doubt that they are specifically designed to shock or draw attention from other players. Although I have no other psychological information about the subjects who choose these names, it is plausible that they indicate real-life antisocial or attention-seeking tendencies.

I found a set of correlations that link these potential antisocial tendencies to the rate and valence of player-player interactions but correlation does not inform about causality. It is tempting to associate report and honor rates with performance and behaviour within the game (since this is the overt purpose of these metrics). By this account, antisocial naming tendencies are associated with antisocial gameplay leading to higher received report rates and lower received honor rates. But equally, it is possible that players with antisocial names receive negative reports solely because those names antagonize other players. In this context, I believe that the 'sent' metrics are particularly interesting because the ANT players themselves initiate these interactions. I found that players with ANT criticise their teammates more and praise them less than controls. In this case, the ANT names are unlikely to *cause* the negative valence of the interactions. Rather, both interaction metrics seem to reflect the underlying personalities of the players.

One intriguing possibility is that antisocial names are used to express affiliation to a particular group (or to differentiate players from their teammates). In this sense, the increased negativity associated with ANT players may be framed in terms of in-group and out-group

behaviour with non-ANT players being more ready to punish and less ready to reward ANT players and *vice-versa*.

A final possibility is that the variables I examine are related through a third 'hidden' factor. For example, I considered the possibility that ANT players tend to perform worse than controls for other reasons and that their antisocial in-game behaviour was a result of this poor performance. Support for this hypothesis comes from recent studies showing that in-game antisocial behaviour is related to losing games (players who lose games are more likely to trade negative reports with their team mates) (Breuer, Scharkow, & Quandt, 2015). A full analysis of this type is beyond the scope of the current chapter but I did examine this possibility in general by comparing the Match Making Rank (MMR) scores of ANT and control players: a proxy for player success. I found a very small, (but statistically significant: $p < .001$) *increase* in ANT MMRs compared to controls suggesting that increased failure levels per se were unlikely to account for the reduction in interaction valence that I measured for ANT players.

It should be noted that after the original publication of the article that that is partially used in this thesis chapter, Riot examined the chat logs and the Reports that its ex-employees received. They identified that approximately 25 % of those fired showed a disproportionate number of toxic/negative behaviours, namely passive-aggressive comments as well as authoritative tendencies against normal players (abusing their Riot affiliation in their nickname to force people to play in a certain way) ("Riot Games: Assessing toxicity in the work place", n.d.). This has lead Riot to ask for players' in game IDs so they can screen in-game chat logs for any display of negative/toxic behaviour. Thus in-game behaviour should not be always be viewed as separate, since it has real life and ecologically predictive applications.

2.5.3. Age

I found significant changes in all our interaction metrics as a function of age. In summary, players become more pro-social as they age: negative interactions decrease and positive interactions increase (for additional analyses on the difference among the metrics and the full paper see Appendix B). The effect is small at the individual level but extremely robust and significant at the group level. Adolescence is a period characterised by significant changes in important brain structures (amygdala, frontal lobes) that govern decision making (Galvan et al., 2006; Giedd et al., 1999). The faster maturation of the limbic system, when compared to that of the frontal lobe structures, may make adolescents more prone to react to emotionally salient situations/stimuli even when their logical reasoning is intact (Casey, Jones, & Hare, 2008; Gardner & Steinberg, 2005; Steinberg, 2004; Steinberg et al., 2009) thus driving the overall higher level of negative interactions in younger players.

Again, it is possible that the names (which embed the age data), rather than the behaviours are causal: older players may bully younger players during game play, thereby leading them to resort more to negative reporting as a retaliation strategy. Very young players may have played fewer games than older players and therefore be unaware of the societal norms within the game or become frustrated by playing against more expert opponents.

These factors are unlikely to explain the trends I observed. The data I examine here consist only of players with accounts opened in a relatively short time window between November 1st 2012 and June 13th 2013. All players therefore had approximately equal experience with the game and there was no effect of age on the number of games played. Very young players are in the minority - only 14% are less than 14 years old for example and the correlation between negativity and age is weak at the low age range, becoming stronger within the 14-27 age group. This supports the hypothesis that negative interaction rates reflect

age-dependent cognitive changes in the players rather than a reaction to out-group discrimination based on their apparent age.

The increased ratio of negative to positive interactions in younger players may be due to the reduced cognitive control present in this age group. For example, Dreyfuss and his colleagues (2014) found increased sensitivity to threatening stimuli in adolescents, especially males which are the main demographic of LoL, even when they were instructed to ignore them. Thus it is possible that adolescents are unable to inhibit possible threatening stimuli leading to communicational escalations. The stimuli could relate to in-game events (for example getting killed), or to social interactions (for example, being criticised by another player).

The change in interaction valence could also be attributed to increases in Agreeableness/Benevolence with age; a trait related to cooperation as well as to the attribution of hostile intent to other agents' actions. Young people are more prone to misjudge a neutral message as a hostile one ((Digman, 1997; Klimstra, Hale, Raaijmakers, Branje, & Meeus, 2009; Van den Akker, Deković, Asscher, & Prinzie, 2014) and a similar pattern has been observed in studies looking at both proactive and reactive aggression in young adolescents (Fite, Colder, Lochman, & Wells, 2008; van Bokhoven et al., 2006). Thus, in the context of a highly demanding competitive match, an otherwise neutral chat message could be misconstrued as offensive leading to increased reporting.

Age-dependent changes in interaction valence may also be driven by changes in cognition as well as personality (Blakemore, 2008). For example, according to (Dumontheil, Apperly, & Blakemore, 2010) adolescents commit more errors in a Theory of Mind task (ToM), when compared to adults while other studies have shown that tasks requiring ToM activate brain networks similar to those involved in empathy and forgiving (Farrow &

Woodruff, 2005); Hayashi et al., 2010; Strang, Utikal, Fischbacher, Iber, & Falk, 2014).

Dumontheil and colleagues (2010) concluded that the interaction between ToM and executive functions is still developing in late adolescence and I hypothesize that this is a factor in the slow increase in the valence of the interactions that I observe over age because younger players are unable to contextualize the actions of others correctly and may misattribute actions (such as accidental poor play) to a deliberate threat or collusion ('Intentional Feeding').

There is some anatomical basis for the changes in impulsivity and risk-taking seen in adolescence. A dominant theory is that the developmental trajectories of subcortical structures involved in reward (for example, the nucleus accumbens) are faster than those of more frontal cortical regions providing inhibition and cognitive control (Casey et al., 2008; Dreyfuss et al., 2014; Galvan et al., 2006). Again, this hypothesis predicts a slow but steady increase in pro-social behaviour and a decrease in impulsivity across the time frame covered by our data.

In the context of cognitive development, adolescent deficits in ToM might also be enhanced because they are deprived of valuable information such as facial and vocal cues which are an important source of information about other players' motives and emotions (Achim, Guitton, Jackson, Boutin, & Monetta, 2013).

Finally, an alternative possibility is that cognitive and behavioural difference are inherent to different birth cohorts rather than different ages per se. In other words, the increase in in-game antisocial behaviour that I observe in younger players will remain constant as those players become older: The millennials are simply more antisocial than those born before the turn of the century. Evidence for this hypothesis is mixed – largely because of the difficulty in performing well-controlled personality experiments spanning multiple

generations. Recent work by Twenge et al., (Twenge & Foster, 2008, 2010) suggests that millennials score higher on at least one antisocial personality trait (Narcissism) than age-matched cohorts from previous generations but this result has been disputed on methodological grounds and other researchers studying similar datasets indicate that any effect that may be present is very small and that a measure strongly related to narcissism (“self-enhancement”) is stable across birth cohorts. At the moment, therefore, I believe that the most parsimonious explanation for our data is based on a developmental change in personality across adolescence rather than a systematic difference in pre- and post-millennial birth cohorts.

2.5.4. Conclusions

Our data show that video games can provide a wealth of useful population-level information on developmental cognitive and psychological processes. Although the individual data points may be noisy, the overall conclusions are highly robust due to the sheer number of subjects. Similar analysis techniques have been used to examine the relationship between practise and performance in a custom-built online game as well as in MMORPGs (Stafford & Dewar, 2014) but I believe I am the first to examine player-player interactions in a-MOBA game using this methodology (Drachen, Sifa, & Thureau, 2014; Guitton, 2010).

It is intriguing to ask if other subclinical psychiatric disorders such as autism, sociopathy or addictive personality traits might be evident in these types of data. For example, since personality influences responses in experiments probing economic choice (Berg, Lilienfeld, & Waldman, 2013), can the same results be observed in video-games? Campbell and his colleagues supported the notion that in a classic “tragedy of the commons” game, where the individual needs to exert self-discipline and harvest a limited amount of the resources in order to allow for the continuous survival of all the players, the optimal strategy

at a group level requires players to delay reward (Campbell, Bush, Brunell, & Shelton, 2005). Here, I expect players who have limited abilities to discount immediate gratification to have a stereotypical profile in complex online games such as LoL, which may alter long-term, in-game success rates both for themselves and for other team members. I will further pursue Inhibition Control and psychometric Intelligence in the following chapter concentrating on Matchmaking Rating, a measure of how good one is at winning. It should also be noted that positive in-game behaviour such as rapid learning, team building or leadership might correlate both with positive usernames and with positive personality traits in the real world.

Finally, I have assumed here that real-world personality attributes are the *cause* of the online behavior patterns I observe. But it is possible that strategies learnt in the online environment may also provide cues to appropriate (or successful) behavior in the real world. It should be noted that after the completion of this article Riot individually experimented with changing the negative nicknames to see whether something randomly generated or even positive could have a different effect on Reports and Honours. For instance, someone who was previously named “Rengar’s Penis” might have had their name forcibly changed to “Purple Manatee” or “Nice Manatee”. However, they have not released any results from this experiment (“r/leagueoflegends - What’s In A Name?,” n.d.).

Video game training alters a wide range of visual, cognitive and attentional mechanisms (Adachi & Willoughby, 2013; Appelbaum, Cain, Darling, & Mitroff, 2013; Boot, Kramer, Simons, Fabiani, & Gratton, 2008a; Granic, Lobel, & Engels, 2014; C Shawn Green & Bavelier, 2003; Li, Polat, Makous, & Bavelier, 2009) and regimes emphasizing different strategies within the same game can lead to changes in real-world behavior (Greitemeyer & Osswald, 2010; Yoon & Vargas, 2014) and cortical activation patterns in subsequent test periods (H. Lee et al., 2012). It has been suggested that the remarkable

plasticity evidenced in such studies is due in part to the highly arousing nature of the games themselves (Bavelier, Levi, Li, Dan, & Hensch, 2010).

2.6. Limitations and Disadvantages

These experiments have some limitations imposed by the nature of the metrics used. In the “old” League of Legends report system there was an additional category specifically targeted towards individuals with offensive names called “Inappropriate Name”. This category was not included in the analysis since it was not given to us by our industry partners; even if given however, I would still not include it since it could be argued that it would “inflate” our results and it would not count as a “true” negative in-game behavioural metric. However, Riot have now changed this system: currently, players can only report an individual for a *single* offence. Thus, if an individual had an extremely offensive name (for instance “Hitler_had_the_best_KDA69”) and, he/she displayed a highly negative attitude in-game, for instance by using racial slurs or by harassing individuals, many players would later report the individual for having an “Inappropriate Name” thus essentially deducting a report from the other categories.

The second problem that arose is that Riot used an all-or-none/binary report system. This means that we cannot infer the magnitude of the behaviour, since it does not conform to a Likert Scale or any large scale for that matter. Thus the “simplistic” report category “Harassment: Offensive Language” loses a significant amount of meaningful information when a single report could be triggered by both a simple chat taunt or by multiple death-threats (e.g. “Easy game, Easy Life” vs, say, “Malphite you are a f***ing worthless s***f***k b*****d pile of trash mental d***face that should be gunned down in the street like the degenerate you are”). While Riot did have access to the chat telemetry that accompanied each report, these data were not provided to us.

Future studies could focus on the in-game chat that was the reason of the report. Do individuals with offensive names have “lower lows” when they “rage” (lose their temper) when compared to the average player or do they show a more continuous “negative attitude”, which could be categorised as trait hostility, throughout the game or even before the match starts, at the champion select? Natural Language Processing could be employed to identify key in-game events (for instance the ANT player’s death) and cross-compare that to an average person. An additional event that may yield results due to the potential to “trigger” an ANT is in-game criticism of a play event (Zagal et al., 2011). It could be that if ANT have a higher form of trait hostility they will respond more negatively even to the most innocuous comments or suggestions. Finally, even though there is a double dissociation between reports and commends, researchers should examine if ANTs are treated differently due to their names, even if they are actively trying to change or if they are not even harassing (albeit this is improbable since some of the names are highly offensive and aggressive to the reader). It is possible that even if these individual actively try to change their behaviour their name traps them in a loop since they are treated as worse by their teammates (nominative determinism).

3. Videogame Rank and its relationship to Age and Fluid Intelligence

3.11. Chapter Introduction

In the previous chapter I focused on how nicknames can be used to extract meaningful information about in-game social behaviour as well as real world demographics such as Age. I subsequently examined Aging and Interaction Rates (Reports-Honour Tokens) within the game League of Legends. In this chapter rather than only using a pre-existing dataset I will perform my own experiments in a Laboratory setting which I will subsequently complement with datasets that are either publicly available or that I obtained by contacting our industrial partners or co-authors. Moreover, the new variables that I will try to link will not just be simple demographics (e.g. Age), rather they will be more complex and have higher utility or real-world applications.

The new key real-world variables that I am going to obtain an in-game proxy of, are: Visuospatial Intelligence, as measured by the WASI-II Matrix Subtest, as well as Working Memory measured through three standard Complex Span Tasks. The key in-game variable that correlates with the aforementioned concepts is Matchmaking Rating (MMR) or how good one is at winning matches with four *human* teammates against five *human* opponents (5 vs 5). When one wins a game their MMR increases, similar to ELO in chess, making their place in the competitive ladder go up. This means that they will subsequently face 'harder' to beat opponents. I will expand on this later in the chapter.

I will additionally show that MMR behaves in a similar manner as fluid Intelligence across one's lifespan by using a cross-sectional sample; MMR shows a peak in approximately one's mid-twenties similar to Cattell-Horn-Carroll model of Intelligence (Horn & Cattell, 1996; McGrew, 2005; Kaufman & Horn, 1996). I will subsequently replicate this with another game of the same genre (Dota 2) while showing that players' performance (MMR) in two

First Person Shooters (FPS), namely Destiny and Battlefield 2, show a differential age pattern, suggesting that they rely more on reaction times rather than problem solving ability effectively acting as “control” videogames; at least at the period the samples were obtained since games tend to evolve across time. Real-life applications of this knowledge will be discussed, namely their relation to Cognitive Epidemiology as well as the identification of gifted young children from an early age just from their videogame scores/performance.

Finally, I will elaborate on many methodological points and problems that are not obvious to the layman or even to the expert videogamer. These methodological points span a wide-range of topics such as: the selection of the battery of tests, the selection of the videogames as well as sampling among others. Real world applications in combination with future research will also be discussed.

The full article can be found in Appendix C.

3.2. Introduction

Games of strategy, such as chess or mancala, can be found across cultures and skilled performance in these games has been associated with intelligence historically (Bilalić, McLeod, & Gobet, 2007; Spitz & DeRisi, 1978; Spitz, 1978; Spitz & Winters, 1977). Spitz formalized this connection with specific subpopulations, pointing out that performance in a wide variety of strategy games such as Tic-Tac-Toe or the Towers of Hanoi can be linked to mental ability (Byrnes & Spitz, 1977; Spitz & Winters, 1977). He went on to suggest that strategy games tap a number of facets of intelligence: visualization of possible moves, short-term memory rehearsal and the ability delay immediate gratification to increase future rewards (for example, sacrificing a piece in chess in order to win the game in a later turn) (Herman H. Spitz, 1978). Later studies consolidated the link between intelligence and game performance. For example, expert chess players have above average intelligence and that the correlation between skill level as approximated by rank and IQ scores (fluid and crystallised intelligence measurements) explains up to 30% of the variance (Burgoyne et al., 2016; de Bruin, Kok, Leppink, & Camp, 2014; Grabner, 2014; Grabner, Stern, & Neubauer, 2007).

This link was later extended to the domain of video games by Rabbitt et al. (1989) who correlated scores from the Alice-Heim (AH-4) IQ test with performance in ‘Space Fortress’; an arcade-like single player game developed by psychologists (Donchin, 1995; Mané & Donchin, 1989; Rabbitt, Banerji, & Szymanski, 1989). While individual player IQs did not predict initial performance in Space Fortress, they did predict learning rates and, therefore, performance once players had engaged with the game long enough to become practised. More recent studies have suggested that IQ can be measured in a subset of simple single-player video games as well as through tasks embedded in game-like environments such as Portal 2 (Foroughi, Serraino, Parasuraman, & Boehm-Davis, 2016; Quiroga et al., 2015, 2009). In this chapter I extend previous findings linking psychometric intelligence with a specific genre with tens of millions of players that uses human opponents rather than

preprogrammed puzzles (aRTS/MOBA). I additionally show that aging and performance in these videogames closely matches to how fluid intelligence behaves across the lifespan (Kaufman, 2001; Salthouse, 2010). Furthermore, I provide insight on more nuanced metrics and methodology that are not easily seen at first glance.

3.2.1. Intelligence and Task Selection

Intelligence is a highly contentious topic surrounded by many myths and fallacies (Ackerman, 2014a; Gottfredson, 1998). The very definition of it also seems to vary a lot depending on who one asks. However Gottfredson's (1997) provides a more than adequate definition that agrees with the author's opinion : 'Intelligence is a very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience. It is not merely book learning, a narrow academic skill, or test-taking smarts. Rather, it reflects a broader and deeper capability for comprehending our surroundings —“catching on,” “making sense” of things, or “figuring out” what to do.’ (p.13).

Historically, there are many models that tried to approximate Intelligence: the Cattell-Horn-Carroll framework, the Spearman's-Holzinger's 'g' model, the Berlin model of Intelligence structure (BIS) and the Verbal-Perceptual-Rotation Model (VPR) (Bucik & Neubauer, 1996; Johnson & Bouchard, 2005; McGrew, 2009; Spearman, 1904). Moreover, working memory has been linked to Intelligence, as well as executive control in general, with certain researchers believing that they are near identical (Chuderski, 2015; Conway & Kovacs, 2013). No matter what theory one ascribes to, everyone can agree that working memory and fluid intelligence are partially overlapping cognitive traits that correlate strongly with multiple life events such as scholastic achievement and professional business success

(Allen, 2004; Alloway & Alloway, 2010; Logie, Baddeley, Mané, Donchin, & Sheptak, 1989).

The full description of these models is beyond the scope of this thesis. I will briefly discuss the influence of the Cattell–Horn–Carroll (CHC) and the VPR models on the selection of the battery of tests. I will not focus on Verbal or Crystallised Intelligence since games that are primarily visuospatial in nature are more likely to tap Visuospatial Abilities; that is not to say that game performance or MMR is not correlated positively with verbal or crystallised intelligence since we know that these games are multiplayer and good communication amongst team members is a highly valuable asset. More generally, all Intelligence subtests tend to positively intercorrelate (Jensen, 1998). Thus there might be some weak correlation between verbal intelligence and videogame performance. However, that would require a much bigger sample that is lingually uniform, excluding non-native English speakers, and that is beyond the topic of this thesis. Finally, it is possible to gain MMR or to improve one’s performance by studying in-game elements or “strategies”; this could be likened to chess players studying and memorising chess openings rather than theorising on the fly. However, the cataloguing, the development and the analysing of this topic could be a secondary PhD topic and thus I opted not to attempt it.

Before one starts describing these frameworks it should be noted that the semantics regarding *g*, fluid intelligence, visuospatial abilities vary a lot (Gignac, 2014, 2015; Salthouse, Pink, & Tucker-Drob, 2008). In his “Contextual Analysis of Fluid Intelligence” Salthouse et al., (2008) state: “*Whether this dimension is labelled Gf (fluid Intelligence), working memory, executive processing, or some form of cognitive control may reflect the research tradition of the investigator more than any fundamental differences among the concepts because it appears that individuals would be ordered in nearly the same way with variables from each of these perspectives.*” (p. 16). Thus these factors are not always clear

cut, still Salthouse as well as many other researchers believe that Working Memory is an extremely good predictor of Fluid Intelligence calling them isomorphic concepts nearly equating the two, which is further backed by their really high intercorrelations (Salthouse et al., 2008). However, having read Ackerman's position on this topic I would never fully equate Working Memory to Fluid Intelligence (Ackerman, Beier, & Boyle, 2005). Even still Working Memory is a highly robust measurement of cognitive ability that needs to be implemented in a battery of tests to have a more complete understanding of an individual. This is why I included 3 Complex Span Tasks created by Foster and his colleagues (Foster et al., 2014a). These were a Symmetry, a Rotation and an Operation Span Task (for more details please see the Methods section).

The first theory that will be discussed was developed by Cattell, based on Thurstone's factor analytic work, and divides Intelligence into two categories : Crystallised (Gc) and Fluid (Gf). He postulated that Fluid Intelligence is involved with deductive and inductive reasoning being primarily related to biological and neural factors, as opposed to Crystallised Intelligence that is related to social factors (acculturation) (McGrew, 2005, 2009).

Cattell's theory was later expanded by Horn in 1965, with the addition of the following abilities, visual perception (Gv), short-term memory (Short term Acquisition and Retrieval— Gsm), long term storage and retrieval (Glr), an auditory ability Ga, as well as speed of processing (Gs) (Flanagan, 2008; Flanagan & Dixon, 2014). Moreover, Horn started the capitalization of g (gf→Gf) since the factors became narrower when compared to their previous iteration (W. Schneider & McGrew, 2012). These two types of Intelligence have an important distinction: Gf reflects the ability to solve new tasks that prior experience is not of any use while Gc taps information that could be simplified as 'experience';

preacquired information based on one’s culture, practised skills and education among others (Johnson & Bouchard, 2005).

This theory would later further be cemented by Carroll’s “Human Cognitive Abilities: A Survey Factor-Analytic Studies” (Carroll, 1993). In this book Carroll analysed a large number of studies (over 400 datasets) leading him to assign a general factor *g* that is involved in more complex

procedures as well as many subdivisions for *Gf* and *Gc* (Flanagan & Dixon, 2014; McGrew, 2009). Carroll’s contributions to this theory were significant leading McGrew to parallelise his work with that of Newton’s (for the field of Psychometrics) (McGrew, 2005). An illustration of this theory can be seen in Figure 3-1 below.

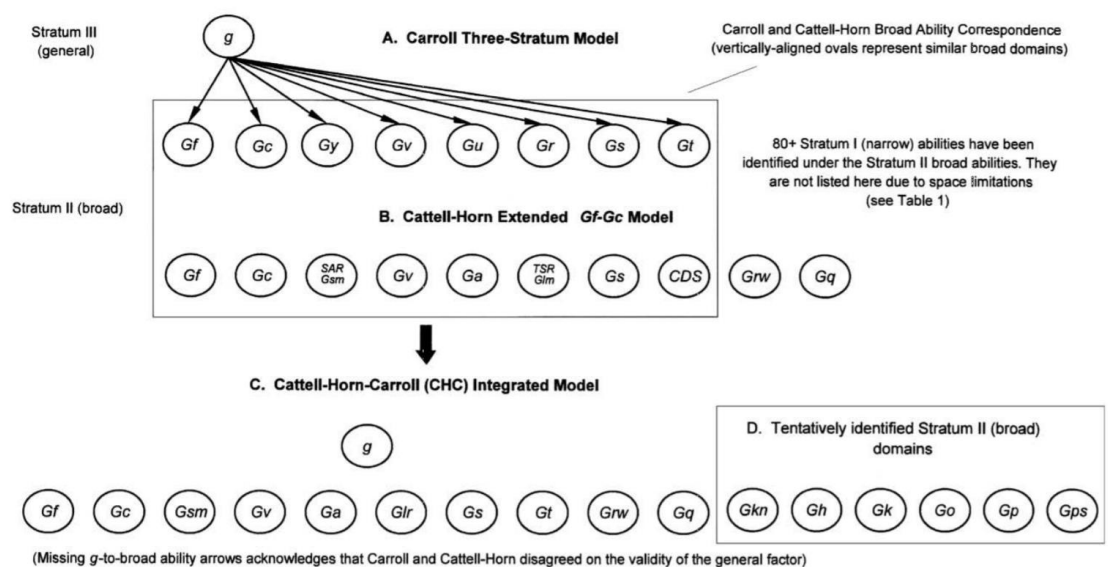


Figure 3-1. The Cattell-Horn-Carroll Framework. Different factors are subdivided into different Strata with a *g* factor presiding over all other factors at Stratum III. Adapted from McGrew (2005).

The second influential model that will be discussed is the Verbal-Perceptual-Rotation model (Johnson & Bouchard, 2005). Johnson & Bouchard (2005) used multiple theoretical models trying to identify the best fit with a wide battery of tests with subjects being Twins reared apart allowing them to make additional analyses. The three models they examined were: the original Cattell-Horn model (without the g factor), the Carroll model (g factor over Gf and Gc) and the Vernon Model. The Vernon Model is relatively old and consists of a general factor that governs all mental abilities. Once that factor is extracted, the residuals fall into two categories: 1) A verbal-educational factor (verbal fluency, divergent thinking, verbal scholastic knowledge and numerical abilities) 2) A visuospatial Mechanical factor (perceptual speed, psychomotor and physical abilities such as proprioception as well as spatial and mechanical abilities).

Their key conclusions after they tried all three fits were that none of the of the models accurately depicted the data although Vernon’s model outperformed the other two. The author subsequently created their own model after Vernon’s due to its superiority. The new model was called Verbal-Perceptual-Rotation) and it can be seen in Figure 3-2, below.

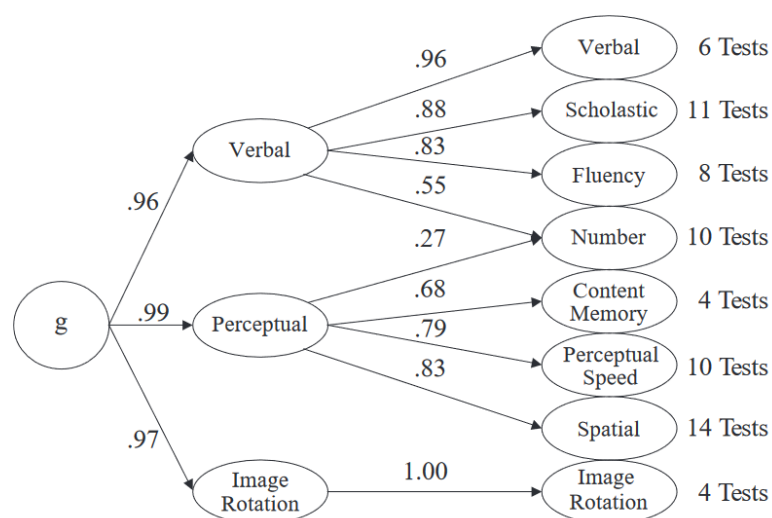


Figure 3-2. The Verbal Perceptual Rotation model. Note that Image Rotation is distinct from Perceptual abilities. Adapted from Johnson et al. (2005).

As can be seen in the graphs both the VPR and the CHC frameworks have a g (general) factor that is the distillation of the large number of tests used in those studies and presides over all abilities. However, their substrata (subdivisions) differ. In the CHC framework mental rotation is a more narrow ability (Stratum III) while in the VPR model there is an important Mental Rotation – Perception distinction. This sounds absurd at first; as Schneider and Newman (2015) mention "This is a bit like saying that there are three kinds of sports: team sports, individual sports, and badminton" (p.13). Even if it does appear as counterintuitive however certain factor analytic studies have shown that these two abilities cluster differently (Hegarty & Waller, 2004; Mix & Cheng, 2012). Moreover, this distinction between mental rotation and perceptual ability has also been backed by brain lesion studies where the patients exhibit a clear dissociation between these two abilities (Morton & Morris, 1995).

In conclusion, as previously mentioned in Salthouse's article on Fluid Intelligence, things are not always so clear cut, with each researcher having their own opinions on the topic of fluid intelligence and working memory (Salthouse et al., 2008). This is echoed by Gignac who uses gf/gv interchangeably (Gignac, 2015). Moreover, in his thesis, "Spatial Ability and g", Lohman clearly explains that visuospatial ability may not be as distinct or low-level as it is described in the CHC model (Lohman, 1996). Therefore I decided to use the WASI-II Matrices which is an acceptable alternative to Raven's Matrices and taps a broad number of constructs such perceptual differences, inductive reasoning and broad visual intelligence with its various subitems (Maccow, 2011; "Wechsler Abbreviated Scale of Intelligence - Second Edition (WASI-II) | Pearson Assessment," n.d.). In conclusion, Fluid Intelligence is a not a well defined topic and due to this lack of clarity with these semantics, I choose to use a wide range of metrics, as previously done in the literature, that encompass a wide range of cognitive abilities. Moreover, due to abstractness of the aforementioned term I

clarify that in this chapter rotational, perceptual or “deductive” abilities will be referred to as visuospatial abilities, visuospatial Intelligence or fluid Intelligence.

Aside from the aforementioned conventional Intelligence frameworks there exists the notion of Collective Intelligence or how good a group is at collaborating effectively influencing how efficiently a group task is done beyond each individual’s psychometric Intelligence scores (Woolley, Aggarwal, & Malone, 2015; Woolley, Chabris, Pentland, Hashmi, & Malone, 2010) . Woolley and her colleagues (2010) noted that Social Sensitivity correlated at $r=.26, p<.002$ with their extracted c factor (collective intelligence factor similar to g). They measured using the ‘Mind in the Eyes Test’, which is a test where the participant tries to identify a person’s Emotions just by looking at their eyes (Woolley et al., 2010). Beyond this study, Engel and his colleagues (2014) support the notion that the averaged Reading the Mind in the Eyes Test (RME) scores of a group can accurately predict their performance in multiple tasks such as a Sudoku puzzle or at a word-scrambling task. As MOBAs have grown to be highly team-based games I included this test to see if theory of mind plays a key role. My addition of this test was further supported by the fact that Baker and her colleagues ((Baker, Peterson, Pulos, & Kirkland, 2014) in their meta-analysis found that it partially correlates with performance IQ ($r=.25, p<.05$). This also allows us to run partial correlations excluding this variable social variable allowing us to focus on the other constructs.

Despite the aforementioned studies regarding RME some research shows that individual’s with autistic tendencies (Asperger’s) have increased fluid Intelligence and Visuospatial working memory (Caron, Mottron, Rainville, & Chouinard, 2004; Shah & Frith, 1993). Therefore autistic tendencies may confer some visuospatial advantage that negates or overpowers ToM as measured by the RME test. This holds especially true if one takes into account the very fast nature of the game which might not allow enough time for

understanding an opponent's intentions. A large number of in-game behaviours, including the movement of characters follow a straight line or some form of Newtonian motion. Thus I included the Folk Physics test which included Newtonian and mechanical motion problems (Simon Baron-Cohen, 1997).

3.2.2. Videogame Selection

Many video games can make demands on one or more dimensions of cognitive ability. In this sense, they act to measure some of the same psychological constructs that are probed by more conventional psychometric tests. One of the first demonstrations of this was the work of Rabbit et al. looking at an early space fighting game "Space Fortress". This experiment showed that performance on this game correlated strongly ($r=.69$, $p<.001$) with a pen and paper IQ test, Alice Heim 4 (Rabbit, Banerji & Szymanski, 1989; Mane & Donchin, 1989). These correlations are further supported by recent research showing that the scores in even simple so-called "casual games" are strongly correlated with fluid intelligence, working memory and perceptual speed as measured by a battery of psychometric tests (Baniqued et al., 2013). These results are further supplemented with a study from Quiroga and her colleagues (2015), which find high correlations between intelligence tests and commercially created Wii videogames.

Thus the first issue I faced was the selection of a game. There are many genres with each genre having hundreds or even thousands of games. I did not want to use a 'reskinned' Psychological task because it has been previously done as mentioned above. Most importantly, not many (if any) individuals play these games for an extended amount of time; note that I do not doubt that a sizeable number of individuals may have played them at one point in time, as shown by their sales, however these games are not applicable for any serious

type of research that may require a constant stream of longitudinal data (in order for instance to track aging effects or performance after specific life-events). Therefore I chose not to focus on these type of 'brain' games since gamers do not keep playing these games at neither the rate nor the frequency of commercial games (as a rough metric the most popular hero in Dota 2 has been collectively played in matches for over 51 **thousand years**, while LoL has over 27 million daily users) ("Pudge - Melee, Disabler, Durable, Initiator, Nuker," n.d.; Tassi, n.d.). Moreover, many of this type of software may have been abandoned by their creators in terms of development thus they are not available in a format that is easily useable but rather they might rely on proprietary software (as opposed to MOBAs which are available with a simple download) (Ryan, 1994). Some examples that were mentioned in the literature review are Space Fortress, Memory Goblins as well as many of the live management worlds (simulation), the defunct military projects such as Star, the digit-RT games and finally the program as well as the forgotten test (C. Graham, Cook, Cohen, Phelps, & Gerkovich, 1985; J. McPherson & Burns, 2007; Ryan, 1994; Strohschneider & Güss, 1999). In conclusion, I chose the action real-time strategy genre (aRTS/MOBAS) as my object of study because it involves resource management, fast reaction times, adapting to one's opponent (inductive and deductive abilities), thinking ahead as well as population of gamers that dedicate a sizable amount of time to the game each day.

The second requirement that was needed from the videogame of choice was a huge player base thus increasing my chances of finding participants that are actively willing to be tested in a laboratory environment. This may sound quite obvious, however from my personal experience, even though DOTA 2 had a playerbase of over 450 thousand people, the players did not show up to be tested; I only acquired a small number of individuals, a total of 10, even after I advertised the experiment at multiple sites and I tried to recruit at multiple gaming events ("Dota 2 - Steam Charts," n.d.). A point that researchers may need to take

into account is that, even though DOTA 2 has a large number of players, they are likely not concentrated in the United Kingdom thus making participant selection suboptimal.

The third issue was the selection of the in-game metric. There are theoretically many statistics one can derive from strategy games. These include but are not limited to creep score, vision score, Kills, Deaths, Assists among others (“API Documentation - Riot Games API,” n.d.). There are multiple reasons why I selected Rank rather than the more traditional KDA (Kills plus Assists divided by one’s Deaths). There are many heroes or champions that have different average KDAs; for instance a champion or hero whose job is to receive a lot of in-game damage might have more deaths than a character whose job is to “clean-up” after him (scoop up enemies with low hp and deliver the final blow). Thus the total end game contribution to a victory or defeat is not so clear cut for each individual, which is why when MMR, the aforementioned approximation of one’s skill, is calculated the core variable is winrate and not individual performance (“Game stats by champion - League of Legends,” n.d.). The second reason for not selecting more advanced metrics is that the features were not available in the game’s API and they might change throughout the year as the game changes. In an elegant study Thompson and his colleagues (2014) used the multiplayer online game Starcraft 2 and analyzed data from 3360 participants after they carefully cross examined them for objective rank criteria. They found that actions-per-minute (APM) was a strong predictor of expertise (note that in MOBAS this is not always the case due to the diversification of heroes and hero roles, for instance Supports may have lower APM than Ranged Ad Carries since they do not last hit creeps). However, in a more elegant analysis they also found that Action Latency (a measurement very similar to standard psychophysical reaction times), and Perception-Action-Cycle (PAC), which is a shift of the screen followed by a sequence of actions, are both more highly predictive of a player’s expertise. From a psychological point of view this resembles Jensen’s and Eysenck’s theories about the interlinkage of inspections

times, reaction times and IQ (Bates & Eysenck, 1993; Sheppard & Vernon, 2008). PAC therefore would be a good supplement aside from Rank however that would have required either more extensive support from the REST API or some form of parsing all replay matches, which was not supported, combined with some form of computer vision. In conclusion stabilised Rank is one of the best metrics we could derive; after all one could make the case that all the aforementioned in-game statistics culminate into a Victory or a Defeat which define one's Rank.

The fourth important point is that the game should have a relative diverse ranking system where the various ranks are meaningfully different from each other. If a game has hard boundaries that do not allow for variability it would be akin to an IQ test having floor or ceiling effects; note that I do not claim that a diverse ranking system completely guards against ceiling effects, however it is a good first step; after all Kasparov who was arguably one of the best players had an IQ of 120 which signifies that there are ceiling effects in Chess even when there is a diversification of ranks (Grabner, 2014). The game's rankings should not have any hard or soft limit. To make an analogy between videogames and chess, ceiling effects would be extreme if the chess ladder had a hard limit of 1800 ELO in order to save matchmaking queue time (how long players have to wait for a match). Similarly not allowing players to fall below 1000 ELO because it will make the players feel bad and thus influence player retention or sales in general would create some floor effects ("Ranked Play FAQ," n.d.). Of course, in this scenario due to the imposition of hard limits or "caps", a significant amount of noise is introduced to the Ranking System; or at least at specific parts of the Rank Ladder. LoL has correctly introduced a form of Rank Decay in order to counter this, where if the participant does not play a game for too long their Rank starts to drop as mentioned above. Moreover, they recently introduced a new Rank below Bronze, called Iron (League of Legends, 2018). At the time of the experiment DOTA 2 had no such system, even though

they later introduced one which is based on a seasonal soft reset (“Seasonal Ranked Update | Dota 2,” n.d.).

The fifth point and perhaps one of the most important ones is that the game’s Rank should largely, if not solely, depend on individual “skill”. In MOBAs even if there are multiple agents (players), after a lot of games usually 100 to 300 as a rule of thumb, a decent approximation of peak performance can be inferred (Aung et al., 2018; “Matchmaking | Dota 2,” n.d., p. 2). This happens of course due to the noise (different teammates with different performances), however it cancels itself out over a large amount of games. Although, at the time we did not have the knowledge we have now, since Riot has given us some longitudinal data on individual performance, one can predict the Final MMR for an individual with extreme accuracy just by using the first 30 games they played which echoes Stafford’s and Dewar’s findings (Stafford & Dewar, 2014). This should of course be taken with a grain of salt since individuals have different learning rates and performance in different games stabilises at different rates based on their complexity, especially when the game includes many subtasks of different complexity (Ackerman, 1987; Bonny & Castaneda, 2017). In addition, there are lots of individual factors such as the age of acquisition of a skill or whether someone starts playing again after stopping for a sizeable amount of time (Bilalić et al., 2007; Blanch, Aluja, & Cornadó, 2015; Gaschler, Progscha, Smallbone, Ram, & Bilalić, 2014). Moreover, the type of play is important since deliberate practice and rate of play are important factor in reaching one's true "skill level" (Ackerman, 2014a, 2014b).

To clarify the fifth point above, one can achieve the same Ranking as someone else in a videogame through not Solo Matchmaking (the player getting randomly paired with another 4 random individuals) but with a non-random team where the other individuals are the player’s friends. Many games including LoL include a second type of Ranking called “Team” or “Group” Rank where the participant can play the game with their friends and

“stake” his Team Rank or MMR rather than their Solo Rank/MMR (“Matchmaking Guide,” n.d.). This can be problematic for many reasons. First, if the sample size is too small it can introduce a large amount of noise which can lead to Type 1 error. For instance, one could have teammates that are disproportionately good which could “boost” them to higher Rankings that they could not attain by themselves or vice-versa (their teammates which are also their real-life friends might be the cause they have a lower win-rate). Having acknowledged this point, given a big enough sample this should not be an issue since having cross-correlated Dota 2 Solo and Group MMR they are highly intercorrelated at around ($r=.888$, $p<.001$) (Bonny & Castaneda, 2017). For the reasons describe above, I focused on a player’s solo performance.

In conclusion if one takes into account all the above information the only game that fits the imposed criteria is LoL due to its popularity, the nature of its Ranking system and its gameplay. Do note that given a large enough sample one could obtain similar results with Dota 2 or Starcraft/Starcraft II. However, a major advantage LoL has over Dota 2 is that of the lack of crystallised information. Dota 2 has too many complex hero and item interactions and thus requires a larger number of games for one to internalise them when compared to LoL (<http://blog.dota2.com/2013/12/matchmaking/>). This makes it harder to find an individual that plays Dota 2 and has reached their final peak or stable performance. Moreover, LoL allows for “surrendering”, where team members can decide whether they can quit the match earlier with no repercussions (“Surrender at 15 available now | League of Legends,” n.d.). This allows for faster matches allowing certain individuals to play more games thus reaching their peak stable practiced state.

3.3. Study 1: Fluid Intelligence and associated measures

3.3.1. Methods

3.3.1.1. Ethics

All participants in our laboratory experiments provided informed consent and approval for the study was provided by the ethics board of the Psychology Department of the University of York. All data were anonymized and participants were informed that they could withdraw from the study at any time.

3.3.1.2. Participants

Participants (N = 56, 51 males, mean age 20.5 years) were recruited via adverts from multiple sites within the UK in and around the Universities of Leeds, Essex and York. All subjects were experienced LoL players who had played a large number (>100) of both ‘ranked’ and ‘unranked’ matches in order to reach a stabilised performance).

3.3.1.3. Rank/ MMR /Rating Extraction and Interpretation

Online videogames such as the ones examined here provide detailed telemetry to the coordinating game servers. Companies such as Riot therefore have databases of real-time information about the behaviour and performance of game players. We asked participants to provide their online nicknames so that we could access their game history and rank through a publicly-accessible website that interfaces directly to the Riot Games API database (<https://euw.op.gg/>). Each player’s video game rank (computed from their position in an ELO-like ranking system similar to that used by the United States Chess Federation) was extracted from an online database (Elo, 2008; ‘LoL Stats, Record Replay, Database, Guide, MMR - OP.GG,” n.d.) . In LoL players are divided into ranked ‘tiers’ with each tier having five ‘divisions’. A player’s position in these divisions and ranks depends solely on the ratio of matches won and lost over time and not the performance within each match. Our participants’

rank ranged from ‘Silver Division 5’ up to the ‘Masters Division’ (“Matchmaking Guide,” n.d.).

After 10 placement games LoL players are assigned a Tier and a Division based on their win/loss performance. They subsequently move up and down on that Division and move between Tiers based on their win ratio. Tiers and Divisions therefore correspond to MMR ranges (“/dev,” n.d.; “Matchmaking Guide,” n.d.). In addition, there may be non-linearities in MMR within and between Tiers and data in our relatively small sample do not pass standard tests for normality. For these reasons, we show parametric statistics here only for visualisation purposes and all correlations are computed using Spearman’s rho. We note that the results from parametric and non-parametric analyses are almost identical. Moreover, the dataset is anonymised and freely available for download on the OSF website:

<https://osf.io/dsbx4/> .

3.3.1.4. Instruments

In this analysis we obtained psychometric test scores from subjects under laboratory conditions. We then compared those score with performance as measured by the subjects’ League of Legends rankings. We used the WASI-II Matrix Subtest (which is similar to Raven’s Matrices) as a standardized measure of fluid intelligence, along with three complex span working memory tasks (Symmetry, Rotation and the Operation Span task) that have been validated extensively (Maccow, 2011). A visual depiction of the complex span tasks can be seen below :

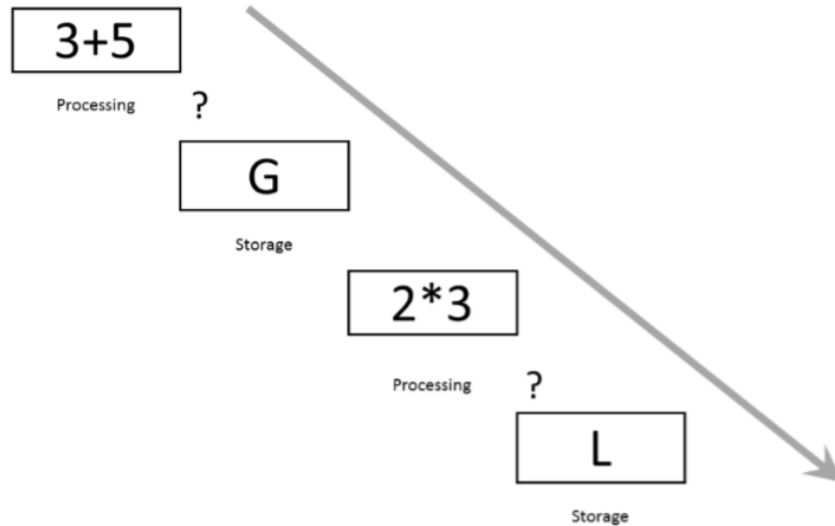


Figure 3-3. The Operation Span task (OSPAN). The participant needs to solve a simple arithmetic operation while they remember a sequence of letters. During the question mark the potential answer appears and they need to reply with whether than answer is True or False.

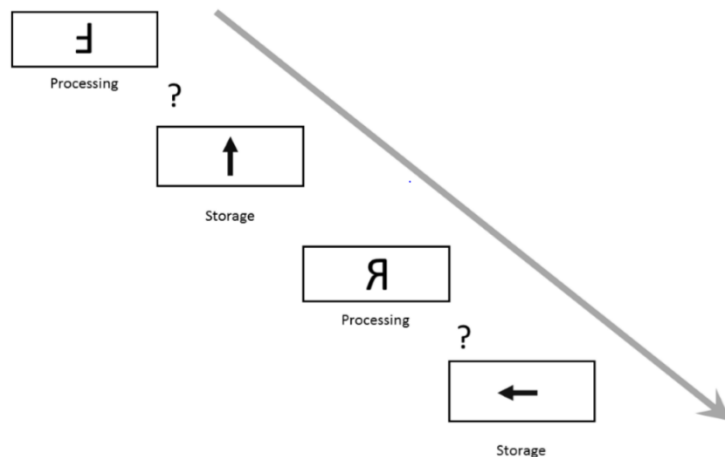


Figure 3-4. The Rotation Span Task (ROTSPAN). In this task the participant has to remember the sequence of arrows (storage) while they are asked to mentally rotate a letter to see if it is flipped or not (processing).

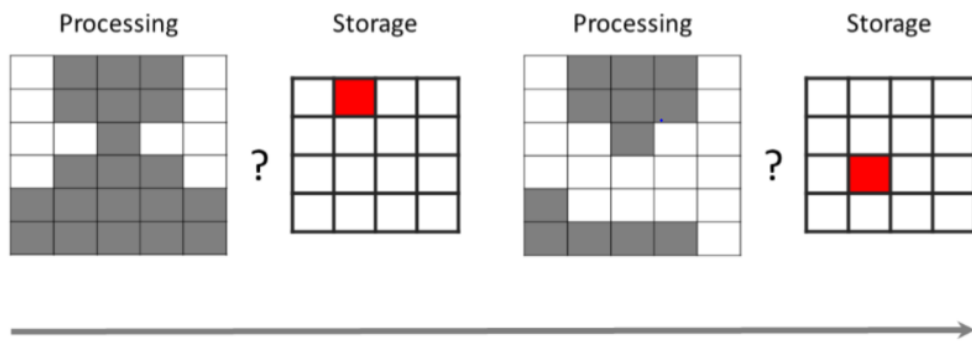


Figure 3-5. The Symmetry Span Task (SymSpan). The participant needs to determine whether something is symmetrical on the vertical axis (processing) while they remember where the order the order and the place the red block appeared (storage). Note that the first processing matrix is symmetrical however the second one is not.

3.3.2. Results

Since our data were both not normally distributed and because Rank in LoL represents an MMR range as mentioned above, we computed the non-parametric ‘Spearman rho’ correlation between Rank and performance on the standardized psychometric tasks. The correlations can be seen in Table 3-1. Aside from the storage component which is what is commonly used to measure one’s cognitive capacity I additionally have provided the speed components which can be seen in the Appendix D.

Table 3-1. The correlation matrix of the cognitive tests used.

Correlation Matrix

		League of Legends Ranking	Rotation Span Score	Symmetry Span Score	Operation Span Score	WASI II Matrices Raw Score	Folk Physics Score	Reading the Mind in the Eyes Score	
Spearman's rho	League of Legends Ranking	Correlation Coefficient	1.000	.260*	.117	.025	.440**	.091	-.096
		Sig. (1-tailed)	.	.031	.199	.427	.000	.253	.242
		N	56	52	54	56	56	56	55
	Rotation Span Score	Correlation Coefficient	.260*	1.000	.440**	.198	.303*	.178	.050
		Sig. (1-tailed)	.031	.	.001	.080	.014	.104	.365
		N	52	52	50	52	52	51	
	Symmetry Span Score	Correlation Coefficient	.117	.440**	1.000	.197	.247*	.108	.050
		Sig. (1-tailed)	.199	.001	.	.077	.036	.218	.361
		N	54	50	54	54	54	53	
	Operation Span Score	Correlation Coefficient	.025	.198	.197	1.000	.071	-.202	-.131
		Sig. (1-tailed)	.427	.080	.077	.	.302	.068	.171
		N	56	52	54	56	56	55	
	WASI II Matrices Raw Score	Correlation Coefficient	.440**	.303*	.247*	1.000	.214	.082	.082
		Sig. (1-tailed)	.000	.014	.036	.302	.	.056	.275
		N	56	52	54	56	56	55	
	Folk Physics Score	Correlation Coefficient	.091	.178	.108	-.202	1.000	.278*	
		Sig. (1-tailed)	.253	.104	.218	.068	.056	.	.020
		N	56	52	54	56	56	55	
	Reading the Mind in the Eyes Score	Correlation Coefficient	-.096	.050	.050	-.131	.082	.278*	1.000
		Sig. (1-tailed)	.242	.365	.361	.171	.275	.020	.
		N	55	51	53	55	55	55	

*. Correlation is significant at the 0.05 level (1-tailed).

**. Correlation is significant at the 0.01 level (1-tailed).

I found that fluid intelligence as measured by the WASI II Matrix Reasoning Subtest, correlated significantly with rank (nonparametric rank correlation: $r_s = .44$ (95% CI [.24 .60], $p = .001$). Importantly, we found no significant correlations between rank and scores in the MITE task or the Folk Physics Test and the partial correlation of WASI II scores with rank controlling for MITE was not significantly different to the initial correlation without accounting for MITE. Note that a weak correlation between MITE and Folk Physics was also identified. Similarly, we found only a weak correlation between rank and a test of visuospatial working memory. A scatterplot with the Working Memory tasks, the WASI-II Matrices, the LoL Rank, the MITE and the Folk Physics Test can be seen in Figure 3-6 below. Note that I made some data points see-through as well as adding some minor jitter better illustrate overlapping data points.

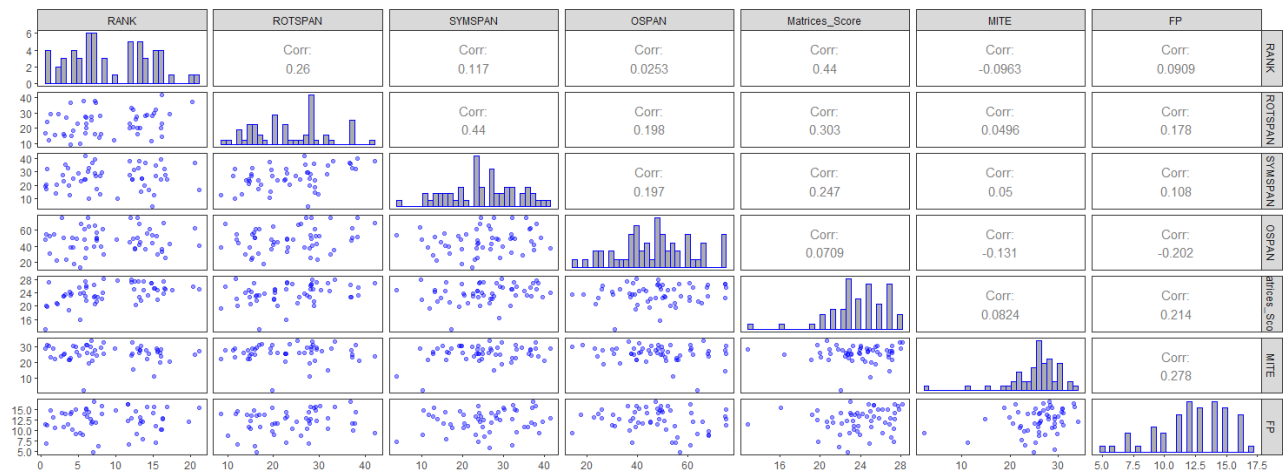


Figure 3- 6. Cross correlations between variables of interest. The leading diagonal shows a histogram with distribution of the data. Numbers above the diagonal show the non-parametric cross correlation coefficient. Scattergrams of the data are also presented. There is a moderately-sized and highly significant correlation between WASI-II Matrices and Rank ($r_s = .44, p = .001$) and a weak but significant correlation between Rank and Rotation Span score with $r_s = .26, p < .05$. The correlations between Rank and OSPAN and MITE task scores were not significantly correlated with with $r_s = 0.3, p = .43$ and $r_s = -.01, p = .242$ respectively. MITE was positively correlated with FP with $r_s = .28, p < .05$.

Outliers are always a concern in correlational analyses. To address this, we computed the Cook's Distance for all points in the MMR vs WASI II analysis. The highest Cook's distance was found for the subject with the lowest WASI II score. However, the Cook's Distance for this subject was .59 (well below the .70 threshold for our sample) and the correlation is virtually unchanged $r_s = .435, p < .001$ even if this player is excluded.

3.3.3. Study 2 : Aging and videogame performance across 4 different videogames

3.3.4. Methods

3.3.4.1. Ethics

We used existing data sources to ask whether performance in different types of games followed the age-dependent trajectory that would be expected if it was highly correlated with fluid intelligence. No player-identifying information was present in any of the datasets and data acquisition procedures were approved by a separate application to the University of York Psychology Ethics Committee. More details on player demographics are presented in Table 3-1.

3.3.4.2. Data sources

All four videogames use the ratio of historical wins to losses as a primary metric for the ‘Matchmaking Ranking’ (MMR) score which we analyse here. MMR is a dynamically-updated measure of player performance that depends solely on the win/loss history of each player and the rank of their opponents. It should be noted that Divisions and Tiers in League correspond to an MMR range that is hidden from the user but provided to us by Riot Games.

3.4.1.2.1. League of Legends

A snapshot of LoL player ranks was provided by Riot Games (Riot Games, a subsidiary of TenCent Holdings, Los Angeles, CA). Other aspects of this dataset have been analysed in a previous paper (Kokkinakis, Lin, Pavlas, & Wade, 2016).

3.4.1.2.2. DOTA II

A dataset from casual players who spectated at the ‘International 5 Dota 2 Tournament 2015’ was provided by the education analysts at Foundry 10 and Valve (Valve LLC, Bellevue, WA) (Bonny & Castaneda, 2017; Bonny, Castaneda, & Swanson, 2016; “Foundry 10 Homepage,” 2016).

3.4.1.2.3. Destiny

The anonymized Destiny dataset were obtained from the developer, Bungie (Bungie, Inc. Bellevue, WA), with age data from a public online survey of approximately 1700 Destiny players who participated on a voluntary basis.

3.4.1.2.4. Battlefield 3

Anonymized Battlefield 3 (Electronic Arts, Redwood City, CA) data were obtained from Tekofsky and colleagues and is available through their website (<http://www.psyopsresearch.com/download/>) (Tekofsky, Spronck, Goudbeek, Plaat, & van den Herik, 2015; Tekofsky, Spronck, Plaat, Van den Herik, & Broersen, 2013). In our analysis we used the data from the structure stats.global.elo.

3.3.4.3. Fluid Intelligence and Aging

Fluid intelligence changes with age (Horn & Cattell, 1966; Kaufman, 2001; Salthouse, 2010). Here, we asked whether performance in four different games had shared similarities in their aging profiles. Our hypothesis was that because MOBA performance correlates with fluid intelligence, it would follow an age trajectory similar to that seen in population-level raw IQ scores—peaking in the early to mid twenties with a decline thereafter (Craik & Bialystok, 2006; Horn & Cattell, 1966; Kaufman & Lichtenberger, 2005; Salthouse, 2009). We chose the first person shooters Destiny and BF 3 as controls since performance on these game might be expected to correlate more with reaction time and therefore peak earlier in the lifespan (Fozard, Vercrijssen, Reynolds, Hancock, & Quilter, 1994). MMR performance measures are available for each of these games but the absolute scaling of this ratio variable differs between games.

3.3.4.4. Grouping and Standardisation

To enable a direct comparison between the four games (Dota2,LoL,Destiny andBF3) I z-scored the MMR distributions for each game separately by removing the means and scaling by the standard deviations. We then separated these normalized within-game scores into three age groups designed to span the point at which raw IQ scores begin to decrease (13–21, 22–27 and 28–40 years old). The lower end of this range was a product of the registration requirements of the games we used. The high end was imposed to ensure that age bins were standardised across games: one of our datasets (DOTA II) had very few participants over this age cutoff. We note that Tekofsky and her colleagues (2015) used a similar strategy for almost identical reasons. Datasets for both studies are available on the Open Science Framework website <https://osf.io/dsbx4/>.

3.3.5. Results

3.3.5.1. Outlier Rejection and Descriptive Statistics

All distributions were inspected for possible outliers. We subsequently used Tukey's outlier technique ($k=2.2$) to identify other candidate points. The illustrations of the distributions before and after outlier rejection can be seen in Appendix E. Note, as mentioned previously, all these terms in the x-axis (ELO, Matchmaking Rating, Combat PVP) are interchangeable (different naming conventions depending on the company) and they use win-ratio as a primary determinant of a player's "skill-level" as mentioned previously. The descriptive statistics for each game post outlier rejection and after the subsequent age limits can be seen in Table 3-2 below.

Table 3-2. Player numbers and ages (in years) for the four games in our analyses after the outlier rejection and the age limits imposed.

Game	N	Minimum	Maximum	Mean	Std. Deviation
Battlefield 3	87443	13	40	23.67	6.53
Destiny	1669	13	40	24.18	6.32
DOTA 2	286	13	40	23.11	4.02
League of Legends	17861	13	40	20.49	5.07

3.3.5.2. Homogeneity of variance

Before performing a one-way ANOVA I performed tests of homogeneity of variance. The only dataset that passed was Battlefield. The Levene's tests can be seen in Table 3-3 below.

Table 3-3. The homogeneity of variance tests for each game.

Game	Levene Statistic	df1	df2	Significance
Battlefield 3	1.816	2	8740	<i>.163</i>
Destiny	4.983	2	1666	<i>.007</i>
DOTA 2	5.156	2	283	<i>.006</i>
League of Legends	15.442	2	17858	<i>.000</i>

3.3.5.3. One way ANOVA - Welch's F

Although ANOVAs are robust to small violations of this assumption, I am still reporting Welch's F due to the lack of homogeneity of variance. There was a significant effect of age on MMR for Destiny ($F(2,1076.55) = 40.21, p < .001$), for Battlefield 3 ($F(2,5185.93) = 122.87, p < .001$), for Dota 2 ($F(2,90.9) = 5.19, p < .05$) and LoL ($F(2,4114.37) = 57.44, p < .001$). Results are qualitatively identical for all games.

3.3.5.4. Contrast Coefficients

I performed two planned contrasts as functions of age band on our ANOVA data to test for effects of age. The first contrast [0 1 -1] asks whether there is a significant difference between the last ('elderly') age band and the mean of the second group. Effectively, it asks whether we see a general age-related fall-off in all games after the mid-20. We hypothesized that all games would show this effect because both IQ and reaction time measures begin to decrease after the mid-20s. The second contrast [-1 1 0] asks whether we also see a performance advantage in the mid-20s compared to the teens. We hypothesized that we would see this effect for MOBAs (which, appear to depend strongly on fluid intelligence) but not for FPS games (which may depend more on rapid reaction times and hand-eye coordination). These contrasts can be seen below, at the Table 3-4.

Table 3-4. Multiple planned contrasts for each game and age group.

Game	Contrast	Value of Contrast	Std. Error	t	df	Significance (2-tailed)
Battlefield 3	1	.3958	.4588	8.627	4609.502	.000
	2	.0186	.02470	.752	5318.884	.452
Destiny	1	.3991	.10456	3.817	990.696	.000
	2	-0.350	.05847	-.599	11134.143	.549
DOTA 2	1	.7614	.25968	2.932	113.215	.004
	2	.3743	.12720	2.943	198.484	.004
League of Legends	1	.3816	.03981	9.585	5760.801	.000
	2	.1723	0.1685	10.228	9669.915	.000

All games showed a significant difference between the middle group and the last group indicating that performance in general falls off after the mid 20s. Performance of the middle group when compared to the older group was significantly higher for Destiny, $t(990.7) = 3.82, p < .001, d = 0.45$, for BF3 $t(4609.5) = 8.63, p < .001, d = 0.4$, for LoL, $t(5760.8) = 9.59, p < .001, d = 0.17$ and for Dota 2, $t(113.22) = 2.93, p = .002, d = 0.38$. Only the MOBAs also showed a significant increase ($p < .001$) between the first and second age group. This increase is consistent with the hypothesis that performance in MOBAs (but not FPS games) is more (heavily) correlated with fluid intelligence due to the complexity of the genre which also exhibited in this age profile. Additional statistics and alternative figures can be found in Appendix C. A depiction of this relationship can be seen in Figure 3-7 below.

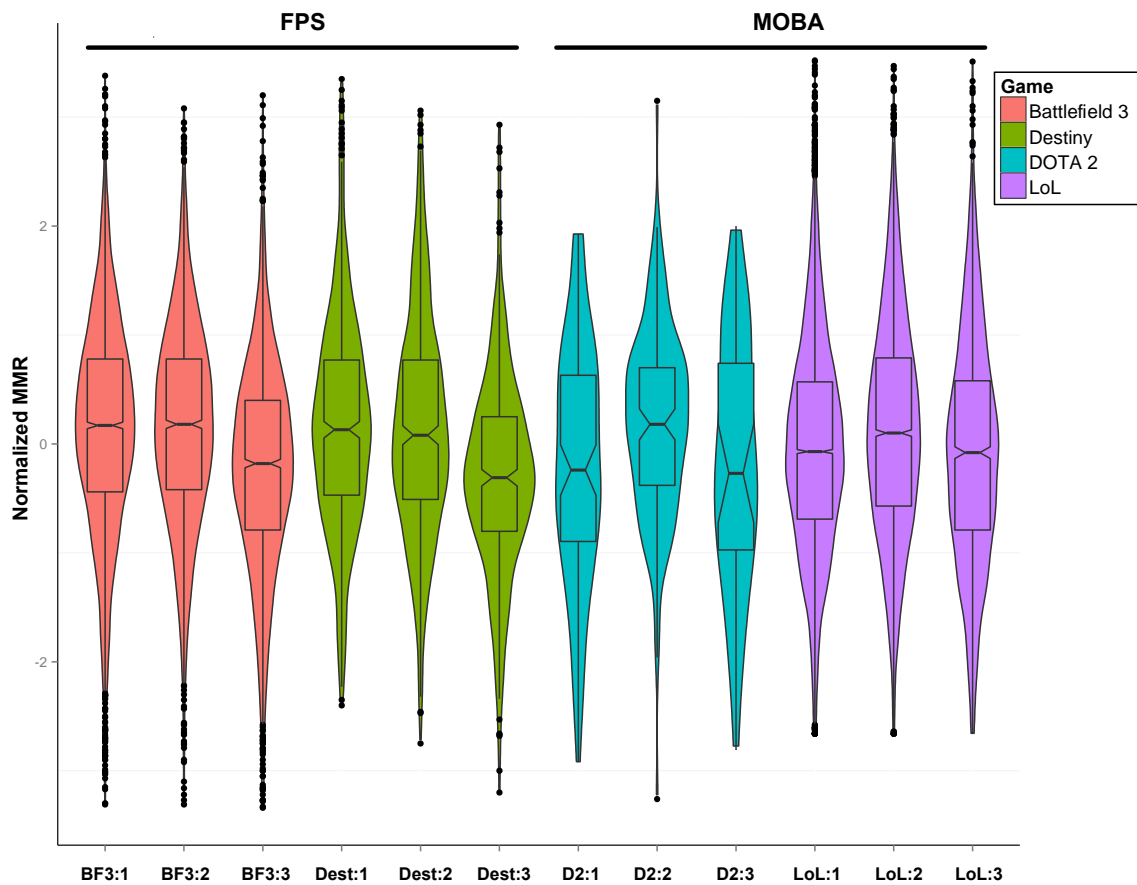


Figure 3-7. Age profiles of MMR in four different games. Three age groups for each game are plotted: (1) 13-21, (2) 22-27 and (3) 28 years and over. In two popular 'First Person Shooter' games (Battlefield 3 and Destiny), performance decreases monotonically with age following a 'high, high, low' profile. In comparison, two of the most popular multiplayer online battle arena (MOBA) games exhibit a 'low, high, low' profile suggesting that performance peaks in the mid-20s. Distributions whose boxplot notches do not overlap are different at $p < .05$.

Overall, we found that MOBA-genre performance profiles followed a 'low, high, low' pattern where performance peaked in the 22–27 year old age group. In comparison, FPS performance followed a 'high, high, low' pattern suggesting that younger players had a relative advantage in this genre and that performance decreases monotonically with age.

3.4. Discussion

The literature around video games, psychology and neurophysiology (much of which focuses on FPS games) is extremely diverse and populated (Palau, Marron, Viejo-Sobera, & Redolar-Ripoll, 2017). Green and Bavelier's seminal work in the early 2000s identified perceptual effects of FPS play and later studies extended this to attentional effects and cognitive tasks such as response inhibition, task switching and working memory (Cain, Landau, & Shimamura, 2012; Colzato, van den Wildenberg, Zmigrod, & Hommel, 2013; Colzato, van Leeuwen, van den Wildenberg, & Hommel, 2010). Furthermore, extended FPS play may lead to improvements in visuospatial processing (C. Shawn Green & Bavelier, 2003) (although see (Boot, Kramer, Simons, Fabiani, & Gratton, 2008b; Latham, Patston, & Tippett, 2013; N. Unsworth et al., 2015) for contesting these claims). Despite this vast literature this is the first study that examines fluid Intelligence and working memory and its relationship to rank against human agents. Moreover, this is the first study that provides an age profile for the specific MOBA subgenre juxtaposing it against the popular FPS genre. I will briefly reiterate the results and how they relate to previous literature. I will also consider different explanations such as practice and age while I finally explain the important implications of my findings in the fields of Health and Education.

3.4.1. Study 1

The correlation between rank and fluid intelligence ($r_s = .44, p < .001$) is slightly smaller than those reported by other groups that have used other videogames as instruments. For example, Foroughi and his colleagues (2016) report a correlation coefficient of .65 between IQ and performance in a custom-made game based on Portal 2 while Quiroga and her colleagues (2015) report a remarkable correlation level of .96 adjusted variance explained

(r^2) (equivalent to a correlation coefficient of around .98) in a study using games from the ‘Big Brain Academy’. These results further support that it is possible to measure fluid intelligence using computer based tests as a proxy, and we have no doubt that some single player puzzle games like Portal test both abstract problem solving and visuospatial working memory, for instance through the number of rules or problems that one needs to maintain to reach the solution. In both cases however, the games used were specifically selected (or designed) a priori to probe IQ. More specifically, in the case of the Quiroga battery, the Big Brain Academy ‘brain training’ games resemble ‘gameified’ intelligence tests and the degree of correlation with actual tests is, perhaps, less surprising (see also Banquied et al., 2013).

One of the key advantages of this study is that the emergent correlations between Rank and fluid Intelligence are remarkable because they are based on data from an unmodified, widely played commercial video game that has no prior links to intelligence testing. Thus these data may also be noisier than the single-player data obtained by the groups discussed above because of the multiplayer nature of the MOBA genre—the outcome of each game depends on a team effort, which stabilises after a large number of games and the proficiency of the opposing team contributes to additional variance.

Even though the visuospatial tests yielded interesting results, I failed to reject the null hypothesis for the MITE and LoL Rank; I found no correlation between performance in LoL and the MITE test which taps theory-of-mind (S. Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001). This was unexpected: MOBAs are social games and I believed that the ability to model the motives of other players enhances performance. In addition, as mentioned in the Introduction of this chapter, scores on the MITE task have been shown to correlate with IQ and Engel and his colleagues (2014) showed that MITE scores predict performance in cognitively demanding tasks such as solving Sudoku puzzles. Moreover, MITE was implicated in a study with a much larger sample that examined how teams perform in LoL

with higher MITE scores being linked to better LoL performances (Kim et al., 2016). I conclude that, at least in our relatively small sample population, any weak correlation between MITE and intelligence scores that may exist may have been overridden by other factors. In addition, if the relationship between performance and MITE score is driven by subjects with poor TOM, we may not have sampled an MITE dataset with sufficient variance over the low end of the range to expose the effect. Finally, it should be mentioned that even though we treat Theory of Mind of mind as a single construct that may not be the case. It is possibly subdivided into an automatic system that processes information fast and efficiently and a more elaborate system and slower system that performs higher order operations; thus this more tests need to be performed, perhaps including perspective taking ones, in order to clarify the exact link between performance and ToM or Tom subsystems (D. Schneider, Slaughter, & Dux, 2017).

I also failed to reject the null hypothesis for the Folk Physics test for both Rank and MITE. This was highly surprising since the FP test has been negatively correlated with MITE albeit it was in a non-typically developing population (Simon Baron-Cohen, 1997). More surprisingly, it did not correlate with any of the cognitive performance variables which is highly surprising. As mentioned in the Introduction visuospatial Intelligence has been previously characterised as “mechanical” intelligence by Vernon as well as others (Johnson & Bouchard, 2005). This failure to reject the null hypothesis can possibly be attributed to our small sample size; a replication of this with a bigger sample may be useful.

Finally, I failed to reject the null hypothesis for the Operation Span task (OSPAN). This is not surprising due to our sample as well as due to the fact that OSPAN does not tap the visuospatial abilities that are required in LoL or in the aRTS genre. However, I believe that due to all cognitive performance tests intercorrelating (positive manifold) other researchers should observe a small correlation with a much bigger sample (Jensen, 1998).

It should also be noted that the reviewers of this paper prior to its publication also asked to see the relationship of the processing variables (reaction times) aside from the storage part which is more commonly used (Foster et al., 2014a; Nash Unsworth & Engle, 2006). I did not find a correlation between the processing aspects of the complex span tasks and MMR in Study 1. This might indicate that MMR depends most critically on the working memory aspects of these tasks. However, I also note that there was little variance in the processing scores—all subjects tended to score well in these tasks. It is possible therefore that a correlation between processing and MMR might emerge in a larger, more diverse sample population.

Previous research supports the idea that fluid Intelligence can be extracted from videogames scores. Rabbitt and his colleagues (1989) showed that intelligence correlated with practised performance on “Space Fortress”: a simple, arcade-like, non-commercial video game designed by psychologists. This link is also supported by recent work showing that the rate of early learning in an online video game and the stable final performance are correlated, lending support to the idea that a single factor (presumably related to cognitive capacity) underlies both metrics (Stafford and Dewar, 2014). Importantly, Stafford and Dewar’s work (2014) suggests that cognitive capacity is assayed by a final, stable performance metric which is ultimately invariant to increasing practice. This correlation between performance after large amounts of practise and cognitive capacity is further supported by Adams and Mayer (2012) who identified a correlation between scores in a first-person shooter (Unreal Tournament) and two mental rotation tasks in non-videogame players (Shepard-Metzler & paper-folding). Finally, similar results have been observed by Bonny & Castaneda (2017) who observed that number processing ability not only correlates with Dota 2 MMR but is also predictive of MMR improvement over time.

3.4.2. Practice as an alternative explanation

The first point the cautious reader should raise, is that even though our data indicate a link between intelligence and video game performance, the relationship is correlational. Moreover, playing video games is known to positively impact cognition, for instance by improving attentional allocation or modifying visual perception abilities (Green & Bavelier, 2003). Thus an alternative explanation is that the correlation might arise because playing video games causes an increase in intelligence. Although IQ scores are believed to be relatively stable, training on action video games does improve visuospatial performance and the general mechanism (an improvement in probabilistic inference based on visual input) could, potentially, translate to a wide range of cognitive tasks (Deary, Penke, & Johnson, 2010; C. S. Green & Bavelier, 2008; C Shawn Green, Pouget, & Bavelier, 2010). Addressing this possibility robustly would require a set of large-scale longitudinal experiments and is beyond the scope of the current study. However, we did look for practice effects, which will be further elaborated in Study 2 below and practice does not appear to be an issue. It should also be noted that the effects videogames have on cognition have been disputed by Unsworth and his colleagues (2015) using a much larger sample (N. Unsworth et al., 2015). Finally, employees from Valve and Riot as well as separate scholars have confirmed that past a certain number of games players reach some form of Rank stability finding it hard, if not impossible, to progress past that Rank (Aung et al., 2018; “Matchmaking | Dota 2,” n.d.). Thus even though practice is extremely important in this game (at least initially), it seems to be more of a critical factor in simpler or more repetitive games that allow for multiple trials; that is not to say that practice may not yield effects up to a specific Rank or MMR bracket but it could be viewed as a variable with diminishing returns.

3.4.3. Study 2

In our second study I juxtaposed 2 distinct videogame genres showing that the performance in them follow a distinct developmental pattern. Performance in all videogames seems to drop off in the final age group. This agrees with previous literature that shows a decrease in multiple skills related to fluid intelligence and visuospatial abilities past one's late twenties (Kaufman & Lichtenberger, 2005). MOBAs also seem to have a peak in approximately one's mid to late twenties as opposed to FPSs that seem to have a peak in one's late teens or early twenties. In a sample of 3,305 Starcraft II players, aged between 16-44, Thompson and his colleagues (2014) estimated that one reaches their peak performance at approximately one's mid-twenties with a steady decline in performance after that period. This is further echoed in certain chess studies where ELO, the equivalent MMR of chess, drops in a similar fashion as one ages (Vaci, Gula, & Bilalić, 2014).

3.4.4. Age and Practice as confounding variables

An important possibility that needs to be addressed is that some third factor is driving variance in both intelligence and video game expertise. One candidate is age. Raw (un-normalized) fluid intelligence scores usually peak in the mid-20s (Craik & Bialystok, 2006). This also appears to be the approximate peak of video game performance in MOBAs that depend on a mixture of memory, tactics, strategy and reaction time while games that emphasise more reaction times and hand-eye coordination (for example, FPS-type games) appear to advantage younger players. However, while we did find a significant correlation between WASI II scores and age in Study 1 ($r = .28, p = .035$) our LoL data showed no correlation between expertise and age and a partial correlation of expertise with WASI II accounting for age was still highly significant ($p = .001, r = .45$). Age may also correlate with

practise: older players may have had more time to practise any particular game (although it is also possible that older players are more restricted in the amount of free time that they can devote to game play). To examine whether pure practise effects determine rank, we examined the relationship between rank and games played in our large ($N > 17000$) dataset of LoL players. After the initial learning stage during which the players attained a relatively stable rank, the magnitude of the correlation between games played and expertise (indicated by MMR) was $r = .02$. While still significant ($p < .001$) this suggests that games played explains only a small amount of the variance found between experienced players (Ackerman, 2014b; Bonny & Castaneda, 2017). This agrees with Stafford and Dewar's (2014) finding that final, stable performance levels are determined largely by the rate of learning in the initial phase rather than the total number of games played overall. Having acknowledged this stability, one should not diminish the role of targeted deliberate practice in achieving or at least maintaining one's performance.

3.5. Limitations

The two studies presented here have very different sample sizes. The first study had an N of 56, while the second study had many thousands of data points from different datasets. The two analyses therefore have different strengths and weaknesses. In Study 1, although N is small, the data are collected under controlled laboratory conditions using standardized instruments. This also partially justifies the small number of test that were used to assess cognitive performance, since both time restraints in combination with fatigue are real variables that need to be taken into consideration. It should also be noted that this relatively small sample size cautions us that this work is still exploratory, the large effect size and strong significance are encouraging. The data from Study 2 are collected from larger cohorts but the provenance of each data point is less certain. Issues such as selection bias are potentially problematic and we expect to find some noise in measures such as age due to

participants deliberately or carelessly reporting false information—although there is evidence that large web-based samples such as these can be relatively reliable (Gosling, Vazire, Srivastava, & John, 2004). The hope with large datasets such as these is that the huge number of participants more than compensates for the increase in noise caused by the sampling methods. An important challenge for future research is to test this assumption by broadening the use of validated tests to a wider subject group—perhaps through careful use of online crowdsourcing platforms (Crump, McDonnell, & Gureckis, 2013).

I also recognize that although fluid intelligence seems to be one factor in obtaining a higher MMR, it does not explain all the variance—practice, dedication and learning must still confer significant advantages—particularly during the early stages of skill acquisition. This effect is mitigated to some extent in Study 1 by the fact that all our subjects were relatively well-practised (over 100 games excluding casual and unranked matches) and had demonstrated a willingness to engage with the game intellectually over a long period of time.

An important methodological point that some researchers seem to not take into account is that each individual genre as well as game within that genre is highly distinct with its own features and quirks. Unfortunately, as Latham and his colleagues (2013) mention, there is a tendency to lump many videogames of different genres or different difficulty even together, not respecting their individual game features. Even though I did cross-compare them above, I did not delve into highly specific comparisons (for instance older group in Destiny vs older group Dota 2) since as mentioned above such comparisons would not be fruitful due to the difference in genre and the possible difference in the populations that play them. Rather I only focused on overarching general themes (a decline in performance past a certain age as well as a peak in one's mid-twenties). Moreover, I did so due to the fact that I obtained a large sample that partially negates many individual differences.

To further clarify the impact of game features on performance one could look at Starcraft II. Thompson and his colleagues (2014) mention that older individuals in Starcraft II may use different compensatory mechanisms that allow them to maintain their Rank or at least maintain a less steep decline. These older individuals may have the same cognitive attributes as another individual that plays a different strategy game, even of the same genre, that has no similar game features that allow for the same compensatory mechanisms to occur; an example would be that of an older individual that has slower reaction times who chooses a really strong unit or hero that has both a high winrate as well as high armour and therefore is really strong. This older individual could maintain his or her rank only on this unit or hero that is available only in this specific game title. A similar strategy game that is theoretically perfectly balanced with heroes that are much less forgiving would not allow for such an individual to maintain their Ranking. Thus game features should be taken into account in future studies, at least if researcher want to perform more elaborate statistical comparisons between different games.

Following the idea of game features being important I need to highlight the idea of meta-strategy or metagaming. The dictionary defines this “an overarching strategy determining which other strategies to use in a given situation” (“Metastrategy dictionary definition | metastrategy defined,” n.d.). Metagaming is a highly similar and more encompassing term but more focused on games (Cunningham,” 2007). As Cunningham (2007) states it has to do with the viability of the strategies other people will use in a game or tournament; note that this could include actions outside of the game for instance losing a match on purpose to be matched with an easier chain of opponents.

Many games such as chess are not perfectly balanced in terms of winrate with white having a slight (approximately 5%) advantage over black due to the first move advantage (Moul & Nye, 2009). MOBAs are also imbalanced with different heroes having different

winrates (“Meta - OpenDota - Dota 2 Statistics,” n.d.). Thus since certain heroes or tactics are better than others or more viable (more likely to win), it comes to no surprise that they keep getting used over and over again. This leads to the designers creating game updates known as patches where they try to rebalance the game by bring each hero closer to a 50% winrate which represents balance (similar to a cointoss resulting in tails 50 % of the time) (“Patch 4.4 Notes | League of Legends,” n.d.). This of course could be problematic if one has a limited sample or attempts to focus on a single individual based on their MMR since a valid strategy this individual has used may be rebalanced by the game developers. As an example if an individual achieves a really high Rank with hero that has 50 % winrate but later that hero’s winrate is diminished through game changes, whether those are direct or indirect, that individuals true performance will not be reflected. In conclusion, one should be wary of extrapolating cognitive ability strictly through game performance, at least at the individual level, unless they have accounted for rank stability, strategies used to acquire said rank, competitiveness among others.

3.6. Implications/Contribution

I proposed that videogame expertise in commercial MOBAs correlates with fluid intelligence or visuospatial abilities and the developmental trajectory of expertise mimics that of fluid intelligence across adolescence and early adulthood (Craik & Bialystok, 2006; Kaufman, 2001; Kaufman & Lichtenberger, 2005). This could give a number of important cross-population demographics such as reaction times, working memory as well as fluid intelligence which we could further link to geographic locations.

The first implication is that we could use videogames in a non-threatening and non-costly manner to identify children’s visuospatial abilities. Visuospatial abilities have been shown to both facilitate the learning of new concepts as well as help in increasing

performance in a number of academic subjects such as mathematics, chemistry or programming ; for instance through visualisation of chemical connections in a molecule or through helping interpreting complex graphs (Lohman, 1996; Mix & Cheng, 2012; Stieff, 2013; Wu & Shah, n.d.). This early identification of children that are lagging behind in the visuospatial domain would also lead to an earlier more effective intervention that would boost children's scores and thus improve their academic performance later in life (Lowrie, Logan, & Ramful, 2017). Moreover, we could identify potentially "gifted" children assigning them to better classes at an earlier time, effectively making them take full advantage of their cognitive ability helping those individuals attain an overall better life satisfaction (Gross, 2006) .

Aside from cross-population inferences and educational inferences, if we establish the developmental trajectories of a healthy population (as indicated by videogame scores), we could in theory detect deviations from normality/outliers and identify at risk populations since the onset of a large number of diseases is associated with declines in attention, motor performance as well as fluid Intelligence; all overlapping constructs which are tapped to various degrees by videogames (Mesholam-Gately, Giuliano, Goff, Faraone, & Seidman, 2009; Nuechterlein et al., 2004; Thomas, Goudemand, & Rousseaux, 1999).

Extending this idea we could apply this theory to a typically developing population as well. The field of Cognitive Epidemiology supports the notion that psychometric Intelligence can predict mortality even when measured at an early age, for instance when the participants were 11 years of old, as evidenced by the now classic experiment of the Scottish Mental Surveys, (Deary, 2010; Deary, Whiteman, Starr, Whalley, & Fox, 2004). Deary (2010) claims that a major strength of Cognitive Epidemiology so far has been its large samples, with their numbers ranging from a few thousands to one million. With videogames gaining immense popularity it is not unreasonable to expect samples of tens of millions so the use of

videogame datasets is imperative. Moreover, the accessibility of videogames in developing areas with games that do not require high graphics will help eliminate any economic barriers; gaining access to less privileged samples, where it is highly likely that fluid Intelligence may have more protective effects amplifying longevity. Having already established a link between already existing and validated psychometric tests that measure working memory and fluid intelligence, future research should focus on employing a bigger sample with a wider variety of tests. This wider variety of tests would help us refine the relationship between cognitive performance and videogame scores, allowing us to probe different strata of visuospatial abilities.

As a real-life example, Professor Wade (personal communication) has expressed interest in looking at the average MMR of the town of Flint in Michigan. This town has had a water crisis with lead poisoning affecting part of its population. Professor Wade believes that by doing spatial correlational analyses we can see meaningful decreases in MMR when compared to another control town, since lead poisoning has been linked with increased reaction times in exposed workers which might influence the town's players' MMR (Stollery, 1996).

In conclusion an important problem with the current field of video-game research is the fixation on using videogames to “enhance” cognitive abilities or delay cognitive decline (Baniqued et al., 2014). Despite cognitive training games being popular and therefore lucrative products, evidence of their effectiveness is mixed (Baniqued et al., 2014). A recent study even reported that a commercial videogame, Portal 2, that was not designed for cognitive enhancement might be a better substitute than those that explicitly advertised for ‘brain training’ (Shute, Ventura, & Ke, 2015). Unfortunately, these claims for cognitive enhancement have obscured an alternative measure that videogames may be ideally suited to: evaluating visuospatial working memory, fluid Intelligence, reaction speed and oculomotor

control (M. B. Jones, 1984; J. L. McPherson, 2008; Rabbitt et al., 1989). In this chapter I have proposed that the expertise in a massively played videogame can predict visuospatial working memory as well as intelligence, as measured by the WASI-II, among other psychometric constructs. Thus video games may offer an insight into a wider range of cognitive abilities and function helping researchers instantly overcome existing issues with small sample sizes, and potentially allowing us to examine dynamic changes in performance at a population level in almost real time. This could lead to a wealth of new information of higher quality that may revitalise the field of Psychometrics.

4. Emotionality in Videogames

4.1. Chapter Introduction

In the previous chapter I discussed how fluid or visuospatial intelligence and working memory correlated positively with videogame rank. Moreover, I presented data from different age groups to show that videogame performance in aRTS or MOBAs mimics the trajectory of fluid intelligence across one's lifespan. In this chapter I will focus on a specific personality trait and how it influences in game performance (videogame rank). The reason I focus on videogame rank has been previously explained in Chapter 2. However to reiterate briefly : videogame rank is robust and stable assuming one individual has played a large number of games, especially because it is the distillation of each match's outcome.

The Psychology of individual differences is an extremely wide topic with thousands of papers. Thus creating a fully descriptive model of each individual core trait and how it relates to videogame performance is near impossible, since it would require both in-game data that is not available to company outsiders as well as a full understanding of each individual personality trait, aside from having to identify how it is expressed in videogames. Thus, I focused on one of the most stable and robust personality traits that is present in multiple personality models albeit with minor differences in its subtraits, that of Neuroticism or Emotionality. I have performed two consecutive experiments. The first one was done online with many thousands of participants and it served as an exploratory basis for my later experiments. In that experiment I focused on a turn based card game rather than an aRTS. In my second experiment is focused on a MOBA (LoL) and it serves as a replication of the previous finding in a laboratory environment with a more visually demanding videogame. This is the same sample that was used for Chapter 3.

4.2. Introduction

Emotionality, Neuroticism or negative affect is a stable personality trait that is prevalent in multiple theories of personality (Mincic, 2015). Although there are slight differences in how these factors are defined, a stressful or a nervous predisposition is almost always the core in these models (K. Lee, Ogunfowora, & Ashton, 2005). This trait can be problematic in certain situations inhibiting performance in multiple fields, ranging from sports to academic subjects (Moutafi, Furnham, & Tsaousis, 2006; Smith, Smoll, Cumming, & Grossbard, 2006).

The HEXACO framework is an alternative to the Big Five with similar, yet not identical, constructs. It is an acronym for its traits which are Humility-Honesty, Emotionality, eXtraversion, Agreeableness, Conscientiousness and Openness to experience (K. Lee & Ashton, 2016). Emotionality can be further broken down into four subtraits : Fearfulness, Anxiety, Dependence and Sentimentality. The first main difference between HEXACO's Emotionality compared to the Big Five's Neuroticism is that Emotionality is less negative, pejorative term (Ashton & Lee, 2007) . The second is that Neuroticism does not include the aforementioned Fearfulness subtrait which measures physical harm avoidance, but rather it includes elements that are weakly associated with it (Ashton, Lee, Visser, & Pozzebon, 2008). This might explain why HEXACO's Emotionality has overall higher correlations with the Phobic Stimuli Response Scales than the Big Five's Neuroticism (Ashton et al., 2008). It should be noted that despite this distinction, the two traits are closely related, with Neuroticism being moderately correlated with Emotionality ($r=.41$, and $r=.52$ $p<.001$) (Ashton & Lee, 2009; Bashiri, Barahmand, Akabri, Hossein Ghamari, & Vusugi, 2011). Due to their overlap and the general underlying commonality of the Anxiety factor I will be using Neuroticism and Emotionality interchangeably (Mincic, 2015).

Neuroticism is extremely interesting from a non-academic or psychometric point of view due to its importance in influencing our choices. Anger and fearfulness, both associated with Neuroticism, have been shown to influence risk taking perspective in multiple theoretical problems, making people risk optimistic and risk averse respectively (Lerner & Keltner, 2001). Moreover, Neuroticism or its related subtraits, has been to shown to influence performance in a number of studies ranging from economic games to simulated gambling tasks (de Visser et al., 2010; Hartley & Phelps, 2012; Miu, Heilman, & Houser, 2008). This is echoed in the game of poker where if a player loses a significant amount of money, they may experience negative emotions that will further impede their decision making; this phenomenon is known as tilting (Palomäki, Laakasuo, & Salmela, 2013).

In a further study by Laakasuo and his colleagues (2014), used the 60 item version of the scales (HEXACO-60) and found that Emotionality, correlates negatively with poker proficiency and the monetary “stakes” individuals played at (Ashton & Lee, 2009). Laakasuo and his colleagues (2015) followed up on that research by priming poker players with a story that elicited anger. Player who were primed that way were indeed more prone to commit numerical miscalculations and thus commit misplays in a poker setting. Thus in-game performance is influenced to a significant degree by one’s emotions which are a direct response to one’s performance. The fact that even videogames incorporate some amount of luck either in their features similar to poker, in combination with them eliciting strong reactions to verbal or in-game graphic stimuli, such as blood, makes this an interesting topic of research.

I hypothesized that Emotionality can be a factor in videogame performance since all of the aforementioned stimuli and conditions are present in a videogame setting making the aforementioned hypotheses testable. Videogames include fast attentionally commanding or demanding projectiles, highly threatening stimuli (aggressive faces, blood) and a

progressively demanding environment through matchmaking where your opponent's skill progressively becomes higher until it reaches your own ("Matchmaking Guide," n.d.).

Our first study was modeled after the Laakasuo et al., (2014) poker study and served as a replication of his experiment a videogame setting that is turn-based. I asked whether the 'Max Rank' of a player in Hearthstone correlates with Emotionality using the HEXACO-60. It also served as an exploratory experiment in order to establish the strength of the relationship in a less threatening environment. I subsequently replicate this using a League of Legends sample and the HEXACO-100. My reasoning behind the second replication of this study was that by using the HEXACO-100 we could deconstruct which specific sub-trait might have been responsible with Rank achieved since the authors of the HEXACO said that the version with the 100 items is better and more stable when one wants to measure subtraits (Ashton & Lee, 2009). Moreover, LoL is more visually demanding than Hearthstone which might give us a stronger effect.

4.3.Study 1

4.3.1. Methods and Materials

4.3.1.1. Ethics

The experiment was approved by the Departmental Ethics Committee prior to any data collection. The participants were informed that they are free to withdraw from the study without any consequences at any time and they were debriefed at the end of the online form. Furthermore, one of the experimenters answered their questions in the forum during the on-going data collection.

4.3.3.1.2. HEXACO-60

The HEXACO-60 is a shortened variant of the HEXACO-PI-R (Ashton & Lee, 2009). It contains 10 items for each scale (Honesty-Humility, Emotionality, eXtraversion, Agreeableness, Conscientiousness and Openness to Experience). It was derived by selecting the questions with the highest loadings after running a principal component analysis (varimax rotation) on the HEXACO-PI-R (Ashton & Lee, 2009). The reason we used the HEXACO-60 rather than its 100-questions counterpart was to reduce drop-out rates, fatigue effects as well as increase the participants' response rate. Question number 25 was omitted due to technical difficulties.

4.3.3.1.3. Gaming Questionnaire

A short gaming questionnaire was also administered with questions asking about their attitudes towards the game's features (for instance art, ranks and gameplay) among others. More importantly I asked them about their in-game performance at different points in time as well as their Maximum Rank ('Max Rank') Achieved. The reason we examined Max Rank Achieved in Hearthstone as opposed to current Rank in League is that Rank rewards in League reset yearly. However, once a player reaches their maximal Rank in Hearthstone and earns their reward, for instance the "legend cardback" or the knowledge of what their peak performance is, they might not be motivated enough to continue (it should be noted that this has changed currently with Blizzard giving in-game rewards monthly based on Rank) ("Ranked," n.d.). Moreover, in the chess literature peak rating is a more reliable variable due to the fact that individuals may stop trying after they attain their maximum rank or because they get older and life events may influence their practice regiments (Howard, 2005, 2009, 2014).

4.3.3.1.4. Participants

In our experiment, the majority of participants were recruited from an online Hearthstone forum: www.reddit.com/r/hearthstone. In order to protect their anonymity, no individually identifying questions were asked such as a Hearthstone username, user ID or exact age; participants were asked for an approximation of their age instead. The original sample had N=3,123 participants and it was disproportionately male with 3,012 individuals identifying as male. Cases that had joke answers, for instance the words “420” in all answer boxes (a codeword for marijuana usage) or that were highly unlikely to be true (being 12 year old and having achieved the top rank in Hearthstone for multiple seasons) were deleted. Mahalanobi’s distances were additionally used to remove outliers leaving me with a sample of N=2,973 participants.

4.3.2. Results

4.3.2.1. Missing Value Analysis

I firstly performed a Missing Value Analysis to see if my data are skewed in a certain way. It is possible that if individuals deem a question too embarrassing they might ignore it, with only a specific members of the population answering it e.g. individuals may not answer a question about how extraverted they are, with only extraverted people responding thus misrepresenting the population mean. I firstly inspected the percentage of missing values which can be seen in Figure 4-1.

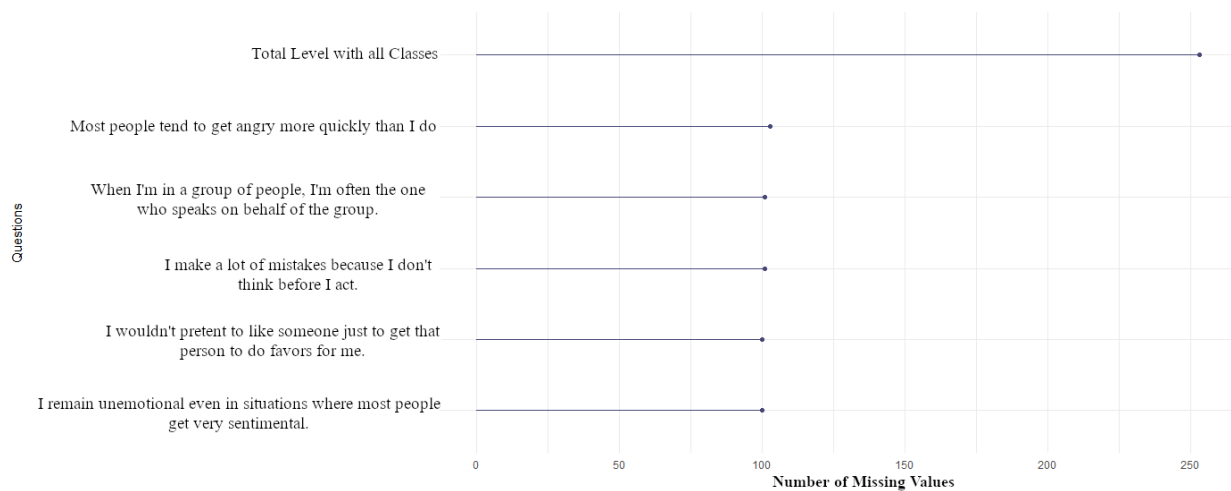


Figure 4-1. The top 6 questions with missing values. Note that only the “Total Level with all classes” is at 8.1 % with the rest of the questions missing approximately 3.2 % of their values.

Even if the numbers of missing values are negligible some researchers recommend examining to see if they are randomly occurring (J. W. Graham, 2009).

There are three possibilities : The first is that the data are Missing Completely at Random (MCAR) which is the optimal possibility, the second is that the data are Missing at Random (MAR) which is that the data are missing at Random if one account’s for a third variable or event and finally that the data are Not Missing at Random (NMAR) which means that the sample is biased. I will briefly clarify the distinction between MCAR and MAR since the terms are relatively confusing and it has been a subject of some articles (J. W. Graham, 2009; Zhang, 2015; Zygmunt & R. Smith, 2014). If a subset of the data has been compromised with missing data points and they are not missing at random but the missingness can be accounted for and explained through another variable, for instance a power outage, then the data are MAR; we know that the power outage has not influenced the normal population mean and the data are ignorable. Little's MCAR test obtained for this

study's data resulted in a chi-square= 27435.64 (df = 26201, $p < .001$). This means that the data are not MCAR and are either MAR and can be accounted for or NMAR and there is an inherent bias in the questionnaire.

A highly likely explanation for this inconsistency in the data is people checking in and then dropping out due to fatigue or disinterest. This could create clusters of missing values in the later parts of the questionnaire thus explaining the above results. Thus I performed a correlation between the position of the question in the questionnaire and the total missing values. The number of missing values were strongly correlated with their position in the questionnaire with $r_s = .87, p < .001$. Note that even though this highly explains our missing data, since the allocation of questions in the questionnaire was random i.e. not all eXtraversion questions were towards the end biasing the result, I performed some additional analyses to find the number of missing values in our sample as well as any underlying patterns. The top 6 variables with missing values can be seen in Figure 4-2 below.

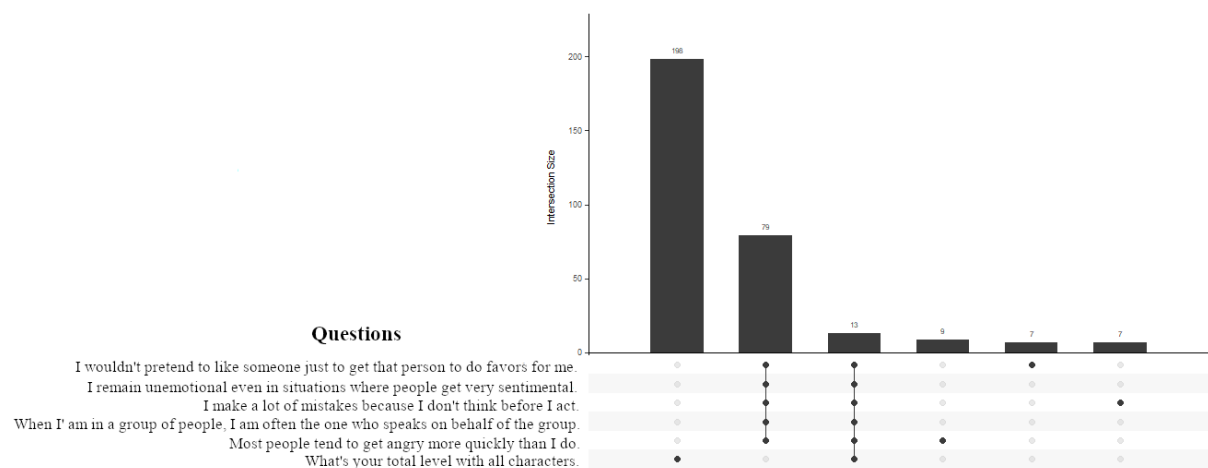


Figure 4-2. Missingness across the various questions. Do note that all these questions seem to not belong to a single facet/trait but rather they seem to be spread across multiple distinct ones.

The questions above, with the exceptions of the Hearthstone question, could be relatively sensitive for some individuals. Even so they do not exceed the 5% cutmark some

researchers have set as a soft limit, and therefore their influence on our large sample seems negligible (J. W. Graham, 2009). Thus we can use listwise deletion to remove them without being concerned with loss of power due to our big sample.

4.3.2.2. Descriptive Statistics and Distributions

I firstly present the descriptive statistics of our sample juxtaposed against those of a male college aged sample obtained by the HEXACO's creators in parentheses (Ashton & Lee, 2009).

Table 4-1. Our HEXACO-60 descriptives with the alternative means of a college aged male sample in parentheses. Note that they are near identical with the exception of eXtraversion.

	N	Means of two Samples	Std. Deviation
Honesty-Humility	2973	3.17 (3.04)	0.76 (0.71)
Emotionality	2973	2.76 (2.93)	0.58 (0.61)
eXtraversion	2973	2.88 (3.47)	0.79 (0.63)
Agreeableness	2973	3.23 (3.19)	0.7 (0.65)
Conscientiousness	2973	3.21 (3.31)	0.67 (0.62)
Openness to Experience	2973	3.50 (3.73)	0.69 (0.68)
Maximum Rank Achieved	2973	7.12	5.13

I subsequently inspected the two distributions of the variables in question, Emotionality and Maximum Rank Achieved. The two distributions can be seen in Figures 4-3 and 4-4.

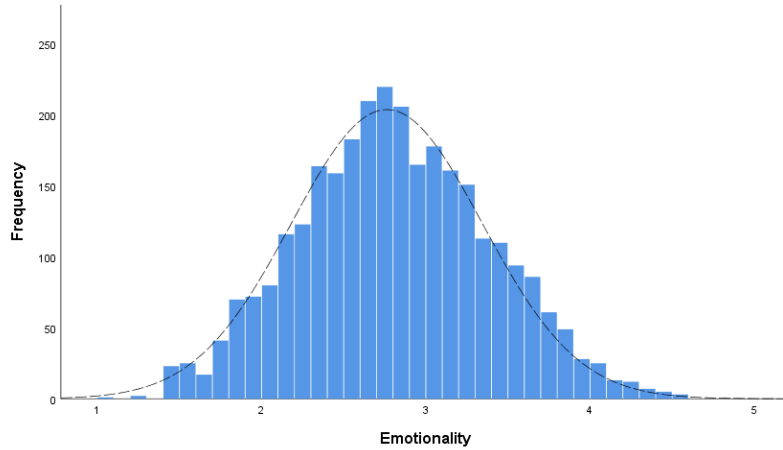


Figure 4-3. The distribution of Emotionality in our online sample. The black dotted line signifies the normal distribution.

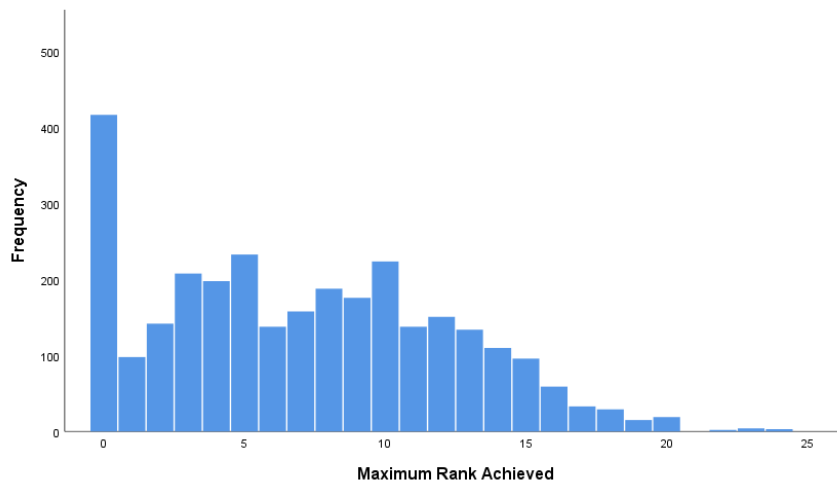


Figure 4-4. The distribution of the Maximum Rank Achieved in Hearthstone in our online sample. Ranks in Hearthstone are inverse with 5 for instance being a better rank than 10. Note that this variable does not appear to be normally distributed with a large number of individuals having achieved the rank of “Legend” (“0”) which is the maximal rank attainable.

Maximum Rank Achieved seems not to be normally distributed which is why I used a non-parametric correlation.

4.3.2.3. Rank and Emotionality Correlations

There was a weak significant correlation between Emotionality and Maximum Rank achieved in Hearthstone with $r_s=.11, p<.001$ (similar if we use a parametric correlation). As mentioned above Ranks are inverse in Hearthstone explaining the positive correlation.

4.4. Study 2

4.4.1. Methods and Materials

4.4.1.1. Ethics

The experiment was approved by the Departmental Ethics Committee prior to any data collection. The participants were informed that they are free to withdraw from the study without any consequences. They were compensated with either course credit or 8 pounds. They were debriefed at the end of the experiment. They detailed demographics have been published elsewhere and can be found in Kokkinakis and his colleagues (2017).

4.4.1.2. Participants

The sample consisted of 56 individuals (51 male, mean age 20.5 years). Participants completed a the WASI-II Matrix Task, three working memory span tasks, the MITE test, the Folk Physics test and the HEXACO-PI-R 100 in a lab environment (Foster et al., 2014b; K. Lee & Ashton, 2016; “Wechsler Abbreviated Scale of Intelligence - Second Edition (WASI-II) | Pearson Assessment,” n.d.). This is the same sample that was used for Chapter 3.

4.4.2. Results

It should be noted that no correlations emerged between IQ-WM and Emotionality. In this sample the Little’s MCAR test was Chi-Square = 7.125, (df = 1086, Sig. = 1.000). Thus

we fail to reject the null hypothesis meaning that our missing values are indeed Missing Completely at Random. Due to the small amount of our sample, listwise deletion did not appear as the optimal solution because it would result in a loss of power, so I opted for imputing the missing values through regression (it should be noted that the amount of missing values was less than 4 %). The resulting Distributions of Emotionality and LoL Rank can be seen in Figures 4-5.

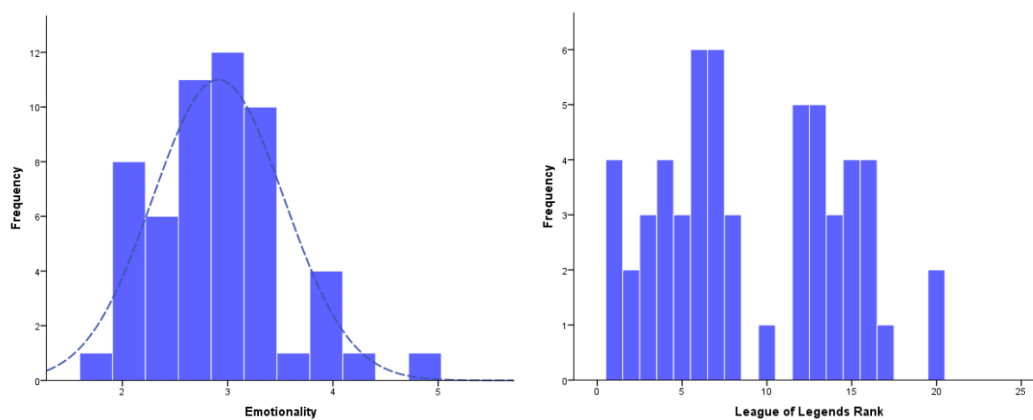


Figure 4-5. The distribution of Emotionality and LoL Rank in our laboratory sample. The black dotted line signifies the normal distribution. Note that the Emotionality distribution seems to be skewed towards the right.

Due to the nature of the distributions as well as the nature of the data (Ranking Data which are discrete and not continuous), I opted for a non-parametric correlation. There was a significant negative correlation between a player's Ranking and Emotionality with $r_s = -.27$, $p < .05$. The fact that I used the 100 item version allows us to examine the Emotionality subtraits/facets more carefully. Out of the 4 subtraits only 2 Sentimentality and Dependence were significant at the $p < .05$ with $r_s = -.25$ and $r_s = -.27$.

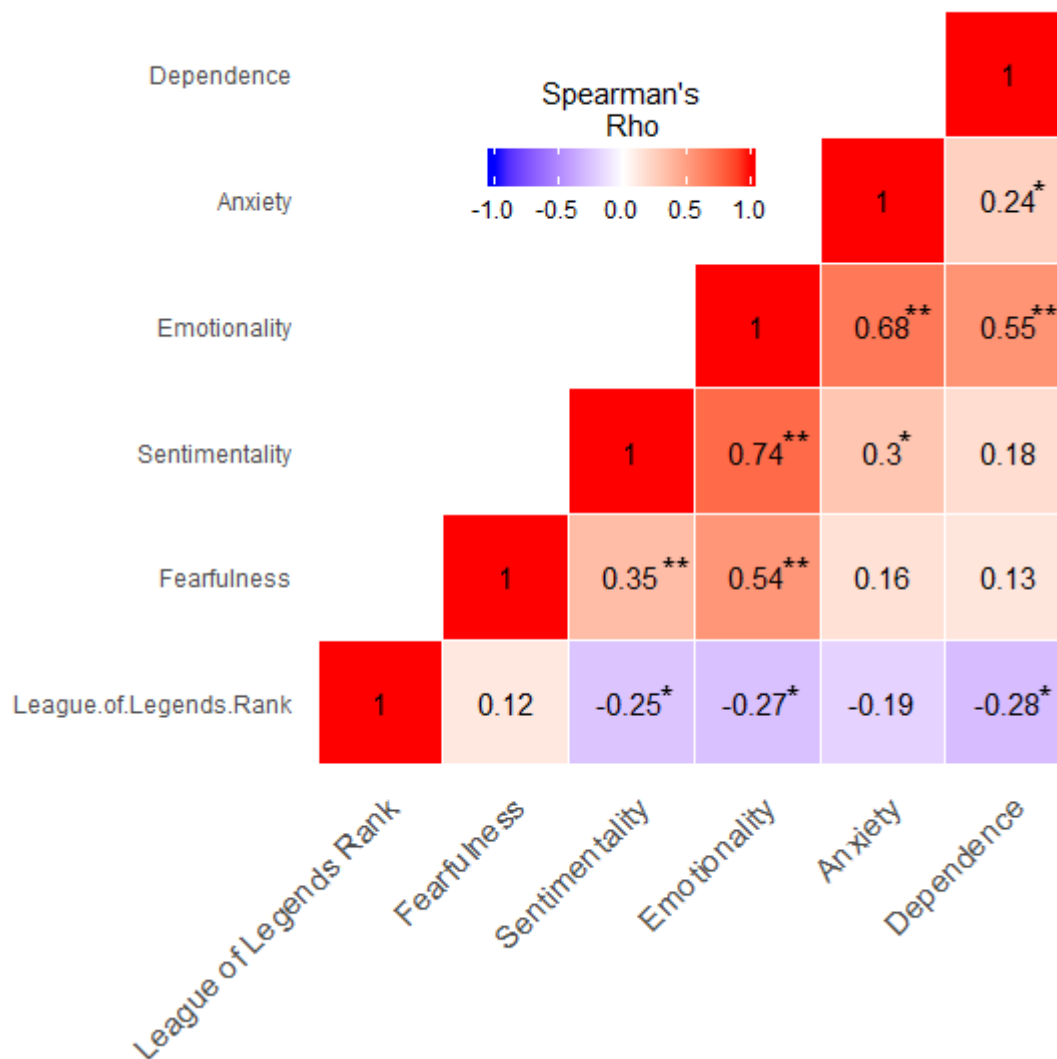


Figure 4-6. A correlation heatmap of the Emotionality Subtraits and League of Legends Rank. One can see that Sentimentality and Dependence seem to be driving the interaction.

The full HEXACO-100 descriptives and their intercorrelations can be found in Appendix F.

4.5. Discussion

4.5.1. Emotionality, its facets and Rank

In the work described above, I have essentially replicated Laakasuo's et al., (2014) experiments. In addition, we showed that relationship between Emotionality and Performance

is more pronounced in more visually demanding videogames with more attentionally demanding stimuli and threatening situations (LoL). I additionally showed that this relationship relies more on Sentimentality and Dependence, although a possible relationship between the other two subtraits should not be excluded due to our small sample size and the fact that our sample lacked participants at the lower spectrum of the Rankings. I have avoided looking at any other correlations between the various facets

Aside from Laakasuo, there have been other studies that have examined HEXACO constructs with risk-taking (Ashton et al., 2008; Vries, Vries, & Feij, 2009; Weller & Thulin, 2012; Weller & Tikir, 2011). These include Conscientiousness, Honesty-Humility and Emotionality. In our studies only Emotionality has been consistently linked with performance as measured by Rank. Similarly, Weller & Tikir (2011) identify Emotionality as a robust trait accounting for variance in multiple risk taking domains. They conclude that more Emotional individuals may have biased risk perception attitudes due to a biological ‘hypersensitivity’ to the perceived outcomes, and possibly punishments, of an action.

Using this framework it is not unexpected that we found these negative correlations with Rank, with more Emotional individuals performing worse at Ranked matches. If players do not take the necessary risks to get ahead or perceive situations as risky when they are not, that will influence their gameplay making it non-optimal which in turn will lead to a loss. The fact that Hearthstone losses are less punishing, both in terms of time investment and in a social-context, could be an additional factor of the smaller correlation.

The first subtrait of interest which correlates negatively with LoL Rank is Sentimentality. The authors define this facet the following way: “The Sentimentality scale assesses a tendency to feel strong emotional bonds with others. Low scorers feel little emotion when saying good-bye or in reaction to the concerns of others, whereas high scorers

feel strong emotional attachments and an empathic sensitivity to the feelings of others.” . In the Big 5 Sentimentality is a more closely related with Agreeableness, which is also why the authors of HEXACO, believe that Emotionality can be viewed through the lens of kin selection or reciprocal altruism (Ashton & Lee, 2007; Vries et al., 2009). A possible impact this facet may have in the in-game environment is that of unfavourable trades. If a teammate dies in an unfavourable condition/battle, an individual with high Sentimentality may attempt to participate in that ongoing battle even if they would die resulting in their death as well leading to a bigger resource delta between their team and their opponents’. However, I partially disagree with the authors if one looks at the nature of the questions; Question 95 focuses on one’s own emotions to events rather than how one reacts to others : “I remain unemotional even in situations where most people get very sentimental.”. Moreover, the other questions that are used for the Sentimentality facet are not necessarily linked to a reciprocal relationship e.g. a person feeling like crying after watching someone else cry (Question 23), maybe more prone to experiencing stronger emotions in general rather than in response to someone else. Thus Sentimentality could be viewed as a general Emotional Lability/Stability/Mood control trait that does not necessarily apply to any reciprocal relationship. This is echoed with the aforementioned poker studies where rational poker players that do not tilt have better performances in various simulated tasks since poker requires near constant risk assessment of various hands (Laakasuo et al., 2015; Palomäki et al., 2013; Siler, 2010).

The second facet implicated is Dependence. The authors define Dependence as : “The Dependence scale assesses one's need for emotional support from others. Low scorers feel self-assured and able to deal with problems without any help or advice, whereas high scorers want to share their difficulties with those who will provide encouragement and comfort.”. This facet seems highly plausible to be the core in obtaining a higher rank, since individuals

who do not rely on other's people performance may be more proactive with in-game decisions that rely on others. An example would be an individuals who buys wards, an item that provides information, rather than waiting for another character whose in-game job is to buy them do it for them. This attitude which is more goal oriented e.g. this task needs to be done rather than relying on someone else could result in a higher winrate. This is echoed by a new study where they link MMR in DOTA 2 with grit (Röhlcke, Bäcklund, Sörman, & Jonsson, 2018). Grit is defined as a form of perseverance, that is linked to high achievement, despite failure or hardships along the way (A. Duckworth, Peterson, D Matthews, & R Kelly, 2007).

Thus grit may be related to Dependence through loss resistance or resilience as shown with the following HEXACO-100 questions that are used to derive the facet : "When I suffer from a painful experience, I need someone to make me feel comfortable" and "I can handle difficult situations without needing emotional support from anyone else.". It should be noted that grit seems to not be correlated with IQ however it is positively correlated with the big five's Conscientiousness and negatively correlated with the big five's Neuroticism which further strengthens this theory (A. L. Duckworth & Quinn, 2009; A. Duckworth et al., 2007).

Although, we offered two different possibilities they are not necessarily mutually non-exclusive. It could be that since individuals have realised their own processing inefficiency when it comes to highly demanding or threatening situations they might choose to adjust their playstyle accordingly by not taking risks; essentially opting for a more low-risk-low-reward playstyle even when that is not the optimal choice. This suboptimal playstyle may cost them victories since different characters/heroes have different playstyles which may demand more aggression as part of their playstyle in order to win. By not being aggressive or taking risks a high scorer may lose the game. This low-risk-low-reward playstyle should be easily identifiable through a player's position in League of Legends by the frequency of trading hits

with one's opponent and their relative position to their opponents' movement since a more neurotic individual will avoid conflict unless the odds are disproportionately in their favour.

Finally, we should note that Hearthstone is a relatively less "intense" videogame which may account for the weaker relationship. Not only it is turn based but the game revolves around simple mathematics at the earlier stages of each match. Thus the manifestation of Emotionality may never become apparent except at the highest level of competition or at the latter part of each match when the stakes are high or the cards and the situations matchups require more complicated mathematics.

4.6. Limitations and Criticisms

The first criticism that could be made is that I used an internet sample in order to replicate Laakasuo's findings; even though Laakasuo's study used an online sample from a poker forum itself. Current literature suggests that internet samples are highly robust and reliable and perhaps even more accurate with certain personality traits due to anonymity facilitating self-disclosure (Buhrmester, Kwang, & Gosling, 2011; Davis, 1999; Gosling et al., 2004). Moreover, I completed a Missing Value Analysis and juxtaposed the descriptives of our sample with a sample acquired by the authors. Thus I believe that the nature of our online sample was not a limitation by itself but rather its very nature allowed us to find a really subtle effect.

Unfortunately, our second study has a relatively small sample size, even though it was a more controlled laboratory setting which I think had an effect on my results. Even though Anxiety did not appear to be a major component in Study 2, I believe that the failure to reject the null hypothesis could be attributed to our small sample size. Anxiety has been shown to affect performance in a number of visuospatial and working memory tasks (de Visser et al., 2010; Hartley & Phelps, 2012; Shackman et al., 2006; Wetherell, Reynolds, Gatz, &

Pedersen, 2002). Therefore due to the vast literature showing that it can have an inhibitory effect on performance I believe that a replication of my experiment with a much bigger sample, and possibly a more threatening videogame, will yield fruitful results.

Eysenck and Calvo (1992) make the distinction between performance effectiveness, which is simply put how well one completes a task, and processing efficiency, which is the amount of effort or resources expended in order to successfully complete a task. More specifically, even if highly neurotic individuals perform in the same way as low neurotic participants in terms of accuracy, self-report measures reveal that they were significantly more fatigued in order to maintain their high performance (Hadwin, Brogan, & Stevenson, 2005; Wilson, Smith, Chattington, Ford, & Marple-Horvat, 2006). This might account for the lack of results.

A later theory developed by Eysenck et al., (2007) that builds on this previous model is the Attentional Control Theory (ACT) which supposes that anxiety may compromise the shifting and inhibition of attention in threatening situations. The authors further create multiple hypotheses based on this theory. The first one is that Anxiety may impair one's ability to disengage from visually demanding stimuli therefore making it harder for those individuals to attend to a more higher order task at hand (stimulus-driven vs goal-directed). The second one is that if distractors are of threatening nature one can expect a decrease in both efficiency and effectiveness since the user might focus on them too much and not disinhibit in order to attend to something more meaningful. Thus if a replication is attempted it maybe of note that the new additional metrics that capture this aforementioned processing inefficiency are used and are paired in an event like manner with highly demanding in-game events; these additional metrics could be heart rate, pupil dilation and skin conductivity synchronized with in-game player deaths or big teamfights.

4.7. Implications

The first major implication is that personality itself can have an influence on the performance of a highly visuospatial videogame. This means that a competitive individual could theoretically boost their performance past the limitations of their own visuospatial intelligence by adopting or cultivating habits that limit their in-game reliance on other players or by trying to actively frame losses in a positive manner so they can be more resistant to them e.g. “learning from one’s mistakes”.

This raises the question of whether standardized techniques that modulate one’s mood or way of thinking may have an impact on player performance at high-stress in-game situations (“clutch plays”). These techniques aside from general “positive” or “grit” promoting attitudes could be influencing the allocation of attention that is mismanaged in high-stressful/emotional situations. Techniques such as quiet-eye-training or mediation, which have been shown to be effective at weakening the Attentional Blink, could be really useful at enhancing the performance of non-elite videogame players (van Leeuwen, Müller, & Melloni, 2009; Vine & Wilson, 2011).

5. Synthesis

5.1. Introduction

This chapter aims to synthesize the problems, the limitations, the potential impact of my studies and by extension this field of research itself. I will focus on the limitations of my data and videogame data in general, ranging from the way they are provided to the researcher by their industrial partners, as well as the fleeting and dynamic way that videogames and their playerbase change. I will discuss in more detail how the types of sampling I, and many other videogame researchers, used in my/our studies can influence conclusions and in what way.

Aside from limitations, I will discuss future research and the possible, direct impact of my findings in health and education. More specifically, I will explain why conventional psychometric testing may be problematic, costly and arduous in certain cases. Finally, I will create a brief theoretical framework on how this technology would seamlessly work in a health setting with sensitive populations and how we can avoid or at least limit practice effects through task periodization.

5.2. Conclusion

In my thesis I show convincingly that data extracted from videogames can be a reliable and valid way to make inferences about the general population. I linked multiple game variables with real life attributes such as Demographics, Personality and visuospatial Intelligence. Due to the nature of videogames billions of data points are constantly generated and are not fully taken advantage of. These data are mostly wasted or interpreted through the narrow lense of videogame playing. In contrast, we could use this wealth of videogame data to create fully functional individual psychological profiles that have a wide range of utility. These profiles could be used in a health and education settings (through outlier detection) or used to compliment advertising profiles by helping establish sleeping patterns or other variables. Moreover, we could use these massive data to make inferences about whole geographical areas, for instance one could wonder whether children from the lead-polluted Flint area would perform worse in videogames since lead can influence reaction times.

Another example would be if we managed to get an individual's Neuroticism score through a social media profile quiz or through voluntary submission of data. If we had that score in combination with their MMR we could make some reasonable estimations using machine learning models about their visuospatial abilities (and vice versa, if we know their visuospatial abilities and how they behave in-game, in terms of conflict we could derive an approximation of Neuroticism/Emotionality). These data could be used to help construct a better (or more complete) user profile that could be used for either targeted advertising or for improving user experience (through better and faster learning) (Hern, 2018).

5.3. Limitations and problems of videogame research

5.3.1. The Blackbox problem and its reasoning

Most videogame companies will usually provide thej users ample access to many of their detailed statistics via their API (“API Documentation - Riot Games API,” n.d.). However, even if large amounts of data are provided that does not change the core problem : many researchers have no idea how these metrics are created. An easy and obvious example would be MMR which was a core methodological point of my research (Kokkinakis, Cowling, Drachen, & Wade, 2017). Although, companies will gladly inform individuals that MMR is a win-ratio based measure, they will often omit that in the first 10 games give additional MMR points and that they tend to use additional metrics such as KDA, damage and healing done for player MMR calibration (Valve, n.d.). This applies to LoL as well with the first 10 games giving more LP; a developer, “Sapmagic”, even admits that the whole placement process can appear as a *black box* (League of Legends, 2018).

Valve’s Dota II MMR calibrations provides an interesting example where the players managed to reverse engineer, possibly through trial and error, the aforementioned additional metrics. They went on to take advantage of it by playing heroes that either do inflated or near unmissable damage (Zeus ultimate ability does magic damage to all visible opponents). A second hero named Oracle can damage and heal his own teammates (Valve, n.d.; “Zeus abusers,” n.d.). It appears that the Dota 2 developers did not limit the metric “Damage and Healing done to heroes” that was used for calibration to one’s opponents. Thus an oracle player could consistently and repeatedly spam one skill on one’s own teammates and achieve a relative high MMR even when losing multiple games.

Many games have metrics or systems that are unavailable to the public, essentially a blackbox which companies tweak all the time without even telling the players. This of course makes sense from a practical point of view. A player could adopt a “selfish playstyle”, as mentioned before, where they avoid dying at the expense of his team’s victory thus “padding” their own stats and rising in the ranks regardless of the game’s outcome. This playstyle would then start to spread out with other users following a similar meta-strategy thus rendering the parts of the game unplayable. Thus, a (noble) lie regarding a metric such as MMR calibration both prevents abuse while it also helps the company keep its “secret sauce” secret. However, the occasional unreliability of these measures could potentially lead academics who are using these metrics to many Type 1 or Type 2 errors since they act on false assumptions and take a company’s word as the truth. In conclusion many videogame companies avoid revealing their secret sauce and that is a rational choice, in most cases at least. Researchers should be mindful when looking at the data and always examine them from a bottom-up way rather than a top-down manner based on what the company says the metrics do or are supposed to do.

5.3.2. Different Genres of video games may tap different abilities and conclusions may not generalise even in the same game across patches and servers.

As previously mentioned by Latham and his colleagues (2013), an important issue is that of conflating videogame genres. An FPS may tap more reaction times when compared to Dota 2 and Dota 2 might rely on more crystallised knowledge due to its ‘deeper’ nature, at least in regards of complexity. Moreover, conflating the same patches can be problematic even within the same game. There will be patches that will completely change game effectively forcing individuals to relearn the game which might cause a drop in their

performance requiring a new phase of adaptation to the game's new dynamic environment or variables (Valve, 2016).

This is specifically important when one takes into consideration the link I found between visuospatial intelligence and MMR in LoL, as shown in Chapter 3. The first obvious point is that this relationship could not be found in a much simpler game or game genre that does not allow for individual differences to manifest themselves. If I played "Tic-tac-toe" or a game of "Snakes and Ladders", against Albert Einstein and won, no meaningful comparisons could be made. "Tic-tac-toe" is limited in each moves which does not allow for a superior opponent to use their creative thinking, faster reaction times or working memory. Similarly, "Snakes and Ladders" is simply based on luck and no "skill" is involved. The second problem is that this link could vary in its strength even in a game of a different genre such as DOTA 2. Due to DOTA's complexity, experience might play a much bigger role than fluid intelligence, although individual differences should manifest themselves in critical, demanding moments such as teamfights.

A study by Röhlicke and his colleagues (2018) that focused on DOTA 2 and MMR could not replicate my findings. I believe that a problem with their methodology lies with them selecting individuals that have only played 110 ranked games of DOTA 2. Even though 110 games maybe enough LoL games to gauge some form of final stable MMR in LoL, that could not translate to a different game (of the same genre) with a much steeper learning curve such as DOTA 2. This is echoed by Bonny & Castaneda (2017), where they found that number processing was predictive of MMR gain (and a slight higher correlation between MMR and number ability for "practiced" participants).

Cross-comparisons can be problematic between different servers as well. Different in terms of skill/performance or a reaction to a stimulus (for instance punishment in response to

negative in-game behaviour). Various professional players have consistently stated that the Korean servers may be more competitive even at lower Ranks and that for instance the Philippine servers and some professional North American DOTA 2 players have stated the “scene” is way more problematic than the European one (Babaev, 2019; Dager, 2019; Seok & Hong, 2018). Thus a researcher should recognise that different servers correspond to different regions with different “native” populations which influence those statistics . A western server could be more individualistic when compared to an Eastern one, and that might also apply to they way these individuals play the game, how often the quit and why they report or commend an individual . Thus merging behavioural data from different servers should be avoided unless the researcher knows that the statistics will not be influenced by a culture that predominantly uses those servers.

5.3.3. The depreciation of videogame data

As mentioned previously videogame data tend to become available through some form of API (“API Documentation - Riot Games API,” n.d.). Data storage creates costs and companies are always looking for a way to cut down on costs. These creates a number of issues which may be problematic for a researcher.

The first problem, I and many of my colleagues faced, is that a company might remove its data or partially degrade it to a less detailed form, either willingly or unwillingly. In my nickname study, I did not receive the “offensive name category” that existed at that time as a report option, probably because the company wanted me to find a different way to identify “toxic” players (Kokkinakis, Lin, Pavlas, & Wade, 2016). This degradation of data had an important effect on the experiments power; one could only report someone for one

issue and thus this category which was removed probably “absorbed” multiple reports since these ANTPs had offensive names.

Accidental degradation of data is also a problem. Riot API does not publicly prove the full matches a player has played many years ago, which means that unless the researcher was provident then that data might have been lost for ever. Similarly, although the Riot API provides the user with a player’s *exact* rank now, after the season has ended it defaults to providing only the general division. This of course is problematic because as the developers have admitted themselves the skill-gap between divisions even in the same tier can be huge (League of Legends, 2018). In conclusion, the research should have knowledge of the game so they can understand how subtle differences or omissions in the data they are given (lack of “Offensive Name” category”). Moreover, constant backups and web-scraping, from multiple different sources, needs to be done diligently to ensure that the data used in studies are valid and do not become lost or degraded.

5.4. Sampling

5.4.1. Survivorship Bias and Special Populations

The “Survivor Bias” is an important point that individuals need to take into account when looking at videogame populations (Austin, Mamdani, Walraven, & Tu, 2006).

Videogames may require of the player some (relatively) fast reaction times as well as some form of visuospatial working memory/intelligence for them to be successful, as shown in Chapter 3. Individuals that lack these characteristics may face consecutive losses that discourage them from playing the game. This leaves behind the survivors, who were also better performers either in terms of cognitive ability or emotional resilience (increased tenacity/higher grit/decreased Emotionality). Thus participants that get some early wins to

keep on playing, creating an illusory practice effects in which more games correspond to more MMR (simply because the bad performers have opted out of Rank play). A researcher should thus try and be mindful to not make generalisations between Ranked and Normal modes.

This expands to a distinction between the videogame population and the “normal” population. The reader will quickly point out that a bad performer will play ranked and their MMR will calibrate at below average effectively making the aforementioned point null. I am stating that players that are below average in terms of MMR/skill (Silver 3-Silver 4), may still score higher than their peers. If one examines the WASI-II tables, for the 20-24 age range, the average raw score is 20 while in my sample it is 24 (“Wechsler Abbreviated Scale of Intelligence - Second Edition (WASI-II) | Pearson Assessment,” n.d.). This could mean that an average MMR in LoL, and by extension videogames in general, corresponds to higher visuospatial intelligence, when compared with the general population. This has direct applications in the education and health applications that will be discussed in a later part below.

5.4.2. Range Restriction

Range Restriction is extremely important because it can lead to Type 1 Error (Ackerman, 2014a, 2014b; Detterman, 2014; Vaci, Gula, & Bilalić, 2014). Not all bivariate relationships follow a perfectly linear pattern, with some relationships showing diminishing returns at the higher level (Howard, 2009, 2014). My sample in the third chapter was severely range restricted since there were no lower performing participants that played at a lower division (“Bronze Division”). An example of how restricted sampling influences findings comes from Detterman (2014) where he plotted the correlation between height and points scored in the NBA. The correlation was almost non-existent ($r=.009$) with Detterman (2014)

jokingly stating : “Based on these near-zero correlations one would be forced to conclude that height had nothing to do with success in basketball.” Similarly, if we look at chess and IQ the relationship seems to be weak or even non-existent at all (Detterman, 2014). Of course we know that in the case of NBA, stating that height is not important is ludicrous. The problem of course is two-fold. First, we are using a correlation rather than looking at the mean and comparing that to a normal population (the average height in the population is much lower when compared to the NBA , similarly the IQ of expert chess players tend to be higher than that of a normal population (Detterman, 2014; Grabner, 2014). The second is performance in the NBA is multidimensional and height seems to be an important variable other secondary variables may come into play for instance muscle mass or tenacity. To paraphrase Ackerman (2014), in elite or high levels of competition individuals with the less desirable trait (height or visuospatial intelligence) have already been excluded which leads to restricted sampling from the beginning. One can see a pictorial representation of range restriction below for visualisation purposes.



Figure 5-1. A sample before and after restriction. We can see that in point a if we do not sample a range big enough the true relationship is masked. Similarly, if we correlate number of legs with performance in a sport performance in B.

Thus, care should be taken when weak relationships are found between two variables when we do not sample individuals at the lower end of the spectrum, where performance measures

are more accurate (Deary, Penke, & Johnson, 2010). Researchers should not portray lower Rankings in a negative light in order to avoid the social stigma associated with it. Thus participants will admit to their lower rank and will not have a trouble when they need to be tested. In turn this will lead to a more representative sample whose conclusions are more generalisable, while also avoiding Type I error.

5.4.3. Cross-Sectional Vs Longitudinal

Videogame academic research has an old history, however using massive online datasets is arguably novel (“Richard A. Bartle: Players Who Suit MUDs,” n.d.; Tekofsky, Spronck, Plaat, Van den Herik, & Broersen, 2013; Thompson, Blair, & Henrey, 2014; Yee, Ducheneaut, Nelson, & Likarish, 2011). A core detail in many videogame studies that uses human participants is the one-off collection of aggregate stats at the end of a competitive season or a testing session (Kokkinakis et al., 2017, 2016; Tekofsky et al., 2013; Thompson et al., 2014). Most notably all my studies, or the ones I have been involved with, are of cross-sectional nature with one exception that is still limited in terms of sampling duration (Aung et al., 2018; Kokkinakis et al., 2017, 2016). This of course is perfectly understandable due to the infancy of the field as well as due to its practicality. Videogames change by introducing new patches rendering comparisons hard at best and impossible at worst. Moreover, videogame companies go bankrupt or they discontinue their titles and data scraping is arduous and potentially economically inefficient for videogame companies; they can find it hard to justify their data analysts providing a curated dataset for some academic, especially when they think that the academic will not provide any economic gains to the company or that their in-house data analysts are better and thus this process is superfluous at best and economically damaging at worst.

Due to these reasons a large number of studies rely on cross-sectional datasets, which are a one-off collection of aggregated stats, rather than employing a longitudinal design, which is the testing of the same individuals at different points in time, usually three or more (Ployhart & Vandenberg, 2010). Note that longitudinal studies that given us valuable information on training and performance do exist, the time periods examined in them however is quite limited when compared to other human studies that have tested participants across multiple years or even decades (Aung et al., 2018; Stafford, Devlin, Sifa, & Drachen, 2017; Stafford & Dewar, 2013). The lack or rarity of longitudinal datasets is of course not problematic per se, since these datasets can still provide extreme value for such an unexplored field, however care should be taken to avoid some common pitfalls that have been brought to our attention through the psychometric literature that used longitudinal samples with the aid of Dr. Schaie's exceptional work in the 1960s (Schaie, 1993; Schaie & Hofer, 2001; Schaie & Strother, 1968).

Before I go over how this type of sampling may influence the conclusions in videogame studies, I will briefly go over the advantages and disadvantages of each method, as well as how they have influenced the conclusions of other fields. Firstly, cross-sectional samples are based on a one-off "section" of the population meaning that the researcher obtains the end product of a different individuals together at a specific point in time (Ployhart & Vandenberg, 2010; Schaie & Hofer, 2001). Longitudinal samples consist of the same individuals tested across multiple points in time, usually using the same psychometric tests or their equivalent in order to see any notable changes between time A and time B. The advantages of a cross-sectional design are many : there are many participants because there is no participant attrition across testing at different time points, there are less time consuming and thus less costly and more practical and there are no practice effects. The disadvantages are the researcher gets a single snapshot of data in time which makes inferences of cause and

effect harder. Moreover, the full envelop of the sample cannot be seen, for instance a drug that causes a short remission of the disease followed by the disease getting stronger. Finally, the most important part which directly relates to videogames is that the merger of individuals from different generations together which creates problematic trait trajectories.

The advantages of longitudinal samples are many. They have multiple time points, thus provide a richer dataset and with the introduction of an event in-between data collection inferences can be made. Moreover, the individuals are the same which means that individual differences are eliminated which makes for better designs. Finally, since different individuals are not merged together misleading trajectories can be avoided, which was a major point of Schaie & Storther (1968) research. The disadvantages are clearly the cost of multiple testing sessions, the dropouts and the practice effects. However, one can still use specific designs to avoid practice effects such as using similar equivalents of a test that were used in the previous testing session or creating different subsamples to ensure that the administration is balanced. A unique disadvantage in videogame research is the “soft changes” of a game, that come in the form of patches, which may create two unique different environment rendering longitudinal comparisons between Time A (Patch 6.99) and Time B (Patch 7) meaningless.

The Seattle Twin Longitudinal Study used a longitudinal sample rather than a cross-sectional one and elucidated many issues. Most notably and most relevantly, it seems that the peak in fluid intelligence maybe much later in one’s early thirties rather than the one in mid-twenties as previously thought (Schaie & Strother, 1968). This can be clearly seen in Figure 5-2.

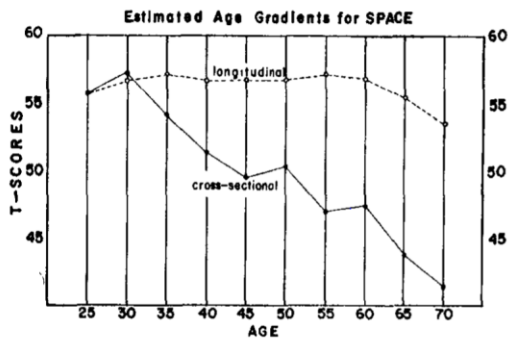


FIG. 2. Estimated age gradients for Space.

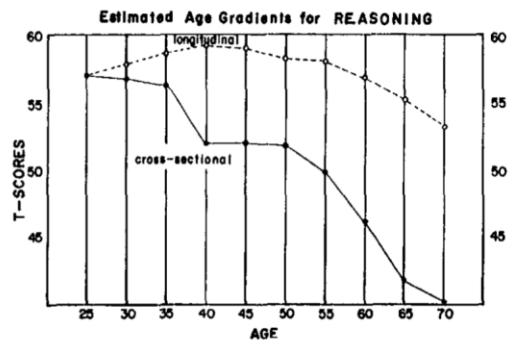


FIG. 3. Estimated age gradients for Reasoning.

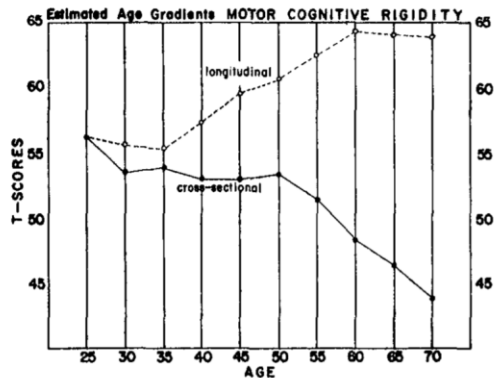


FIG. 8. Estimated age gradients for Motor-Cognitive Rigidity.

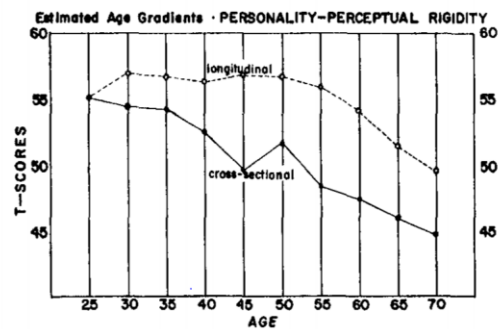


FIG. 9. Estimated age gradients for Personality-Perceptual Rigidity.

Figure 5-2. The difference between a longitudinal and a cross-sectional sample adapted from Schaie, K. W., & Strother, C. R. (1968). A cross-sequential study of age changes in cognitive behavior. *Psychological bulletin*, 70(6p1), 671. All these variables are related to videogame performance since they tap some aspect of it (visuospatial memory, logic and reasoning, reactions times, reasoning and ability to interpret new surroundings). If videogame performance follows a similar pattern we should expect a much smoother decline than what is currently shown with cross-sectional data.

This of course happens because the sample contains individuals from older generations who may have suffered from socioeconomic problems or from the lack of resources. Younger generations tend to outperform their elders at least up to a point as shown by Flynn's research (Flynn, 2013). Thus the true peak is misattributed earlier than it actually is for the younger generations.

This has direct relevance to videogames and my research. Older videogame players did not have access to better nutrition or education, similar to younger individuals; nutrition and education is one of the possible causes for the Flynn effect (Flynn, 2007). Moreover, many games such as Starcraft or Dota are new which means that older gamers may not have had the chance to master them which puts them at a disadvantage since they missed their sensitive period for learning that game (and now due to life obligations they cannot put in the required consistent amount of practice in order to reach their true potential). Finally, a possible explanation of the Flynn effect was proposed by Brand and is that we are simply better at test-taking (Brand, Freshwater, & Dockrell, 1989; Flynn, 2013). Younger people have become better at test taking overall due to the way subjects are taught in school as well becoming better at taking calculated risks when guessing a question, saving on mental resources trying to confirm that that item is the answer (Brand, Freshwater, & Dockrell, 1989; Flynn, 2013). The author shares a similar point of view for videogames. We have become better at strategising about videogames in general and we have started taking calculated risks which lead to unorthodox and new and improved strategies. With the advent of computers, APIs and forums advanced statistics such as heatmaps, effective HP, microing techniques (“cloning”) and value maximisation have become commonplace for anyone willing to learn which was not the case many years ago when an interested individual may have a harder time understanding or even coming across the aforementioned achievements.

In conclusion, psychometric research in combination with Schaie’s findings provide us with three key points. Due to cross-sectional sampling used in many studies, we may be underestimating the peak performance of individuals which may be in one’s early thirties rather than in one’s early to mid twenties as current research shows (at least for this generation). Secondly, the decline will be way slower and more forgiving, so rather than “over the hill past age : 24” the title should read “slow fading away past age 32” (Kokkinakis

et al., 2017; Thompson et al., 2014). Three, it is the author's opinion that younger generations will become better at videogames at least for a decade and a half similar to the Flynn effect (Flynn, 2007). More young pro players that are consistently good will start appearing. This can be attributed to Esports becoming a more legitimate career as well as to younger individuals playing dominant, stable or genre defining games from a younger age (sensitive period) as well as learning how to play them properly (developing good habits and building upon them) which will in turn allow them to revolutionise their respective genres.

5.5. Establishing Rank Correspondence with Standardised IQ ranges based on Age

Average Rank in a videogame that taps Cognitive Abilities does not imply average cognitive abilities. It implies average Cognitive Abilities in relation to the highly specialised population that is playing that videogame. If I am the weakest person at weightlifting in the Olympics, I am still extremely strong in relation to the average person. Similarly, if an older individual has average rank in LoL or DOTA II and the game is played predominantly by individuals in their early-to-mid-twenties where fluid abilities tend to peak (at least using a cross-sectional sample) then that older person has an above average IQ, since IQ is defined based on age (where the raw score is normalised/standardised based on the specific age and socioeconomic group one belongs to) (Raven, 1989, 2000). Thus, if we need to use videogame rank as a more elaborate metric, it is imperative that we use large samples of individuals from various age groups and various socioeconomic backgrounds. Standardising the in-game rank for many games means that we can apply videogame scores as alternatives to psychometric tests. This has uses in education for screening 'gifted' and 'lagging' children or the Health industry, as a preventative measure that identifies at risk populations.

5.6. Future Application Areas

I will further elaborate on how we can use videogame data to create a coherent framework in two distinct areas : Education and Health (Disease Prevention). I will explain the logical first steps that need to be taken, as well as some pitfalls that might not be easily seen by the layman.

5.6.1. Theoretical Framework for Education

Due to the link between videogame scores and fluid intelligence, we can use videogame rank to identify gifted children. Good spatial skills have been linked to S.T.E.M. performance in multiple studies (Sorby, Veurink, & Streiner, 2018; Tam, Wong, & Chan, 2019) Psychometric testing maybe time consuming and costly when done at a larger scale, with tests costing several hundred dollars (“Raven’s Progressive Matrices | Pearson Assessment,” n.d.; “Wechsler Abbreviated Scale of Intelligence - Second Edition (WASI-II) | Pearson Assessment,” n.d.). Moreover, if it is a one off testing session, an IQ test may mischaracterize an individual just by chance (even though at the sample level these random fluctuations may cancel themselves out). Videogame rank is a way to counter this since a) it is inexpensive and non-invasive since the individual is doing all the test work with their peers (effectively it could be characterised as a crowd-sourced IQ test) b) there are lesser individual mischaracterisation, since the final Rank is the product of hundreds of games where the individual is the only stable and consistent variable.

Even though videogame scores could help identify children with high visuospatial abilities through their rank, the case might not be fully straightforward. Although, in theory even the simple rule of 2.5 standard deviation over the mean could be used, there is bound to be an amount of children that will be false positives. Thus what should happen, at least in the early of this technology, is to a) identify the gifted sample through a collection of skill/performance variables (APM, PAC, MMR, MMR trajectory – delta among others) b)

use a full range of psychometric tests including a verbal IQ test c) assign all the children to the gifted class d) use machine learning to refine the identification process based on the children's performance (grades, teacher evaluation) and the psychometric tests collected at the second phase (Aung et al., 2018; Thompson, Blair, Chen, & Henrey, 2013; Thompson et al., 2014).

Note that socioeconomic status should be taken into consideration, for an example a child that achieves Platinum 1 rank from a less privileged region should in theory be better when compared to a child that reached Diamond 5 from a richer environment (better nutrition, education, equipment, better FPS and ping among other factors). In conclusion, videogame rank can be used as an inexpensive way to identify gifted children through their performance in their ladder at an earlier age and through constant iteration and refinements a machine learning model can be created that eliminates any false positives.

5.6.2. Theoretical Framework for Health

A large number of diseases have a genetic component that not only influences the risk of getting the disease but also its manifestation (Pihlstrøm, Wiethoff, & Houlden, 2017; Tsuang, 2000). Thus a body, such as the NHS, could identify sensitive populations through a) questionnaires b) parental history or genetic screening. Some examples of identifying such as populations could be by administering a questionnaire that establishes a baseline of their initial abilities as well as establishing the neurological risk they are at by measuring the constructs of interests for instance schizotypy if the disease of interest was schizophrenia (Bošnjak-Nadž et al., 2011; Mason, Linney, & Claridge, 2005). This population would subsequently be monitored either through their scores in commercial videogames or through an application they have installed in their phone which contains mini-games that are designed

to measure cognitive capacity. These videogames exist today mostly for commercial purposes (to screen job applicants), however they can easily be adapted for this purpose (“About Us,” n.d.; Steger, Schroeders, & Wilhelm, 2019).

At the early stage of this Framework when no big data has been gathered, researcher will need to rely on three possible methods to screen for the expression of a disease. The first and most obvious one is to cross-compare the participant’s scores against that of a healthy population. Cognitive scores are expected to decline as one ages and any sharper deviation could flag the participants as at risk of the disease having expressed itself since the expression of these diseases is accompanied by cognitive and motor deficits (Bošnjak-Nadž et al., 2011; Kensinger, Shearer, Locascio, Growdon, & Corkin, 2003; Larson, Walker, & Compton, 2010). The second way would be to use longitudinal data and have the participant as a control for themselves, meaning that we contrast each new individual score with the participant’s newest performance. We need to be mindful of practice effects, bypassing them by “circling” through the various cognitive tasks, as well as time of day and day of week effects i.e. if the participant underperforms on a late Saturday night it could be flagged as a false positive due to them going out with their friends and becoming intoxicated which would in turn influence their reaction times. Thus, if a participant’s cognitive and motor functions fail over a threshold or if they do not even perform the test at all (especially if they performed the test online like a clock for many years) then a Health body such as the NHS could dispatch an ambulance or send a check-up e-mail based on the severity of the decline; this of course will have to be refined through a machine learning algorithm and the first iterations of this framework are bound to be unstable. In conclusion, after many iterations of this procedure, a rich dataset will be collected. This dataset will include participants’ initial scores, rate of play, rate of decline as well as demographics among other key statistics including age of death and any key events that preceded it. We should be able to use some

form of machine learning in combination with this rich dataset to predict healthy aging trajectories of both at-risk and not-at-risk individuals with a much higher accuracy.

Appendix A

Games used in the study categorized using an expert task analysis.				
Game	Group	Description	Measure	Source
Silversphere	Reas	The goal is to enter the blue vortex in each level by moving a silver sphere around a maze. Players plan how to use blocks with different features to create paths to the vortex while avoiding obstacles.	Max level	miniclip.com
Bloxorz	Reas	The aim is to get a block to fall into a square hole at the end of each stage by rotating and moving the block across platforms of different configurations and features while avoiding falling off from a platform.	Last level	miniclip.com
Sushi-Go-Round	Reas	Players pretend to be a sushi chef. The goal is to learn the different recipes, serve a certain amount of customers with the correct recipes, clean the tables, order ingredients and appease the customers.	Max score	miniclip.com
Blobs	Reas	The aim of the game is to keep jumping blobs until only one remains. A blob can only be jumped in certain directions and a blob that was jumped over is removed from the board.	Last level	miniclip.com
TwoThree	Reas	The aim of the game is to shoot down rapidly presented numbers by subtracting them exactly down to zero using only units of 2 or 3 and sometimes switching between target numbers to shoot.	Mean points	Armor Games
Memotri	WM	Participants uncovered three cards at a time and had to remember the specific items associated with each with the goal of identifying all matching sets by uncovering each set in a single trial.	Max points	Platina Games
Simon Says	WM	The aim is to replicate the whole sequence of light and sound conjunction patterns played in each level.	Mean score	neave.com
Memocubes	WM	Players are presented with nine cubes with forms on each surface. The aim of the game is to match forms of the same color and complementary shape by rotating and remembering the location of matching cubes.	Mean score	Platina Games
Round Table	WM	A table is divided in marked sections that each hide a number of marbles. The table rotates at each turn. The aim is to get more marbles than the opponent by remembering which segments that still have marbles left.	Mean score	Platina Games
Oddball	WM	In each trial, the aim is to identify the new ball in the display before time runs out. The display gets increasingly complex as all previous balls remain on the screen.	Mean score	Armor Games
Filler	ATT	Player has to fill 2/3 of the screen by creating filler balls of different size while avoiding bouncing balls.	Max score	kongregate.com
Enigmata	ATT	Players navigate a ship through space. The aim of the game is to gather objects that provide power or armor, destroy opponents using the collected fire or armor, and avoid enemy fire and debris.	Max score	maxgames.com
Dodge	ATT	The aim of the game is to avoid enemy missiles that are actively chasing the player's ship and destroy enemies by navigating around the enemies so that their missiles destroy each other.	Max level	Armor Games
Cathode	ATT	Players navigate around a space to trace different forms while avoiding colliding with flickers.	Mean score	Armor Games
Music Catch 2	ATT	The aim of the game is to catch certain shapes appearing on the screen while avoiding red shapes.	Max points	reflexive.com
Digital Switch	PS	Players switch digbot positions to correspond to falling targets and collect coins matching the bot color.	Max points	miniclip.com
Crashdown	PS	Players prevent the wall from reaching the top of the display by clicking on three or more adjacent same-colored bricks to remove them.	Max level	miniclip.com

25 Boxes	PS	Two sets of matrices are presented side by side. Players search for a character in the first matrix and indicate its location on the blank matrix.	Max score	Platina Games
Phage Wars	PS	Players spread their parasites and overtake all other parasites to become the dominant species.	Last level	Armor Games
Alphattack	PS	Players prevent bombs from landing by pressing the characters specified on the approaching bombs.	Max points	miniclip.com

Appendix B

Abstract

Multi-player online battle arena games (MOBAs) are large virtual environments requiring complex problem-solving and social interaction. We asked whether these games generate psychologically interesting data about the players themselves. Specifically, we asked whether user names, which are chosen by players outside of the game itself, predicted in-game behaviour. To examine this, we analyzed a large anonymized dataset from a popular MOBA ('League of Legends') – by some measures the most popular game in the world.

We find that user names contain two pieces of information that correlate with in-game social behavior. Both player age (estimated from numerical sequences within name) and the presence of highly anti-social words are correlated with the valences of player/player interactions within the game.

Our findings suggest that players' real-world characteristics influence behavior and interpersonal interactions within online games. Anonymized statistics derived from such games may therefore be a valuable tool for studying psychological traits across global populations.

Keywords:

Video games; data mining; age; personality

1. Introduction

Online video games are played by hundreds of millions of people worldwide and fine-grained statistics on each game are constantly relayed to centralized servers where they can be stored and analyzed. These games often require complex team strategies and permit direct personal interactions mediated by real-time chat, as well as inter-player rating mechanisms. They therefore represent a rich potential source of data for psychological investigation.

Previous research on relating personality traits to video game characteristics have often correlated findings from personality questionnaires with game data: either statistics collected within the game environment, or statistics about the amounts or types of games played (Chory & Goodboy, 2011; King, Delfabbro, & Griffiths, 2013; Park, Song, & Teng, 2011; Teng, 2008; Worth & Book, 2014; Yee, Ducheneaut, Nelson, & Likarish, 2011). This approach is valuable because personality questionnaires provide verified indicators about stable, real-life personality traits. However, respondents may respond untruthfully even to questionnaires administered anonymously across the internet and completing these questionnaires is time-consuming, thereby limiting the number of individuals who can be included in each study.

An alternative approach to the psychological analysis of gaming data is to ‘mine’ very large datasets for scientifically relevant relationships. This approach is interesting for several reasons. First, it is valuable to ask whether large datasets of this type are useful for statistical analysis at all. It may be, for example, that all players adopt a single ‘optimal’ strategy that leaves little room for personal variability, rendering these datasets uninteresting from a psychological viewpoint.

Secondly, if players do seem to exhibit systematic differences in behaviour, it might be that some of this variance is linked to real-world characteristics such as age, gender or personality

(Worth & Book, 2014). Understanding these relationships could provide valuable information about these characteristics at a population level, and this information could be used as a preliminary screen to identify subjects who may be suitable for further testing. Finally, from a system design point of view, if reliable metrics on player behaviour can be established, they can be used to improve the social environment within the game.

1.1. Hypotheses

Here we examined correlations between the valence of in-game interactions and estimates of player age and anti-social tendencies in the massive online battle arena game 'League of Legends' (LoL). Here we define anti-social tendencies as being a propensity to engage in

(FIGURE 1 HERE)

Figure SEQ Figure * ARABIC 1. Gameplay within League of Legends. a) A screenshot of a LoL match in progress. A small portion of the playing arena is shown (illustrated in the small inset box, bottom right). Individual player 'summoner names' or 'usernames' appear above the human-controlled characters along with a health indicator. b) In-game chat. Players are able to communicate with each other both during and after a game. c) Sending negative and positive reports is possible after each game ends. Here, a player is choosing to send a positive 'Honor' report about a teammate.

behaviour that breaches societal norms and which is likely to cause offense to a large proportion of people.

Because we are harvesting a large, anonymized dataset this study is correlational: We use two pieces of data extracted from usernames and use them to make estimates about the players real-world attributes. We then describe how these estimated variables correlate with in-game behaviour as assessed by the game-based reporting system. We discuss methodological issues relating to the accuracy of these inferences in detail at the end of the paper.

In LoL (Figure 1) players join small, competing ‘teams’ that proceed to challenge each other for territory and an objective (overtaking the enemy base) in a relatively short time period (typically < 1 hour). The precise details of the game are beyond the scope of this paper but there are abundant descriptions in online sources (“League of Legends,” 2013). LoL is currently one of the most popular video games on the planet with an estimated 27 million online players every day (Gaudiosi, 2012). There are regular professional LoL tournaments with prizes worth millions of dollars and top players are eligible for US “internationally recognized athletes” visa status (Blake, 2013).

LoL players communicate through a real-time chat facility. This facilitates coordinated game play but it also allows players to interact socially. Players are also encouraged to evaluate their teammates at the end of each game. For example, players can praise each other for their teamwork or friendliness by sending ‘Honor’. Alternatively they can submit ‘Reports’ chastising other players for deliberately playing badly or sending abusive messages through the chat system. This report system allows us to gather information about the average valence of each player’s interpersonal interaction within the game environment. We hypothesized that if players’ real-world personality types predict their behaviour within the game, the valence of these interactions might correlate with factors that are related to real-world behaviour. Two such factors are players’ ages and their tendency to use foul or offensive language in their public usernames (DeWall, Buffardi, Bonser, & Keith Campbell, 2011; Holtzman, Vazire, & Mehl, 2010) – their ‘anti-social naming tendency’ (ANT).

We analyzed players’ self-chosen user names to estimate both age and ANT. Many of these user names contained information that informed us about these parameters. Specifically, players often embed their birth date in their user names (e.g. goodplayer1996) and in a separate analysis we show that these dates are highly-correlated with the self-reported ages of the players in the registration procedure. In addition, many usernames contain explicit or

lightly obfuscated expletives, racial slurs and boasts that are clearly designed to attract attention (e.g. 'g0ats3x'). Players must invest some time in generating these ANT names as multi-player online games typically have simple filters in place to block straightforward examples of offensive language.

Once we had identified user names that appeared to contain either age or ANT information, we asked if there was a relationship between ANT or age and the average valence of reports that each player sent or received within the game. We found that both age and ANT are predictive of in-game interaction valences as measured by honors and reports. Importantly, we find this effect for both incoming and outgoing ratings (in other words, ratings generated by a player and directed towards other teammates or, alternatively, ratings generated by teammates directed to a player).

2. Methods and Materials

2.1. Data sources

Data were provided by the US-based company Riot Games (Santa Monica, CA)—the creators of League of Legends. To improve internet connectivity, Riot Games maintains servers around the world dedicated to particular geographic regions. The data described here were obtained from servers based in North America (NA), Western Europe (EUW), North Eastern Europe (EUNE), Turkey (TK), and Brazil (BR). Riot Games supplied a representative, random sample of 450,000 data sets—one for each player. This large dataset comprised of 100,000 players on each of the NA, EUW, EUNE and BR servers and 50,000 players from the TK servers. The data represent a snapshot of the accounts on the different servers on June 13, 2013. All accounts in the dataset had been created after November 1st, 2012. The number of data sets was chosen to be as large as possible while still remaining computationally tractable.

Our analysis of anti-social user names was based on data from just the NA server (allowing us to identify English language epithets). Our age analysis was based on all available datasets.

Strict controls were imposed of the type of data that were analysed. Data were collected and analysed in accordance with guidelines from both the Association of Internet Researchers (Markham & Buchanan, 2012) and the American Psychological Association (Kraut et al., 2004). It is important to note that only anonymized data sets were analysed. Researchers had no access to personal identifying information and no modification of players' online experience was performed as a result of this research. All players had agreed to Riot's Terms and Conditions as part of the LoL registration procedure and these explicitly allow LoL to use their data for research purposes. All procedures described in this paper were approved by the University of York Department of Psychology ethical review board.

2.2. Interaction valence

After each League of Legends match, players are allowed to generate feedback on the behaviour of other team members via a point and click interface (Figure 1c). Feedback can be positive ('honor') or negative ('reports') and can refer to a range of predefined behaviours (for example, 'Verbal abuse - Report', or 'Teamwork – Honor'). A single click on each of the feedback buttons generates a single instance of a report. Players can honor or report multiple team members at the end of each game but can only send a single feedback (either positive or negative) to each player. The accumulation of negative or positive reports can have consequences to a player. For example, large numbers of negative reports may lead to temporary or even permanent suspension of the player's account. Riot now implement a 'tribunal' procedure that allows other players to vote on these types of punishment and we note that the statistics of these tribunal events provide another rich dataset that may also

relate to player personalities (Blackburn & Kwak, 2014). Although both positive and negative evaluations are nominally assigned to specific categories, in reality the nature of the infraction is sometimes unclear. For example, reporting a player for “intentional feeding” implies that they are deliberately playing poorly to benefit the opposing team but some aggressive players will use this accusation indiscriminately against anyone they consider to be inferior to themselves or to vent frustration when their team loses. There is also evidence that perceptions of toxic behaviour vary somewhat across cultures and geographic domains (Kwak et al., 2015).

Because we were interested in the overall valence of player behaviour, we used the mean of the combined ‘report’ and ‘honour’ metrics as scalar representations of negative and positive interaction. Outliers were removed using a robust outlier labelling heuristic (Banerjee & Iglewicz, 2007; Hoaglin, Iglewicz, & Tukey, 1986) which typically removed fewer than 1% of the data points. Report and Honor values were divided by the total number of games played and then log-scaled. Rates, rather than absolute levels were used to avoid conflating number of games played with average levels of anti-social or altruistic behavior. The log transform was important to ensure that data distributions were approximately normal and therefore amenable to parametric statistical analysis. The resulting datasets were found to have equal variance as assessed by Levene’s statistic.

Because the number of samples in each group was very high, these distributions were still found to be non-Gaussian by standard tests (Kolmogorov-Smirnov and Shapiro-Wilk tests; $p < .001$ in both cases) but inspection of Q-Q plots indicated relatively minor deviations. For the sake of completeness, we performed both parametric (ANOVA) and non-parametric (Kruskal-Wallis – with p-values indicated by ‘KW’) tests on our datasets and the results were found to be almost identical.

2.3. Antisocial names

A script containing a lexicon of common swearwords, slurs and sexual epithets as well as attention-drawing words and simple alphanumerical variations was created in MATLAB (Mathworks, MA). The list of words was derived initially from an online list (<http://www.noswearing.com/dictionary>). Additional common epithets and attention-seeking words were added by experienced game players and alphanumerical variation of the words (e.g. “g0ats3x”) were also added algorithmically because players often use them to by-pass filters (Blashki & Nichol, 2005). Because we used databases of English language epithets, we restricted our search to data from the North-American Server (100,000 names). The full list of substrings used to identify antisocial names is provided in the supplementary material. This list of target words was not exhaustive but it nevertheless identified over 2000 antisocial names from the North American server. We asked whether mean Honor and Reports sent and received (four statistics in total) were different between players with ANT and the control group. To avoid issues of multiple comparisons resulting from performing four separate

Table SEQ Table * ARABIC 1

Note. Values are log-scaled means of incoming and outgoing ‘Reports’ and ‘Honor’ feedback for each player.

t-tests, we used a standard one-way ANOVA to evaluate the statistical significance of pairs of group differences. ANT data were compared to an equal-sized random sample of players with non-antisocial names extracted from the same server. Statistical analysis was performed in SPSS (SPSS IBM, New York, U.S.A) and Matlab (Mathworks, MA). Parametric means testing is generally robust to small deviations from normality when sample sizes are large and equal—as our datasets were (see above). Because significance depends on group size, we also include measure of the raw effect size in our analysis. The descriptive statistics for ANT vs non-ANT data are shown in Table 1.

2.4. Age

Age data were extracted from all servers. Years are conventionally indicated using either four (e.g. 1987) or two ('87) digits. Consequently, an automated script identified dates within an appropriate 2- or 4-digit range (1985 – 2002) at either the beginning or end of the nickname ("1987Nickname", "Nickname87"). Because we were interested primarily in developmental changes up to adulthood, and because statistical tests on very small groups are unreliable, subjects over the age of 20 were not included in our analysis.

Clearly, not all instances of two or four digits matching a 'year' template actually indicate players' birth years. To examine this, we obtained an additional dataset: the years of birth reported to Riot during the game registration procedure. These represent an independent, noisy estimate of player age. Ultimately, we obtained a total of 11,630 players who passed all the criteria with a mean age (estimated from the usernames) of 15.9 years. The distribution of birth years from all servers between the 1985 and 2002 (inclusive) are shown in Figure 2a. We performed two separate analyses based on player ages. For the analysis in Figure 2b (where we compare reported vs extracted dates of birth to assess the reliability of name-derived age estimates) we deliberately excluded ages less than 14 (year of birth 2000) from the final analysis. We did this to ensure that our results were unlikely to have been skewed by players lying about their age deliberately to pass the registration stage (Riot imposes a nominal minimum age of 13).

(FIGURE 2 HERE)

Figure 2. a) Histograms of player ages estimated from usernames from five servers. Data are shown between estimated birth years 1985 and 2003 inclusive to illustrate the full shape of the distribution. b) Joint histogram of age estimates extracted from two different sources (N=10,299) with birth years between 1985 and 1999. The area of each circle represents the number of players in each group. Age extracted from alphanumeric usernames correlates strongly (Pearson's $r=0.60$) with age entered in the registration procedure.

Age estimates from the two independent sources (usernames and registration) were correlated.

Figure 2b shows a joint histogram of 'name derived' vs 'reported' ages for a total of 10,299 players whose birth years lay between 1985 and 1999. The areas of the circles indicate the relative number of players that fall into each year.

Our data clearly show that many players use the same age in both their user names and during registration and we find a statistically significant correlation ($p<.001$) between the two measures with a medium to strong effect size (Pearson's $r=.53$, Spearman's $\rho=.51$). We note several interesting phenomena: The number '88' is an outlier in terms of its frequency of appearance in usernames. We believe this is likely due to its dual use as a cultural signifier (see Discussion). Players also tend to over-report the birth year 1990 during registration and there is a particularly strong

correlation between players who report a birth year of 1990 and use the digits '2000' or '00' in their username. The reason for this is unclear but when this report year is omitted from the analysis the effect size for the correlation between reported and extracted date of birth increases to $r=.6$, $\rho=.58$ ('strong').

In the main analysis examining the relationship between age and interaction valence, we included the full set of 11,630 players with estimated ages as young as 11 (birth year 2002)

because ages were estimated solely from username information which is not vetted. Players therefore have no reason to ‘lie’ about their date of birth in their usernames.

2.5. Analysis summary:

- Players with antisocial usernames were identified using an enhanced dictionary lookup that accounted for alphanumeric substitutions in the NA dataset.
- Ages were estimated from the presence of two- and four-digit strings at the start or end of a name in all datasets. Cross-checking with registration data confirmed a high correlation between reported ages and the ages extracted from the user names.
- For all players, positive and negative interaction rates were computed from the means of the incoming and outgoing ‘Report’ and ‘Honor’ metrics. Rates were log-scaled to achieve near-normal distributions.

3. Results

3.1. Antisocial names

Out of the 3,229 hits in the North American Server, 1031 nicknames were rejected as false positives after visual inspection by an expert English speaker who was blinded to the statistics associated with each name. For example, there would be nothing deliberately anti-social about the name “ThePenIsMightier” despite it generating a hit in the swear word dictionary lookup. After false positive rejection, we obtained a sample of 2,198 users whose names were unequivocally designed to be anti-social – generally containing blatant racial, sexual or scatological epithets. An equal random sample of ‘control’ (non-antisocial) names from the North American server was selected for comparison so that the ANOVA was operating on groups of equal sizes. Human inspection of the control group identified no ‘false negatives’ or missed incidences of antisocial names.

We found that players with antisocial names had significantly higher *sent* ($F(1,4394) = 27.31, p < .001, \epsilon^2 = .0064, r = .08, KW p < .001$) and *received* ($F(1,4394) = 84.2, p < .001, \epsilon^2 = .02, r = .14,$

(FIGURE 3 HERE)

Figure 3 There are significant differences between players with 'normal' and 'ANT nicknames in all the traits we examined. Players with antisocial nicknames tend to have higher levels of negative incoming and outgoing interactions ('reports') and lower levels of positive interactions ('honor'). All differences are significant at $p < .001$. Error bars are $\pm 1SEM$

$KW p < .001$) Report rates compared to the control group, reflecting an increase in anti-social behavior. They also had significantly lower *sent* ($F(1,4394) = 34.517, p < .001, \epsilon^2 = .0081, r = .09, KW p < .001$) and *received* ($F(1,4394) = 23.11, p < .001, \epsilon^2 = .0049, r = .07, KW p < .001$) Honor rates, indicating a reduction in altruistic or prosocial behavior. These differences are illustrated in Figure 3.

Overall, our control group sent and received positive 'Honor' at a rate that was 25% higher than that of their antisocially-named peers. Similarly, antisocial-named players sent and received negative 'Reports' at a 25% higher rate than controls.

3.2. Age

We found a significant relationship between age and online interaction rates. First of all, we note that overall, all games generate around 8 times as many positive interactions as negative ones despite the fact that there are slightly more categories for negative compared to positive interactions. There was also no difference between the average number of games played at different ages.

However, we found that overall interaction rates (interactions per game) increase with age ($p < .001$) so that players at the highest end of the age range considered (26 years old) send, on average, 20% more interactions per game than those at the lowest end (11 years old). This overall increase in interaction is composed of opposing and statistically significant changes in the rates of the four interaction types: [positive (honor) or negative (reports)] x [incoming (initiated by other team members) or outgoing [initiated by the player themselves)].

The rate of all negative interactions decrease with age. Older players (22-26 years old) are significantly less likely to send or receive negative reports compared to younger players (11-15 years old). Conversely, the rate of positive interactions increases with age. These effects were highly significant ($p < .001$) in all cases (Figure 4a).

(FIGURE 4 HERE)

Figure 4. The valence of interaction rates changes with age. a) Bootstrapped ratios of honor or report rates in older (22-26 year old) group compared to younger (11-15 year old) group as estimated from username data. Negative reports become less common (ratio old:young < 1) in the older group while positive 'honor' interactions become more common (ratio old:young > 1). On average, older players send approximately 6% more positive interactions and receive approximately 2% more positive interactions. b) The ratio of positive to negative interactions increases approximately linearly as a function of age.

As in the case of ANT data, the effect sizes of these age-related changes are small ($R^2 < .001$).

As an example, the older players (22-26 years old) sent, on average, only 6% more positive interactions than younger players (11-15 years old). In comparison to the effect sizes seen in

our analysis of username data, our ability to predict the behavior of any individual player based on their estimated age is almost non-existent and the strong significance values we find are the result of having a large number of subjects. Our ability to predict changes in the *overall* behaviour of a particular age group is, however, excellent. A linear regression model fitted to the average ratio of positive to negative interactions (the ‘valence’ of overall interactions) gives an excellent fit ($R^2 = .8$, $p < .001$) – See figure 4b. On average therefore, player behavior within LoL games experiences a slow, significant and linear increase between the ages of 11 and 26. This effect is seen equally strongly in the valence of both incoming and outgoing interactions.

4. Discussion

Although there is evidence from questionnaire-based studies that personality types are reflected to some extent in online game interactions (Worth & Book, 2014) and even in email addresses (Back, Schmukle, & Egloff, 2008), we ask here whether psychologically interesting information could be obtained purely from a large, anonymized gaming dataset. We chose to examine two game-independent attributes associated with individual players (age and antisocial tendencies) because information relating to both of these can be estimated from a single, publicly displayed data string chosen by the players themselves.

Naturally, these data are not perfect reflections of real-life player attributes. Numbers, for example, may reflect culturally significant digits rather than years. For example, ‘88’ is a culturally laden number with Chinese speakers where it represents good luck. In addition, older players may attempt to appear younger to mislead other players with regard to their expertise and younger players may attempt to appear older to gain status. Nevertheless, our comparison of two independent estimates year of birth age (Figure 2b) suggests a strong

correlation between ages extracted from usernames and those provided as part of the registration procedure.

It is possible that players choose user names that reflect a personality that they choose to adopt within the game rather than one that matches their own real-world personality. We find this plausible to some extent (video gamers are, after all, playful) but the extreme nature of some of the obscene usernames makes it unlikely that they are chosen by pro-social individuals even as a form of escapism. The age results are particularly encouraging in this respect as they correlate well with registration data and we believe that players are less likely to systematically choose alternative numerical data codes to propagate an alternative online personality although we are aware that certain numbers can be used to advertise an affiliation with extreme political beliefs (for example the number 88 also has significance within the culture of far-right Nazi sympathizers).

4.1. Antisocial nicknames

Although the actual usernames cannot be reported here for reasons of privacy, they lie well outside the adult societal norms and there can be little doubt that they are specifically designed to shock or draw attention from other players. Although we have no other psychological information about the subjects who choose these names, it is plausible that they indicate real-life antisocial or attention-seeking tendencies and we are currently investigating this hypothesis in ongoing lab-based experiments.

We found a set of correlations that link these potential antisocial tendencies to the rate and valence of player-player interactions but correlation does not inform about causality. It is tempting to associate report and honor rates with performance and behaviour within the game (since this is the overt purpose of these metrics). By this account, antisocial naming tendencies are associated with antisocial gameplay leading to higher received report rates and lower received honor rates. But equally, it is possible that players with antisocial names

receive negative reports solely because those names antagonize other players. In this context, we believe that the 'sent' metrics are particularly interesting because the ANT players themselves initiate these interactions. We found that players with ANT criticise their teammates more and praise them less than controls. In this case, the ANT names are unlikely to *cause* the negative valence of the interactions. Rather, both interaction metrics seem to reflect the underlying personalities of the players.

One intriguing possibility is that antisocial names are used to express affiliation to a particular group (or to differentiate players from their teammates). In this sense, the increased negativity associated with ANT players may be framed in terms of in-group and out-group behaviour with non-ANT players being more ready to punish and less ready to reward ANT players and *vice-versa*.

A final possibility is that the variables we examine are related through a third 'hidden' factor. For example, we considered the possibility that ANT players tend to perform worse than controls for other reasons and that their antisocial in-game behaviour was a result of this poor performance. Support for this hypothesis comes from recent studies showing that in-game antisocial behaviour is related to losing games (players who lose games are more likely to trade negative reports with their team mates) (Breuer, Scharnow, & Quandt, 2015). A full analysis of this type is beyond the scope of the current paper but we did examine this possibility in general by comparing the Match Making Rank (MMR) scores of ANT and control players: a proxy for player success. We found a very small, (but statistically significant: $p < .001$) *increase* in ANT MMRs compared to controls suggesting that increased failure levels per se were unlikely to account for the reduction in interaction valence that we measured for ANT players.

4.2. Age

We found significant changes in all our interaction metrics as a function of age. In summary, players become more pro-social as they age: negative interactions decrease and positive interactions increase. The effect is small at the individual level but extremely robust and significant at the group level. Adolescence is a period characterised by significant changes in important brain structures (amygdala, frontal lobes) that govern decision making (Galvan et al., 2006; Giedd et al., 1999). The faster maturation of the limbic system, when compared to that of the frontal lobe structures, may make adolescents more prone to react to emotionally salient situations/stimuli even when their logical reasoning is intact (Casey, Jones, & Hare, 2008; Gardner & Steinberg, 2005; Steinberg, 2004; Steinberg et al., 2009) thus driving the overall higher level of negative interactions in younger players.

Again, it is possible that the names (which embed the age data), rather than the behaviours are causal: older players may bully younger players during game play, thereby leading them to resort more to negative reporting as a retaliation strategy. Very young players may have played fewer games than older players and therefore be unaware of the societal norms within the game or become frustrated by playing against more expert opponents.

These factors are unlikely to explain the trends we observed. The data we examine here consist only of players with accounts opened in a relatively short time window between November 1st 2012 and June 13th 2013. All players therefore had approximately equal experience with the game and there was no effect of age on the number of games played. Very young players are in the minority - only 14% are less than 14 years old for example and the correlation between negativity and age is weak at the low age range, becoming stronger within the 14-27 age group. This supports the hypothesis that negative interaction rates reflect age-dependent cognitive changes in the players rather than a reaction to out-group discrimination based on their apparent age.

The increased ratio of negative to positive interactions in younger players may be due to the reduced cognitive control present in this age group. For example, (Dreyfuss et al., 2014) found increased sensitivity to threatening stimuli in adolescents, especially males which are the main demographic of LoL, even when they were instructed to ignore them. Thus it is possible that adolescents are unable to inhibit possible threatening stimuli leading to communicational escalations. The stimuli could relate to in-game events (for example getting killed), or to social interactions (for example, being criticised by another player).

The change in interaction valence could also be attributed to increases in Agreeableness/Benevolence with age; a trait related to cooperation as well as to the attribution of hostile intent to other agents' actions. Young people are more prone to misjudge a neutral message as a hostile one ((Digman, 1997; Klimstra, Hale, Raaijmakers, Branje, & Meeus, 2009; Van den Akker, Deković, Asscher, & Prinzie, 2014) and a similar pattern has been observed in studies looking at both proactive and reactive aggression in young adolescents (Fite, Colder, Lochman, & Wells, 2008; van Bokhoven et al., 2006). Thus, in the context of a highly demanding competitive match, an otherwise neutral chat message could be misconstrued as offensive leading to increased reporting.

Age-dependent changes in interaction valence may also be driven by changes in cognition as well as personality (Blakemore, 2008). For example, according to (Dumontheil, Apperly, & Blakemore, 2010) adolescents commit more errors in a Theory of Mind task (ToM), when compared to adults while other studies have shown that tasks requiring ToM activate brain networks similar to those involved in empathy and forgiving (Farrow & Woodruff, 2005); (Hayashi et al., 2010; Strang, Utikal, Fischbacher, Weber, & Falk, 2014). Dumontheil and colleagues (Dumontheil et al., 2010) concluded that the interaction between ToM and executive functions is still developing in late adolescence and we hypothesize that this is a factor in the slow increase in the valence of the interactions that we observe over age because

younger players are unable to contextualize the actions of others correctly and may misattribute actions (such as accidental poor play) to a deliberate threat or collusion ('Intentional Feeding').

There is some anatomical basis for the changes in impulsivity and risk taking seen in adolescence. A dominant theory is that the developmental trajectories of subcortical structures involved in reward (for example, the nucleus accumbens) are faster than those of more frontal cortical regions providing inhibition and cognitive control (Casey et al., 2008; Dreyfuss et al., 2014; Galvan et al., 2006). Again, this hypothesis predicts a slow but steady increase in pro-social behaviour and a decrease in impulsivity across the time frame covered by our data.

In the context of cognitive development, adolescent deficits in ToM might also be enhanced because they are deprived of valuable information such as facial and vocal cues which are an important source of information about other players' motives and emotions (Achim, Guitton, Jackson, Boutin, & Monetta, 2013).

Finally, an alternative possibility is that cognitive and behavioural differences are inherent to different birth cohorts rather than different ages per se. In other words, the increase in in-game antisocial behaviour that we observe in younger players will remain constant as those players become older: The millennials are simply more antisocial than those born before the turn of the century. Evidence for this hypothesis is mixed – largely because of the difficulty in performing well-controlled personality experiments spanning multiple generations. Recent work by Twenge et al (Twenge, Konrath, Foster, Campbell, & Bushman, 2008)(Twenge & Foster, 2010) suggests that millennials score higher on at least one antisocial personality trait (Narcissism) than age-matched cohorts from previous generations but this result has been disputed on methodological grounds and other researchers studying similar datasets indicate

that any effect that may be present is very small and that a measure strongly related to narcissism (“self-enhancement”) is stable across birth cohorts. At the moment, therefore, we believe that the most parsimonious explanation for our data is based on a developmental change in personality across adolescence rather than a systematic difference in pre- and post-millennial birth cohorts.

4.3. Conclusions

Our data show that video games can provide a wealth of useful population-level information on developmental cognitive and psychological processes. Although the individual data points may be noisy, the overall conclusions are highly robust due to the sheer number of subjects. Similar analysis techniques have been used to examine the relationship between practise and performance in a custom-built online game as well as in MMORPGs (Stafford & Dewar, 2014) but we believe we are the first to examine player-player interactions in a-MOBA game using this methodology (Drachen, Sifa, & Thurau, 2014; Guitton, 2010).

It is intriguing to ask if other clinical psychiatric disorders such as autism, sociopathy or addictive personality traits might be evident in these types of data. For example, since personality influences responses in experiments probing economic choice (Berg, Lilienfeld, & Waldman, 2013), can the same results be observed in video-games? Campbell and his colleagues supported the notion that in a classic “tragedy of the commons” game, where the individual needs to exert self-discipline and harvest a limited amount of the resources in order to allow for the continuous survival of all the players, the optimal strategy at a group level requires players to delay reward (Campbell, Bush, Brunell, & Shelton, 2005). Here, we expect players who have limited abilities to discount immediate gratification to have a stereotypical profile in complex online games such as LoL, which may alter long-term, in-game success rates both for themselves and for other team members. Conversely, it is also

possible that positive in-game behaviour such as rapid learning, team building or leadership might correlate both with positive usernames and with positive personality traits in the real world.

Finally, we have assumed here that real-world personality attributes are the *cause* of the online behavior patterns we observe. But it is possible that strategies learnt in the online environment may also provide cues to appropriate (or successful) behavior in the real world.

Video game training alters a wide range of visual, cognitive and attentional mechanisms (Adachi & Willoughby, 2013; Appelbaum, Cain, Darling, & Mitroff, 2013; Boot, Kramer, Simons, Fabiani, & Gratton, 2008; Granic, Lobel, & Engels, 2014; Green & Bavelier, 2003; Li, Polat, Makous, & Bavelier, 2009) and regimes emphasizing different strategies within the same game can lead to changes in real-world behavior (Greitemeyer & Osswald, 2010; Yoon & Vargas, 2014) and cortical activation patterns in subsequent test periods (Lee et al., 2012). It has been suggested that the remarkable plasticity evidenced in such studies is due in part to the highly arousing nature of the games themselves (Bavelier, Levi, Li, Dan, & Hensch, 2010). We are currently investigating the possibility that reinforcing altruistic strategies within a game environment condition players to modify antisocial behavior in their day-to-day life.

References

- Achim, A. M., Guitton, M., Jackson, P. L., Boutin, A., & Monetta, L. (2013). On what ground do we mentalize? Characteristics of current tasks and sources of information that contribute to mentalizing judgments. *Psychological Assessment, 25*(1), 117–126. <http://doi.org/10.1037/a0029137>
- Adachi, P. J. C., & Willoughby, T. (2013). More than just fun and games: the longitudinal relationships between strategic video games, self-reported problem solving skills, and academic grades. *Journal of Youth and Adolescence, 42*(7), 1041–1052. <http://doi.org/10.1007/s10964-013-9913-9>
- Appelbaum, L. G., Cain, M. S., Darling, E. F., & Mitroff, S. R. (2013). Action video game playing is associated with improved visual sensitivity, but not alterations in visual sensory memory. *Attention, Perception, & Psychophysics, 75*(6), 1161–1167. Retrieved from <http://link.springer.com/article/10.3758/s13414-013-0472-7>
- Back, M. D., Schmukle, S. C., & Egloff, B. (2008). How extraverted is honey.bunny77@hotmail.de? Inferring personality from e-mail addresses. *Journal of Research in Personality, 42*(4), 1116–1122. <http://doi.org/10.1016/j.jrp.2008.02.001>
- Banerjee, S., & Iglewicz, B. (2007). A Simple Univariate Outlier Identification Procedure Designed for Large Samples. *Communications in Statistics - Simulation and Computation, 36*(2), 249–263. <http://doi.org/10.1080/03610910601161264>
- Bavelier, D., Levi, D. M., Li, R. W., Dan, Y., & Hensch, T. K. (2010). Removing brakes on adult brain plasticity: from molecular to behavioral interventions. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 30*(45), 14964–14971. <http://doi.org/10.1523/JNEUROSCI.4812-10.2010>
- Berg, J. M., Lilienfeld, S. O., & Waldman, I. D. (2013). Bargaining with the devil: Using economic decision-making tasks to examine the heterogeneity of psychopathic traits. *Journal of Research in Personality, 47*(5), 472–482. <http://doi.org/10.1016/j.jrp.2013.04.003>
- Blackburn, J., & Kwak, H. (2014). STFU NOOB!: predicting crowdsourced decisions on toxic behavior in online games (pp. 877–888). ACM Press. <http://doi.org/10.1145/2566486.2567987>
- Blake, J. (2013, July 15). Video gamers given “athlete” status. *BBC*. Retrieved from <http://www.bbc.co.uk/newsbeat/23321595>
- Blakemore, S.-J. (2008). The social brain in adolescence. *Nature Reviews Neuroscience, 9*(4), 267–277. <http://doi.org/10.1038/nrn2353>
- Blashki, K., & Nichol, S. (2005). Game geek’s goss: linguistic creativity in young males within an online university forum (94/\3933k’5 9055oneone). *Australian Journal of Emerging Technologies and Society, 3*(2), 71–80.
- Boot, W. R., Kramer, A. F., Simons, D. J., Fabiani, M., & Gratton, G. (2008). The effects of video game playing on attention, memory, and executive control. *Acta Psychologica, 129*(3), 387–398. <http://doi.org/10.1016/j.actpsy.2008.09.005>
- Breuer, J., Scharrow, M., & Quandt, T. (2015). Sore losers? A reexamination of the frustration–aggression hypothesis for colocated video game play. *Psychology of Popular Media Culture, 4*(2), 126–137. <http://doi.org/10.1037/ppm0000020>

- Campbell, W. K., Bush, C. P., Brunell, A. B., & Shelton, J. (2005). Understanding the Social Costs of Narcissism: The Case of the Tragedy of the Commons. *Personality and Social Psychology Bulletin*, *31*(10), 1358–1368. <http://doi.org/10.1177/0146167205274855>
- Casey, B. J., Jones, R. M., & Hare, T. A. (2008). The Adolescent Brain. *Annals of the New York Academy of Sciences*, *1124*, 111–126. <http://doi.org/10.1196/annals.1440.010>
- Chory, R. M., & Goodboy, A. K. (2011). Is basic personality related to violent and non-violent video game play and preferences? *Cyberpsychology, Behavior and Social Networking*, *14*(4), 191–198. <http://doi.org/10.1089/cyber.2010.0076>
- DeWall, C. N., Buffardi, L. E., Bonser, I., & Keith Campbell, W. (2011). Narcissism and implicit attention seeking: Evidence from linguistic analyses of social networking and online presentation. *Personality and Individual Differences*, *51*(1), 57–62. <http://doi.org/10.1016/j.paid.2011.03.011>
- Digman, J. M. (1997). Higher-order factors of the Big Five. *Journal of Personality and Social Psychology*, *73*(6), 1246–1256.
- Drachen, A., Sifa, R., & Thurau, C. (2014). The name in the game: Patterns in character names and gamer tags. *Entertainment Computing*, *5*(1), 21–32. <http://doi.org/10.1016/j.entcom.2014.02.001>
- Dreyfuss, M., Caudle, K., Drysdale, A. T., Johnston, N. E., Cohen, A. O., Somerville, L. H., ... Casey, B. J. (2014). Teens impulsively react rather than retreat from threat. *Developmental Neuroscience*, *36*(3-4), 220–227. <http://doi.org/10.1159/000357755>
- Dumontheil, I., Apperly, I. A., & Blakemore, S.-J. (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental Science*, *13*(2), 331–338. <http://doi.org/10.1111/j.1467-7687.2009.00888.x>
- Farrow, T. F. ., & Woodruff, P. W. . (2005). Neuroimaging of Forgivability. In *Handbook of Forgiveness* (pp. 259–273).
- Fite, P. J., Colder, C. R., Lochman, J. E., & Wells, K. C. (2008). Developmental trajectories of proactive and reactive aggression from fifth to ninth grade. *Journal of Clinical Child and Adolescent Psychology: The Official Journal for the Society of Clinical Child and Adolescent Psychology, American Psychological Association, Division 53*, *37*(2), 412–421. <http://doi.org/10.1080/15374410801955920>
- Galvan, A., Hare, T. A., Parra, C. E., Penn, J., Voss, H., Glover, G., & Casey, B. J. (2006). Earlier Development of the Accumbens Relative to Orbitofrontal Cortex Might Underlie Risk-Taking Behavior in Adolescents. *The Journal of Neuroscience*, *26*(25), 6885–6892. <http://doi.org/10.1523/JNEUROSCI.1062-06.2006>
- Gardner, M., & Steinberg, L. (2005). Peer influence on risk taking, risk preference, and risky decision making in adolescence and adulthood: an experimental study. *Developmental Psychology*, *41*(4), 625–635. <http://doi.org/10.1037/0012-1649.41.4.625>
- Gaudiosi, J. (2012). Riot Games' League Of Legends Officially Becomes Most Played PC Game In The World. Retrieved from <http://www.forbes.com/sites/johngaudiosi/2012/07/11/riot-games-league-of-legends-officially-becomes-most-played-pc-game-in-the-world/>
- Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., ... Rapoport, J. L. (1999). Brain development during childhood and adolescence: a longitudinal MRI study. *Nature Neuroscience*, *2*(10), 861–863. <http://doi.org/10.1038/13158>

- Granic, I., Lobel, A., & Engels, R. C. M. E. (2014). The benefits of playing video games. *The American Psychologist*, *69*(1), 66–78. <http://doi.org/10.1037/a0034857>
- Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature*, *423*(6939), 534–537. <http://doi.org/10.1038/nature01647>
- Greitemeyer, T., & Osswald, S. (2010). Effects of prosocial video games on prosocial behavior. *Journal of Personality and Social Psychology*, *98*(2), 211–221. <http://doi.org/10.1037/a0016997>
- Guitton, M. J. (2010). Cross-modal compensation between name and visual aspect in socially active avatars. *Computers in Human Behavior*, *26*(6), 1772–1776. <http://doi.org/10.1016/j.chb.2010.07.004>
- Hayashi, A., Abe, N., Ueno, A., Shigemune, Y., Mori, E., Tashiro, M., & Fujii, T. (2010). Neural correlates of forgiveness for moral transgressions involving deception. *Brain Research*, *1332*, 90–99. <http://doi.org/10.1016/j.brainres.2010.03.045>
- Hoaglin, D. C., Iglewicz, B., & Tukey, J. W. (1986). Performance of Some Resistant Rules for Outlier Labeling. *Journal of the American Statistical Association*, *81*(396), 991–999. <http://doi.org/10.1080/01621459.1986.10478363>
- Holtzman, N. S., Vazire, S., & Mehl, M. R. (2010). Sounds like a Narcissist: Behavioral Manifestations of Narcissism in Everyday Life. *Journal of Research in Personality*, *44*(4), 478–484. <http://doi.org/10.1016/j.jrp.2010.06.001>
- King, D. L., Delfabbro, P. H., & Griffiths, M. D. (2013). Trajectories of problem video gaming among adult regular gamers: an 18-month longitudinal study. *Cyberpsychology, Behavior and Social Networking*, *16*(1), 72–76. <http://doi.org/10.1089/cyber.2012.0062>
- Klimstra, T. A., Hale, W. W., Raaijmakers, Q. A. W., Branje, S. J. T., & Meeus, W. H. J. (2009). Maturation of personality in adolescence. *Journal of Personality and Social Psychology*, *96*(4), 898–912. <http://doi.org/10.1037/a0014746>
- Kraut, R., Olson, J., Banaji, M., Bruckman, A., Cohen, J., & Couper, M. (2004). Psychological research online: report of Board of Scientific Affairs' Advisory Group on the Conduct of Research on the Internet. *American Psychologist*, *59*(2), 105.
- Kwak, H., Blackburn, J., & Han, S. (2015). Exploring Cyberbullying and Other Toxic Behavior in Team Competition Online Games (pp. 3739–3748). ACM Press. <http://doi.org/10.1145/2702123.2702529>
- League of Legends. (2013, October 3). In *Wikipedia, the free encyclopedia*. Retrieved from http://en.wikipedia.org/w/index.php?title=League_of_Legends&oldid=575467374
- Lee, H., Voss, M. W., Prakash, R. S., Boot, W. R., Vo, L. T. K., Basak, C., ... Kramer, A. F. (2012). Videogame training strategy-induced change in brain function during a complex visuomotor task. *Behavioural Brain Research*, *232*(2), 348–357. <http://doi.org/10.1016/j.bbr.2012.03.043>
- Li, R., Polat, U., Makous, W., & Bavelier, D. (2009). Enhancing the contrast sensitivity function through action video game training. *Nature Neuroscience*, *12*(5), 549–551. <http://doi.org/10.1038/nn.2296>
- Markham, A., & Buchanan, E. (2012). Ethical decision-making and Internet research (version 2.0): Recommendations from the AoIR Ethics Working Committee. *Available Online: Aoir.org/reports/ethics2.Pdf*.

- Park, J., Song, Y., & Teng, C.-I. (2011). Exploring the links between personality traits and motivations to play online games. *Cyberpsychology, Behavior and Social Networking*, *14*(12), 747–751. <http://doi.org/10.1089/cyber.2010.0502>
- Stafford, T., & Dewar, M. (2014). Tracing the Trajectory of Skill Learning With a Very Large Sample of Online Game Players. *Psychological Science*, *25*(2), 511–518. <http://doi.org/10.1177/0956797613511466>
- Steinberg, L. (2004). Risk taking in adolescence: what changes, and why? *Annals of the New York Academy of Sciences*, *1021*, 51–58. <http://doi.org/10.1196/annals.1308.005>
- Steinberg, L., Graham, S., O'Brien, L., Woolard, J., Cauffman, E., & Banich, M. (2009). Age Differences in Future Orientation and Delay Discounting. *Child Development*, *80*(1), 28–44. <http://doi.org/10.1111/j.1467-8624.2008.01244.x>
- Strang, S., Utikal, V., Fischbacher, U., Weber, B., & Falk, A. (2014). Neural Correlates of Receiving an Apology and Active Forgiveness: An fMRI Study. *PLoS ONE*, *9*(2), e87654. <http://doi.org/10.1371/journal.pone.0087654>
- Teng, C.-I. (2008). Personality differences between online game players and nonplayers in a student sample. *Cyberpsychology & Behavior*, *11*(2), 232–234. <http://doi.org/10.1089/cpb.2007.0064>
- Twenge, J. M., & Foster, J. D. (2010). Birth cohort increases in narcissistic personality traits among American college students, 1982–2009. *Social Psychological and Personality Science*, *1*(1), 99–106. Retrieved from <http://spp.sagepub.com/content/1/1/99.short>
- Twenge, J. M., Konrath, S., Foster, J. D., Campbell, W. K., & Bushman, B. J. (2008). Egos inflating over time: a cross-temporal meta-analysis of the Narcissistic Personality Inventory. *Journal of Personality*, *76*(4), 875–902; discussion 903–928. <http://doi.org/10.1111/j.1467-6494.2008.00507.x>
- van Bokhoven, I., van Goozen, S. H. M., van Engeland, H., Schaal, B., Arseneault, L., Séguin, J. R., ... Tremblay, R. E. (2006). Salivary testosterone and aggression, delinquency, and social dominance in a population-based longitudinal study of adolescent males. *Hormones and Behavior*, *50*(1), 118–125. <http://doi.org/10.1016/j.yhbeh.2006.02.002>
- Van den Akker, A. L., Deković, M., Asscher, J., & Prinzie, P. (2014). Mean-level personality development across childhood and adolescence: A temporary defiance of the maturity principle and bidirectional associations with parenting. *Journal of Personality and Social Psychology*, *107*(4), 736–750. <http://doi.org/10.1037/a0037248>
- Worth, N. C., & Book, A. S. (2014). Personality and behavior in a massively multiplayer online role-playing game. *Computers in Human Behavior*, *38*, 322–330. <http://doi.org/10.1016/j.chb.2014.06.009>
- Yee, N., Ducheneaut, N., Nelson, L., & Likarish, P. (2011). Introverted elves & conscientious gnomes: the expression of personality in world of warcraft. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 753–762). ACM. Retrieved from <http://dl.acm.org/citation.cfm?id=1979052>
- Yoon, G., & Vargas, P. T. (2014). Know Thy Avatar The Unintended Effect of Virtual-Self Representation on Behavior. *Psychological Science*, *0956797613519271*. <http://doi.org/10.1177/0956797613519271>

Figures

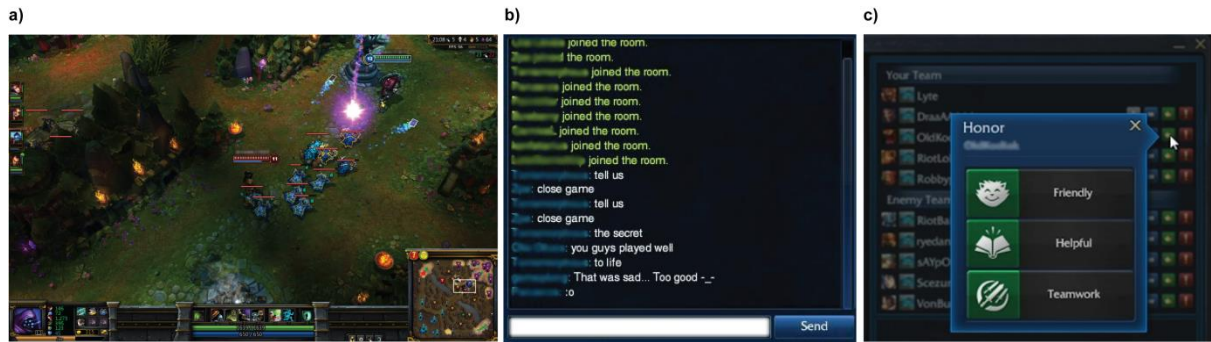


Figure 1.

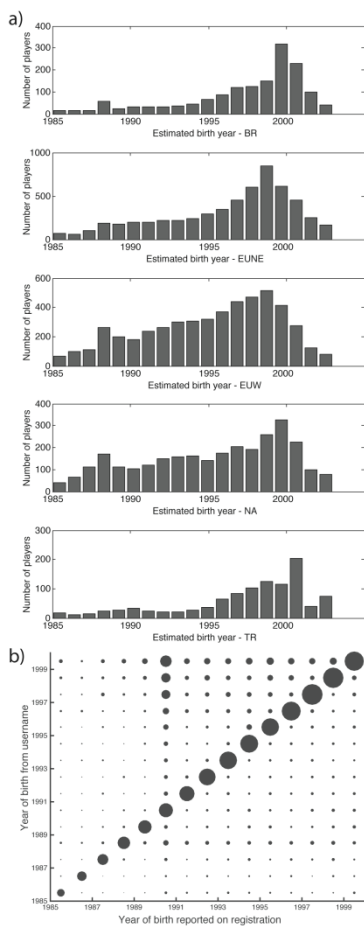


Figure 2

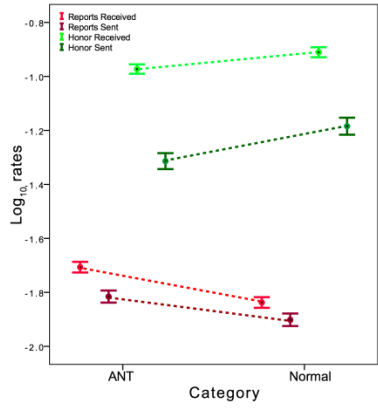


Figure 3.

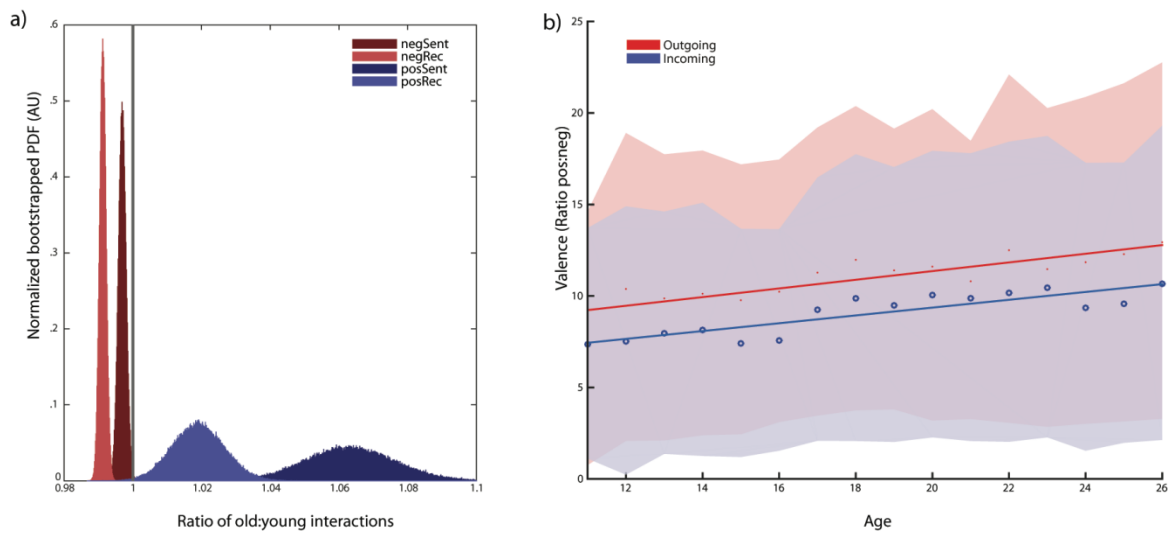


Figure 4.

Tables

Descriptive statistics from ANT and random non-ANT players.

		N	Mean	Std. Dev	Std. Err	95% Conf Interval		Min	Max
						Lower	Upper		
Log(Rep Received)	ANT	2198	-1.707	.473	.010	-1.726	-1.687	-3.371	.572
	Random	2198	-1.838	.472	.010	-1.857	-1.818	-3.083	.166
Log(Reports Sent)	ANT	2198	-1.816	.537	.011	-1.838	-1.793	-3.373	.286
	Random	2198	-1.902	.555	.012	-1.925	-1.878	-3.729	.079
Log(Honor Received)	ANT	2198	-.973	.417	.009	-.990	-.955	-2.538	1.314
	Random	2198	-.910	.441	.009	-.929	-.892	-2.093	1.623
Log(Honor Sent)	ANT	2198	-1.314	.706	.015	-1.343	-1.284	-3.560	1.729
	Random	2198	-1.184	.752	.016	-1.216	-1.153	-3.604	1.975

Appendix C

Exploring the relationship between video game expertise and fluid

Intelligence

A.V. Kokkinakis^{1,2}, P. Cowling², A., Drachen², A.R. Wade^{1*}

1: Department of Psychology, University of York, UK

2: Department of Computer Science, University of York, UK

*Corresponding author

Email: alex.wade@york.ac.uk

Abstract

Hundreds of millions of people play intellectually-demanding video games every day. What does individual performance on these games tell us about cognition? Here, we describe two studies that examine the potential link between intelligence and performance in one of the most popular video games genres in the world (Multiplayer Online Battle Arenas: MOBAs). In the first study, we show that performance in the popular MOBA League of Legends' correlates strongly with fluid intelligence as measured under controlled laboratory conditions. In the second study, we also show that the age profile of performance in the two most widely-played MOBAs (League of Legends and DOTA II) matches that of raw fluid intelligence. We discuss and extend previous videogame literature on intelligence and videogames and suggest that commercial video games can be useful as 'proxy' tests of cognitive performance at a global population level.

Introduction

Games of strategy, such as chess or mancala, can be found across cultures and skilled performance in these games has been associated with intelligence [1–4] historically. Spitz formalized this connection with specific subpopulations, pointing out that performance in a wide variety of strategy games such as *Tic-Tac-Toe* or the *Towers of Hanoi* can be linked to mental ability [3,5]. He went on to suggest that strategy games tap a number of facets of intelligence: visualization of possible moves, short-term memory rehearsal and the ability delay immediate gratification to increase future rewards (for example, sacrificing a piece in chess in order to win the game in a later turn) [1]. Later studies consolidated the link between intelligence and game performance. For example, expert chess players have above average intelligence and that the correlation between skill level as approximated by rank and IQ scores explains approximately 30% of the variance [6–9].

This notion was extended to the domain of video games by Rabbitt et al. [10] who correlated scores from the Alice-Heim (AH-4) IQ test with performance in '*Space Fortress*'; an arcade-like single player game developed by psychologists [11–13]. While individual player IQs did not predict initial performance in *Space Fortress*, they *did* predict learning rates and, therefore, performance once players had engaged with the game long enough to become practised. More recent studies have suggested that IQ can be measured in a subset of simple single-player video games [14,15] as well as through tasks embedded in game-like environments [16]. In our current paper we extend their

findings by asking whether we can establish a link between intelligence and performance in widely-played, commercial, team-based videogames with global reach.

More specifically, we focus on performance in a category of videogame that is played by millions of people: 'Multiplayer Online Battle Arenas' (MOBAs). MOBAs are action strategy games that typically involve two opposing teams of five individuals. Each individual controls one unit in a bounded map and the objective is to destroy the opponents' base [17]. In comparison to the relatively specialized games analysed in some previous studies, MOBAs are, by some measures, the most popular games on the planet with an aggregate of at least 100 million registered active players. Findings based on these games are therefore important because they have relevance to the lives of a significant fraction of the global population. Their complexity also makes them intriguing targets for scientific investigation. While many previous studies have examined the cognitive correlates of playing 'First Person Shooters' (FPS), MOBAs have a reduced emphasis on hand-eye coordination but a far stronger dependence on memory, tactics and strategy which may, in turn, tap cognitive resources more closely linked to fluid intelligence.

Here we perform two separate studies performed using two independent video game datasets both of which address the relationship between intelligence and MOBA video game performance.

Study 1

In **Study 1** we ask whether we can measure a correlation between MOBA video game performance and intelligence directly. We describe an experiment in which we measure psychometric factors related to intelligence in individual players under laboratory conditions and correlate these factors with players' ranks in the popular commercial MOBA '*League of Legends*' (LoL) [18]. Specifically, we ask whether a common measure of Fluid Intelligence (scores on the WASI II Matrix test) correlates with LoL rank. Because working memory and fluid intelligence are highly related [19, 20], we also tested players on a battery of WM tasks to ask whether WM itself was the key driver in performance.

One confound in our IQ/rank results could be players' ability to work socially with other members of their team or to impute the motives of the opposing team members. Poor theory of mind (TOM) processing could therefore affect performance directly. In addition, TOM scores have also been shown to correlate positively with IQ [21] and so any correlation between performance and IQ could potentially be explained by TOM. To control for this, we also presented our subjects with a test that measures aspects of TOM: The 'Reading the Mind in the Eyes Test' (MITE) [21, 22]. We performed correlation and partial correlation analyses to determine whether scores on this test can explain the relationships we find between WASI II scores and performance.

Study 2

Many studies have examined the relationship between age and psychometric measures of intelligence [23, 24]. Raw fluid intelligence scores tend to peak in the early to mid 20s followed by a steady decline [24–27]. We would expect this profile to be mirrored in the performance metrics of video games that correlate highly with IQ.

In **Study 2** we extended the finding of Study 1 to ask if potential effects of Raw IQ can be detected in large sets of data relating to video game performance. In particular, we ask whether performance in MOBAs follows the age profile that would be predicted if it correlated with raw Fluid Intelligence. We analyse data from two MOBAs ('*League of Legends*' and '*Defense of the Ancients 2*' (Dota 2) [28]

which have more than 100 million registered unique players between them. We specifically asked whether performance in our these two MOBA datasets followed a trajectory that peaks in the early to mid-20s.

To control for age-related factors that may not depend on IQ, we compared these games with another popular genre: 'First Person Shooters' (FPS) in which players control a character from the 'first person' perspective and engage in combat within a simulated 3D world. Our comparison data come from two popular exemplars of the FPS genre: *Destiny* [29] and *Battlefield 3* [30] which have peaked at an aggregate of over 25 million registered players. We specifically use FPS games as a comparison because they appear to prioritise speed and targeting accuracy over memory and multifactorial decision making [31–33] and may therefore reflect a different set of cognitive performance characteristics (in particular, reaction times) that peak at earlier ages.

Materials and Methods

Study 1: Fluid Intelligence and associated measures

Ethics

All participants in our laboratory experiments provided informed consent and approval for the study was provided by the ethics board of the Psychology Department of the University of York. All data were anonymized and participants were informed that they could withdraw from the study at any time.

Participants

Participants (N=56, 51 males, mean age 20.5 years) were recruited via adverts from multiple sites within the UK in and around the Universities of Leeds, Essex and York. All subjects were experienced LoL players who had played a large number (>100) of both 'ranked' and 'unranked' matches. For more specific criteria and the advertisement please see the supplementary material.

Instruments

In this analysis we obtained psychometric test scores from subjects under laboratory conditions. We then compared those score with performance as measured by the subjects' League of Legends rankings.

We used the WASI-II [34] Matrix Subtest as a standardized measure of fluid intelligence along with three working memory tasks (Symmetry, Rotation and the Operation Span task) that have been validated extensively [35–40].

Working Memory is closely related to fluid intelligence which we included as a complementary way of measuring cognitive ability [19,20,41,42]. We performed correlation analysis with a set of working memory measures to assess its relationship to the video game rank data. We also asked if a single underlying latent variable constructed from scores on the WASI and the three working memory tests provided a more parsimonious explanation for our results. Using IBM SPSS Amos (version 24, IBM Corp, NY) we ran a confirmatory factor analysis (CFA) and confirmed that a valid single factor could be constructed (see Supplementary Material). However the correlation with our Rank scores was weaker (although still significant) and for simplicity we choose to present correlations with individual test scores here.

We also included the Mind in the Eyes Test (MITE) [22] which is designed to probe subject's understanding of other people's emotional states and which has been shown to correlate with intelligence [21]. In the MITE participants had to identify the emotion a face conveyed just from their

eyes without hints from the rest of the face. Test administration was counterbalanced to eliminate order/presentation effects [43].

Rank

Online videogames such as the ones examined here provide detailed telemetry to the coordinating game servers. Companies such as Riot therefore have databases of real-time information about the behaviour and performance of game players.

We asked participants to provide their online nicknames so that we could access their game history and rank through a publicly-accessible website that interfaces directly to the Riot Games API database (<https://euw.op.gg/>).

Each player's video game rank (computed from their position in an ELO-like ranking system similar to that used by the United States Chess Federation) was extracted from an online database [44]. In LoL players are divided into ranked 'tiers' with each tier having five 'divisions'. A player's position in these divisions and ranks depends solely on the ratio of matches won and lost over time [45] and not the performance within each match. Our participants' rank ranged from 'Silver Division 5' up to the 'Masters Division' (see Supplementary Material for more details).

Study 2 – Age and performance

Ethics

We used existing data sources to ask whether performance in different types of games followed the age-dependent trajectory that would be expected if it was highly correlated with fluid intelligence. No player-identifying information was present in any of the datasets and data acquisition procedures were approved by a separate application to the University of York Psychology Ethics Committee. More details on player demographics are presented in Table 1.

Data Sources

All four videogames use the ratio of historical wins to losses as a primary metric for the 'Matchmaking Ranking' (MMR) score which we analyse here. MMR is a dynamically-updated measure of player performance that depends solely on the win/loss history of each player and the rank of their opponents. It should be noted that Divisions and Tiers in League correspond to an MMR range that is hidden from the user but provided to us by Riot Games.

League of Legends: A snapshot of LoL player ranks was provided by Riot Games (Riot Games, a subsidiary of TenCent Holdings, Los Angeles, CA). Other aspects of this dataset have been analysed in a previous paper [46].

DOTA II: A dataset from casual players who spectated at the 'International 5 Dota 2 Tournament 2015' was provided by the education analysts at Foundry 10 [47–49] and Valve (Valve LLC, Bellevue, WA).

Destiny: The anonymized Destiny dataset were obtained from the developer, Bungie (Bungie, Inc. Bellevue, WA), with age data from a public online survey of approximately 1700 Destiny players who participated on a voluntary basis.

Battlefield 3: Anonymized Battlefield 3 (Electronic Arts, Redwood City, CA) data were obtained from Tekofsky and colleagues [50,51] and is available through their website (<http://www.psyopsresearch.com/download/>). In our analysis we used the data from the structure `stats.global.elo`.

Age Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
BF3	8743	13	40	23.67	6.53
Destiny	1669	13	40	24.18	6.32
Dota 2	286	13	40	23.11	4.02
LoL	17861	13	40	20.49	5.07

Table 1 Player demographics for the four games in our analyses.

Data analysis

Fluid intelligence changes with age [26,52,53]. Here, we asked whether performance in four different games had shared similarities in their aging profiles. Our hypothesis was that because MOBA performance correlates with fluid intelligence, it would follow an age trajectory similar to that seen in population-level raw IQ scores – peaking in the early to mid-twenties with a decline thereafter [25,26,54,55]. We chose the first person shooters *Destiny and BF3* as controls since performance on these game might be expected to correlate more with reaction time and therefore peak earlier in the lifespan [56]. MMR performance measures are available for each of these games but the absolute scaling of this ratio variable differs between games. To enable a direct comparison between the four games (*Dota 2, LoL, Destiny and BF3*) we Z-scored the MMR distributions for each game separately by removing the means and scaling by the standard deviations. We then separated these normalized within-game scores into three age groups designed to span the point at which raw IQ scores begin to decrease (13-21, 22-27 and 28+ years old).

Results

Study 1 – Raw fluid Intelligence scores and player Rank



Figure 3 Cross correlations between variables of interest. The leading diagonal shows the distribution of the data. Numbers above the diagonal show the non-parametric cross correlation coefficient. Scattergrams of the data with best fit lines and error limits are shown below the leading diagonal. There is a moderately-sized and highly significant correlation between WASI-II Matrices and Rank ($r_s=.44$, $p=.001$) and a weak but significant correlation between Rank and Rotation Span score with $r_s=.26$, $p<.05$. The correlations between Rank and OSPAN and MITE task scores were not significantly correlated with with $r_s=0.3$, $p=.43$ and $r_s=-.01$, $p=.242$ respectively.

Because our data were not normally distributed, we computed the non-parametric ‘Spearman ρ ’ correlation between Rank and performance on the standardized psychometric tasks (for more information about Ranking and alternative coding see Supplementary Material). We found that fluid intelligence as measured by the WASI II Matrix Reasoning Subtest, correlated significantly with rank (nonparametric rank correlation: $r_s=.44$ (95% CI [.24 .60], $p=.001$) – See Figure 1. Importantly, we found no significant correlations between rank and scores in the MITE task and the partial correlation of WASI II scores with rank controlling for MITE was not significantly different to the initial correlation without accounting for MITE. Similarly, we found only a weak correlation between rank and a tests of visuospatial working memory.

The Rotation Span scores were weakly but significantly correlated with Rank ($r_s = .26, p < .05$).

Correlations

			WASI II	League of Legends Ranking	Rotation Span Score	Symmetry Span Score	Operation Span Score	Mind in the Eyes Test Score
Spearman's rho	WASI II Raw Score	Correlation Coefficient	1.000	.440**	.303*	.247*	.071	.082
		Sig. (1-tailed)	.	.000	.014	.036	.302	.275
		N	56	56	52	54	56	55
League of Legends Ranking	League of Legends Ranking	Correlation Coefficient	.440**	1.000	.260*	.117	.025	-.096
		Sig. (1-tailed)	.000	.	.031	.199	.427	.242
		N	56	56	52	54	56	55
Rotation Span Score	Rotation Span Score	Correlation Coefficient	.303*	.260*	1.000	.440**	.198	.050
		Sig. (1-tailed)	.014	.031	.	.001	.080	.365
		N	52	52	52	50	52	51
Symmetry Span Score	Symmetry Span Score	Correlation Coefficient	.247*	.117	.440**	1.000	.197	.050
		Sig. (1-tailed)	.036	.199	.001	.	.077	.361
		N	54	54	50	54	54	53
Operation Span Score	Operation Span Score	Correlation Coefficient	.071	.025	.198	.197	1.000	-.131
		Sig. (1-tailed)	.302	.427	.080	.077	.	.171
		N	56	56	52	54	56	55
Mind in the Eyes Test Score	Mind in the Eyes Test Score	Correlation Coefficient	.082	-.096	.050	.050	-.131	1.000
		Sig. (1-tailed)	.275	.242	.365	.361	.171	.
		N	55	55	51	53	55	55

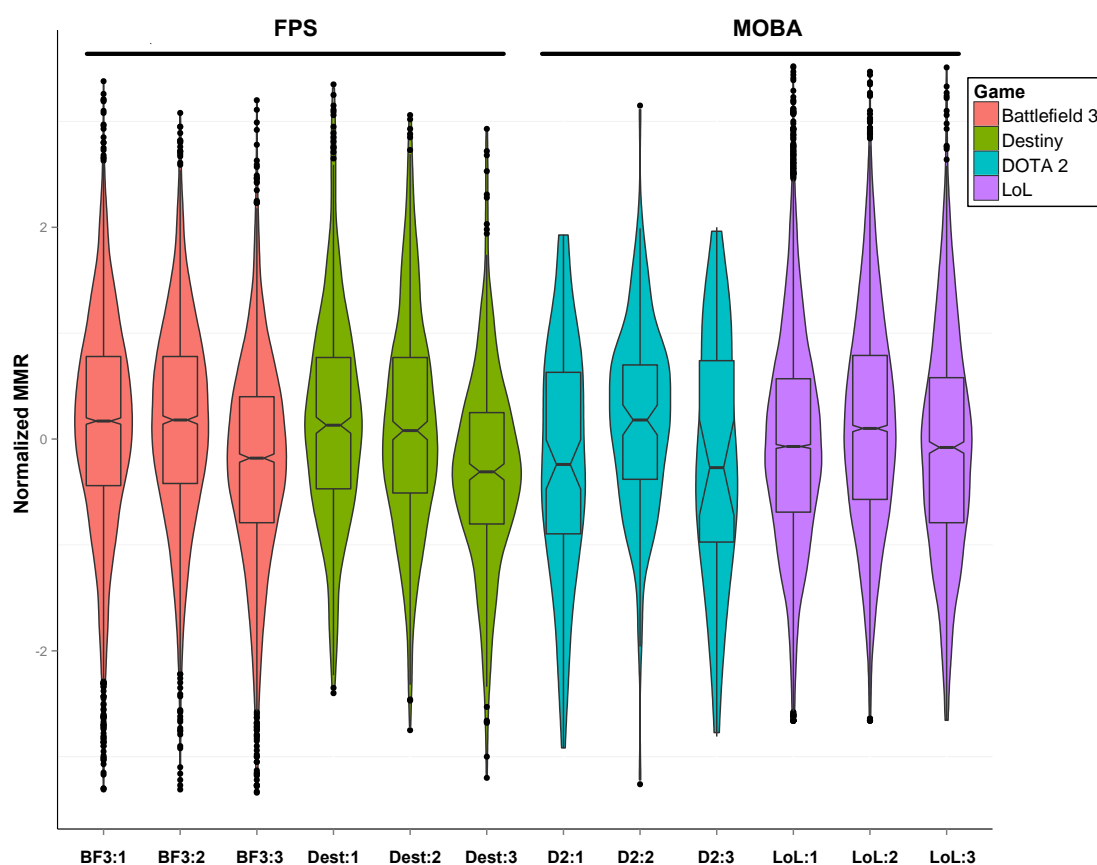
** . Correlation is significant at the 0.01 level (1-tailed).

* . Correlation is significant at the 0.05 level (1-tailed).

Outliers are always a concern in correlational analyses. To address this, we computed the Cook's Distance for all points in the MMR vs WASI II analysis. The highest Cook's distance was found for the subject with the lowest WASI II score. However, the Cook's Distance for this subject was .59 (well

below the .70 threshold for our sample) and the correlation is virtually unchanged $r_5=.435$, $p<.001$ even if this player is excluded.

Study 2 – Performance as a function of age



*Figure 4 Age profiles of MMR in four different games. Three age groups for each game are plotted: (1) 13-21, (2) 22-27 and (3) 28 years an over. In two popular ‘First Person Shooter’ games (Battlefield 3 and Destiny), performance decreases monotonically with age following a ‘high, high, low’ profile. In comparison, two of the most popular multiplayer Boxplots of the age-grouped MMR data are shown in Figure 2. Visual inspection shows that performance scores in the MOBA and FPS games follow a different age trajectory and ANOVA analysis of the data with planned comparisons confirms this observation. All games showed a significant difference ($p<.001$) between the middle group and the last group – indicating that performance in general falls off after the mid 20s. However, only the MOBAs **also** showed a significant increase ($p<.001$) between the first and second age group. This increase is consistent with the hypothesis that performance in MOBAs (but not FPS games) is strongly correlated with fluid intelligence which also exhibits this age profile. For more detailed statistics and tables see the Supplementary material.*

Overall, we found that MOBA-genre performance profiles followed a ‘low,high,low’ pattern where performance peaked in the 22-27 year old age group. In comparison, FPS performance followed a ‘high,high,low’ pattern suggesting that younger players had a relative advantage in this genre and that performance decreases monotonically with age.

Discussion

What can these results tell us about the link between commercial MOBA video games and cognition?

The literature around video games, psychology and neurophysiology (much of which focuses on FPS games) is extremely diverse (see Palau 2017 for a review [57]). Green and Bavelier's work in the early 2000s identified perceptual correlates of FPS play and later studies extended this to attentional effects and cognitive tasks such as response inhibition, task switching and working memory [58–61]. While extended FPS play may lead to improvements in visuospatial processing [62] (although see [63,64]), the same subjects may also exhibit reduced ability to process emotional stimuli [65]. Finally, extended FPS play may reduce and/or increase cortical gray matter thickness and cortical connectivity [66–69] depending, in part, on the game strategies adopted.

These observations are important but the field is largely focused on the question of whether video game practise generates cognitive or perceptual benefits that transfer to other domains. This focus is due to several factors: Clinicians and health scientists are, understandably interested in the potential that video games may hold for neurorehabilitation, educators and parents are interested in the long-term effects that video game play may have on young people and neuroscientists see the extended training and perceptual measurements that video game play affords as an opportunity to learn more about relatively mature fields such as perceptual learning.

Although our data indicate a link between intelligence and video game performance, the relationship is correlational and so the causality is unclear. One possibility is that rather than games modifying cognition, learning to play video games depend on the same cognitive resources underlying performance on intelligence tests. There is some support for this from previous research by, for example, Rabbitt *et al* (9) who showed that intelligence correlated with practised performance on “*Space Fortress*”: a simple, non-commercial video game designed by psychologists. This link is also supported by recent work showing that the rate of early learning in an online video game and the stable final performance are correlated, lending support to the idea that a single factor (presumably related to cognitive capacity) underlies both metrics [70].

Importantly, Stafford and Dewar's work [70] suggest that cognitive capacity is assayed by a final, stable performance metric which is ultimately *invariant* to increasing practice. This correlation between performance after large amounts of practise and cognitive capacity is further supported by Adams and Mayer [71] who identified a correlation between scores in a first-person shooter (*Unreal Tournament*) and two mental rotation tasks in non-videogame players (Shepard-Metzler & paper-folding). Finally, similar results have been observed by Bonny & Castaneda [48] who that number processing ability not only correlates with Dota 2 MMR but is also predictive of MMR improvement over time.

In principle, the correlation might arise because playing video games causes an increase in intelligence. Although IQ scores are believed to be relatively stable [72], training on action video games does improve visuospatial performance [73] and the general mechanism (an improvement in probabilistic inference based on visual input [74] could, potentially, translate to a wide range of cognitive tasks. Addressing this possibility robustly would require a set of large-scale longitudinal experiments and is beyond the scope of the current study.

Age

The final possibility is that some third factor is driving variance in both intelligence and video game expertise. One candidate is age. Raw (un-normalized) fluid intelligence scores usually peak in the

mid-20s [25]. This also appears to be the approximate peak of video game performance in MOBAs that depend on a mixture of memory, tactics, strategy and reaction time (Figure 2) while games that emphasise more reaction times and hand-eye coordination (for example, FPS-type games) appear to advantage younger players. However, while we did find a significant correlation between WASI II scores and age ($r=.28$, $p=.035$) our LoL data showed no correlation between expertise and age ($r=.06$, $p=.056$) and a partial correlation of expertise with WASI II accounting for age was still highly significant ($p=.001$, $r=.45$).

Age may also correlate with practise: older players may have had more time to practise any particular game (although it is also possible that older players are more restricted in the amount of free time that they can devote to game play). To examine whether pure practise effects determine rank, we examined the relationship between rank and games played in our large ($N>17000$) dataset of LoL players. After the initial learning stage during which the players attained a relatively stable rank, the magnitude of the correlation between games played and expertise (indicated by MMR) was $r=.02$. While still significant ($p<.001$) this suggests that games played explains only a small amount of the variance found between experienced players [48, 75]. This agrees with Stafford and Dewer's finding that final, stable performance levels are determined largely by the rate of learning in the initial phase rather than the total number of games played overall [70].

We recognize that although fluid intelligence seems to be one factor in obtaining a higher MMR, it does not explain all the variance – practice, dedication and learning must still confer significant advantages [6,7] – particularly during the early stages of skill acquisition. This effect is mitigated to some extent in Study 1 by the fact that all our subjects were relatively well-practised (over 100 games *excluding* casual and unranked matches) and had demonstrated a willingness to engage with the game intellectually over a long period of time.

Theory of Mind

Finally, we found no correlation between performance in LoL and the MITE test which is a probe of a subject's ability to perform theory-of-mind tasks [22]. This was unexpected: MOBAs are social games and we believe that the ability to model the motives of other players enhances performance. In addition, scores on the MITE task have been shown to correlate with IQ [21] and Engel et al. [76] showed that MITE scores predict performance in cognitively demanding tasks such as solving Sudoku puzzles. We conclude that, at least in our relatively small sample population, any weak correlation between MITE and intelligence scores that may exist may be swamped by other factors. In addition, if the relationship between performance and MITE score is driven by subjects with poor TOM, we may not have sampled an MITE dataset with sufficient variance over the low end of the range to expose the effect.

Sample size

The two studies presented here have very different sample sizes. For Study 1, $N=56$. For Study 2, we have many thousands of data points. The two analyses therefore have different strengths and weaknesses. In Study 1, although N is small, the data are collected under controlled laboratory conditions using standardized instruments. Although the relatively small sample size cautions us that this work is still exploratory, the fact that we find a correlation with such a large effect size and strong significance suggests that the results are likely to replicate across larger cohorts. The data from Study 2 *are* collected from larger cohorts but the provenance of each data point is less certain. Issues such as selection bias are potentially problematic and we expect to find some noise in measures such as age due to participants deliberately or carelessly reporting false information – although there is evidence that large web-based samples such as these can be relatively reliable

[77]. The hope with large datasets such as these is that the huge number of participants more than compensates for the increase in noise caused by the sampling methods. An important challenge for future research is to test this assumption by broadening the use of validated tests to a wider subject group – perhaps through careful use of online crowdsourcing platforms [78].

Conclusion

We propose that videogame expertise in commercial MOBAs correlates with fluid intelligence and the developmental trajectory of expertise mimics that of fluid intelligence across adolescence and early adulthood [24,25,52]. A decline in fluid intelligence and working memory has been linked with the expression of a number of diseases as well as with healthy aging [79–81]. The specific MOBA genre is remarkable in the sense that it already engages a vast number of players across the globe but more generally, complex, socially-interactive and intellectually demanding video games are now ubiquitous and generate a constant stream of performance data that can be normalized against millions of other players. If MOBAs in particular, or even video games in general offer a robust insight into cognitive function, they may be used to study cognitive epidemiology at a massive scale – instantly overcoming existing issues with small sample sizes [81] and potentially allowing us to examine dynamic changes in performance at a population level in almost real time.

Acknowledgements:

The authors extend their sincere gratitude to Dr Bonny and Lisa Castaneda at Foundry10, Prof Cairns, Dr. Fabri and Dr. John at Leeds Beckett, Myat Aung, Laura Helsby, Marisa Bostock, the University of Essex and York Fragsoc for assistance with the data collection. We are extremely grateful to both Riot Games and Valve for giving us access to their Age/MMR datasets. We would also like to thank Dr. Wallner, University of Applied Arts Vienna; Dr. Johnson, Queensland University of Technology; Dr. Kriglstein, Vienna University of Technology, Dr. Nacke, University of Waterloo and Dr. Rafet Sifa, Fraunhofer IAIS, for access to the *Destiny* datasets.

Finally, we thank Chris Ntzokas for valuable discussions and advice.

References

1. Spitz HH. The universal nature of human intelligence: Evidence from games. *Intelligence*. 1978;2: 371–379. doi:10.1016/0160-2896(78)90026-0
2. Bilalić M, McLeod P, Gobet F. Does chess need intelligence? — A study with young chess players. *Intelligence*. 2007;35: 457–470. doi:10.1016/j.intell.2006.09.005
3. Spitz HH, Winters EA. Tic-Tac-Toe performance as a function of maturational level of retarded adolescents and nonretarded children. *Intelligence*. 1977;1: 108–117. doi:10.1016/0160-2896(77)90030-7
4. Spitz HH, DeRisi DT. Porteus Maze Test performance of retarded young adults and nonretarded children. *American journal of mental deficiency*. 1978; Available: <http://psycnet.apa.org/psycinfo/1979-23910-001>
5. Byrnes MM, Spitz HH. Performance of retarded adolescents and nonretarded children on the Tower of Hanoi Problem. *Am J Ment Defic*. 1977;81: 561–569.

6. Grabner RH. The role of intelligence for performance in the prototypical expertise domain of chess. *Intelligence*. 2014;45: 26–33. doi:10.1016/j.intell.2013.07.023
7. de Bruin ABH, Kok EM, Leppink J, Camp G. Practice, intelligence, and enjoyment in novice chess players: A prospective study at the earliest stage of a chess career. *Intelligence*. 2014;45: 18–25. doi:10.1016/j.intell.2013.07.004
8. Burgoyne AP, Sala G, Gobet F, Macnamara BN, Campitelli G, Hambrick DZ. The relationship between cognitive ability and chess skill: A comprehensive meta-analysis. *Intelligence*. 2016;59: 72–83. doi:10.1016/j.intell.2016.08.002
9. Grabner RH, Stern E, Neubauer AC. Individual differences in chess expertise: a psychometric investigation. *Acta Psychol (Amst)*. 2007;124: 398–420. doi:10.1016/j.actpsy.2006.07.008
10. Rabbitt P, Banerji N, Szymanski A. Space fortress as an IQ test? Predictions of learning and of practised performance in a complex interactive video-game. *Acta Psychologica*. 1989;71: 243–257. doi:10.1016/0001-6918(89)90011-5
11. Donchin E. Video games as research tools: The Space Fortress game. *Behavior Research Methods, Instruments, & Computers*. 1995;27: 217–223. doi:10.3758/BF03204735
12. Mané A, Donchin E. The space fortress game. *Acta Psychologica*. 1989;71: 17–22. doi:10.1016/0001-6918(89)90003-6
13. Boot WR. Video games as tools to achieve insight into cognitive processes. *Front Psychol*. 2015;6. doi:10.3389/fpsyg.2015.00003
14. Quiroga MA, Escorial S, Román FJ, Morillo D, Jarabo A, Privado J, et al. Can we reliably measure the general factor of intelligence (g) through commercial video games? Yes, we can! *Intelligence*. 2015;53: 1–7. doi:10.1016/j.intell.2015.08.004
15. Quiroga MA, Herranz M, Gómez-Abad M, Kebir M, Ruiz J, Colom R. Video-games: Do they require general intelligence? *Computers & Education*. 2009;53: 414–418.
16. Foroughi CK, Serraino C, Parasuraman R, Boehm-Davis DA. Can we create a measure of fluid intelligence using Puzzle Creator within Portal 2? *Intelligence*. 2016;56: 58–64. doi:10.1016/j.intell.2016.02.011
17. Kou Y, Gui X. Playing with strangers: understanding temporary teams in league of legends. *ACM Press*; 2014. pp. 161–169. doi:10.1145/2658537.2658538
18. What is League of Legends? | League of Legends [Internet]. [cited 20 Apr 2017]. Available: <http://gameinfo.euw.leagueoflegends.com/en/game-info/get-started/what-is-lol/>
19. Gignac GE. Fluid intelligence shares closer to 60% of its variance with working memory capacity and is a better indicator of general intelligence. *Intelligence*. 2014;47: 122–133. doi:10.1016/j.intell.2014.09.004
20. Colom R, Rebollo I, Palacios A, Juan-Espinosa M, Kyllonen PC. Working memory is (almost) perfectly predicted by g. *Intelligence*. 2004;32: 277–296. doi:10.1016/j.intell.2003.12.002
21. Baker CA, Peterson E, Pulos S, Kirkland RA. Eyes and IQ: A meta-analysis of the relationship between intelligence and “Reading the Mind in the Eyes.” *Intelligence*. 2014;44: 78–92. doi:10.1016/j.intell.2014.03.001

22. Baron-Cohen S, Wheelwright S, Hill J, Raste Y, Plumb I. The “Reading the Mind in the Eyes” Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *J Child Psychol Psychiatry*. 2001;42: 241–251.
23. Warner K, Willis SL. Age difference patterns of psychometric intelligence in adulthood: Generalizability within and across ability domains. *Psychology and Aging*. 1993;8: 44–55. doi:10.1037/0882-7974.8.1.44
24. Kaufman AS, Johnson CK, Liu X. A CHC Theory-Based Analysis of Age Differences on Cognitive Abilities and Academic Skills at Ages 22 to 90 Years. *Journal of Psychoeducational Assessment*. 2008;26: 350–381. doi:10.1177/0734282908314108
25. Craik FIM, Bialystok E. Cognition through the lifespan: mechanisms of change. *Trends in Cognitive Sciences*. 2006;10: 131–138. doi:10.1016/j.tics.2006.01.007
26. Horn JL, D P, Cattell RB, D P, Sc D. Age differences in primary mental ability factors. *Journal of Gerontology*. 1966; 210–220.
27. Kaufman AS, Lichtenberger EO. Age and Intelligence across the adult Life Span. *Assessing Adolescent and Adult Intelligence*. John Wiley & Sons; 2005. pp. 127–189.
28. Dota 2 on Steam [Internet]. [cited 20 Apr 2017]. Available: <http://store.steampowered.com/app/570/>
29. *Destiny* (video game) [Internet]. Wikipedia. 2017. Available: [https://en.wikipedia.org/w/index.php?title=Destiny_\(video_game\)&oldid=774345963](https://en.wikipedia.org/w/index.php?title=Destiny_(video_game)&oldid=774345963)
30. *Battlefield 3* [Internet]. Wikipedia. 2017. Available: https://en.wikipedia.org/w/index.php?title=Battlefield_3&oldid=772726451
31. Thompson JJ, Blair MR, Henrey AJ. Over the hill at 24: persistent age-related cognitive-motor decline in reaction times in an ecologically valid video game task begins in early adulthood. *PLoS ONE*. 2014;9: e94215. doi:10.1371/journal.pone.0094215
32. Wilkinson RT, Allison S. Age and simple reaction time: decade differences for 5,325 subjects. *J Gerontol*. 1989;44: P29-35.
33. Gottsdanker R. Age and simple reaction time. *J Gerontol*. 1982;37: 342–348.
34. Wechsler D. Wechsler Abbreviated Scale of Intelligence—Second Edition (WASI-II). NCS Pearson; 2011.
35. Bonny JW, Lindberg JC, Pacampara MC. Hip Hop Dance Experience Linked to Sociocognitive Ability. *PLOS ONE*. 2017;12: e0169947. doi:10.1371/journal.pone.0169947
36. Conway ARA, Kane MJ, Bunting MF, Hambrick DZ, Wilhelm O, Engle RW. Working memory span tasks: A methodological review and user’s guide. *Psychon Bull Rev*. 2005;12: 769–786.
37. Foster JL, Shipstead Z, Harrison TL, Hicks KL, Redick TS, Engle RW. Shortened complex span tasks can reliably measure working memory capacity. *Mem Cogn*. 2014;43: 226–236. doi:10.3758/s13421-014-0461-7
38. Kane MJ, Hambrick DZ, Tuholski SW, Wilhelm O, Payne TW, Engle RW. The generality of working memory capacity: a latent-variable approach to verbal and visuospatial memory span and reasoning. *J Exp Psychol Gen*. 2004;133: 189–217. doi:10.1037/0096-3445.133.2.189

39. Schweizer S, Dalgleish T. Emotional working memory capacity in posttraumatic stress disorder (PTSD). *Behaviour Research and Therapy*. 2011;49: 498–504. doi:10.1016/j.brat.2011.05.007
40. Wongupparaj P, Kumari V, Morris RG. The relation between a multicomponent working memory and intelligence: The roles of central executive and short-term storage functions. *Intelligence*. 2015;53: 166–180. doi:10.1016/j.intell.2015.10.007
41. Kyllonen PC, Christal RE. Reasoning ability is (little more than) working-memory capacity?! *Intelligence*. 1990;14: 389–433. doi:10.1016/S0160-2896(05)80012-1
42. Colom R, Shih PC, Flores-Mendoza C, Quiroga MÁ. The real relationship between short-term memory and working memory. *Memory*. 2006;14: 804–813.
43. Sharma VK. An easy method of constructing latin square designs balanced for the immediate residual and other order effects. *Can J Statistics*. 1975;3: 119–224. doi:10.2307/3315104
44. Elo AE. *The rating of chessplayers, past and present*. Bronx, N.Y.: Ishi Press International; 2008.
45. Ranked Play FAQ. In: Riot Games Support [Internet]. [cited 28 Mar 2017]. Available: <http://support.riotgames.com/hc/en-us/articles/204010760-Ranked-Play-FAQ>
46. Kokkinakis AV, Lin J, Pavlas D, Wade AR. What’s in a name? Ages and names predict the valence of social interactions in a massive online game. *Computers in Human Behavior*. 2016;55, Part B: 605–613. doi:10.1016/j.chb.2015.09.034
47. Bonny JW, Castaneda LM, Swanson T. Using an International Gaming Tournament to Study Individual Differences in MOBA Expertise and Cognitive Skills. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM; 2016. pp. 3473–3484. doi:10.1145/2858036.2858190
48. Bonny JW, Castaneda LM. Number processing ability is connected to longitudinal changes in multiplayer online battle arena skill. *Computers in Human Behavior*. 2017;66: 377–387. doi:10.1016/j.chb.2016.10.005
49. Foundry 10 Homepage. In: Foundry10 [Internet]. 20 Jun 2016 [cited 20 Aug 2017]. Available: <http://foundry10.org/>
50. Tekofsky S, Spronck P, Goudbeek M, Plaat A, van den Herik J. Past Our Prime: A Study of Age and Play Style Development in Battlefield 3. *IEEE Transactions on Computational Intelligence and AI in Games*. 2015;7: 292–303. doi:10.1109/TCIAIG.2015.2393433
51. Tekofsky S, Spronck P, Plaat A, Van den Herik J, Broersen J. Play style: Showing your age. *2013 IEEE Conference on Computational Intelligence in Games (CIG)*. 2013. pp. 1–8. doi:10.1109/CIG.2013.6633616
52. Kaufman AS. WAIS-III IQs, Horn’s theory, and generational changes from young adulthood to old age. *Intelligence*. 2001;29: 131–167. doi:10.1016/S0160-2896(00)00046-5
53. Salthouse TA. Selective review of cognitive aging. *J Int Neuropsychol Soc*. 2010;16: 754–760. doi:10.1017/S1355617710000706

54. Kaufman AS, Lichtenberger EO. *Assessing Adolescent and Adult Intelligence*. John Wiley & Sons; 2005.
55. Salthouse TA. When does age-related cognitive decline begin? *Neurobiol Aging*. 2009;30: 507–514. doi:10.1016/j.neurobiolaging.2008.09.023
56. Fozard JL, Vercryssen M, Reynolds SL, Hancock PA, Quilter RE. Age differences and changes in reaction time: the Baltimore Longitudinal Study of Aging. *J Gerontol*. 1994;49: P179-189.
57. Palaus M, Marron EM, Viejo-Sobera R, Redolar-Ripoll D. Neural Basis of Video Gaming: A Systematic Review. *Front Hum Neurosci*. 2017;11: 248. doi:10.3389/fnhum.2017.00248
58. Cain MS, Landau AN, Shimamura AP. Action video game experience reduces the cost of switching tasks. *Atten Percept Psychophys*. 2012;74: 641–647. doi:10.3758/s13414-012-0284-1
59. Colzato LS, Wildenberg WPM van den, Zmigrod S, Hommel B. Action video gaming and cognitive control: playing first person shooter games is associated with improvement in working memory but not action inhibition. *Psychological Research*. 2013;77: 234–239. doi:10.1007/s00426-012-0415-2
60. Colzato LS, van Leeuwen PJA, van den Wildenberg WPM, Hommel B. DOOM'd to Switch: Superior Cognitive Flexibility in Players of First Person Shooter Games. *Front Psychol*. 2010;1. doi:10.3389/fpsyg.2010.00008
61. Krishnan L, Kang A, Sperling G, Srinivasan R. Neural strategies for selective attention distinguish fast-action video game players. *Brain Topogr*. 2013;26: 83–97. doi:10.1007/s10548-012-0232-3
62. Green CS, Bavelier D. Action video game modifies visual selective attention. *Nature*. 2003;423: 534–537. doi:10.1038/nature01647
63. Boot WR, Kramer AF, Simons DJ, Fabiani M, Gratton G. The effects of video game playing on attention, memory, and executive control. *Acta Psychol (Amst)*. 2008;129: 387–398. doi:10.1016/j.actpsy.2008.09.005
64. Unsworth N, Redick TS, McMillan BD, Hambrick DZ, Kane MJ, Engle RW. Is Playing Video Games Related to Cognitive Abilities? *Psychological Science*. 2015;26: 759–774. doi:10.1177/0956797615570367
65. Bailey K, West R. The effects of an action video game on visual and affective information processing. *Brain Res*. 2013;1504: 35–46. doi:10.1016/j.brainres.2013.02.019
66. Gong D, He H, Liu D, Ma W, Dong L, Luo C, et al. Enhanced functional connectivity and increased gray matter volume of insula related to action video game playing. *Sci Rep*. 2015;5: 9763. doi:10.1038/srep09763
67. West GL, Konishi K, Diarra M, Benady-Chorney J, Drisdelle BL, Dahmani L, et al. Impact of video games on plasticity of the hippocampus. *Mol Psychiatry*. 2017; doi:10.1038/mp.2017.155
68. Kühn S, Gleich T, Lorenz RC, Lindenberger U, Gallinat J. Playing Super Mario induces structural brain plasticity: gray matter changes resulting from training with a commercial video game. *Mol Psychiatry*. 2014;19: 265–271. doi:10.1038/mp.2013.120

69. Kühn S, Lorenz R, Banaschewski T, Barker GJ, Büchel C, Conrod PJ, et al. Positive Association of Video Game Playing with Left Frontal Cortical Thickness in Adolescents. *PLOS ONE*. 2014;9: e91506. doi:10.1371/journal.pone.0091506
70. Stafford T, Dewar M. Tracing the Trajectory of Skill Learning With a Very Large Sample of Online Game Players. *Psychological Science*. 2013;25: 511–518. doi:10.1177/0956797613511466
71. Adams D, Mayer R. Examining the connection between dynamic and static spatial skills and video game performance. *Proceedings of the Cognitive Science Society*. 2012.
72. Deary IJ, Penke L, Johnson W. The neuroscience of human intelligence differences. *Nat Rev Neurosci*. 2010;11: 201–211. doi:10.1038/nrn2793
73. Green CS, Bavelier D. Exercising your brain: a review of human brain plasticity and training-induced learning. *Psychol Aging*. 2008;23: 692–701. doi:10.1037/a0014345
74. Green CS, Pouget A, Bavelier D. Improved probabilistic inference as a general learning mechanism with action video games. *Curr Biol*. 2010;20: 1573–1579. doi:10.1016/j.cub.2010.07.040
75. Ackerman PL. Nonsense, common sense, and science of expert performance: Talent and individual differences. *Intelligence*. 2014;45: 6–17. doi:10.1016/j.intell.2013.04.009
76. Engel D, Woolley AW, Jing LX, Chabris CF, Malone TW. Reading the Mind in the Eyes or Reading between the Lines? Theory of Mind Predicts Collective Intelligence Equally Well Online and Face-To-Face. *PLoS ONE*. 2014;9: e115212. doi:10.1371/journal.pone.0115212
77. Gosling SD, Vazire S, Srivastava S, John OP. Should We Trust Web-Based Studies? 2004;
78. Crump MJC, McDonnell JV, Gureckis TM. Evaluating Amazon’s Mechanical Turk as a Tool for Experimental Behavioral Research. *PLOS ONE*. 2013;8: e57410. doi:10.1371/journal.pone.0057410
79. Lara J, Cooper R, Nissan J, Ginty AT, Khaw K-T, Deary IJ, et al. A proposed panel of biomarkers of healthy ageing. *BMC Medicine*. 2015;13: 222. doi:10.1186/s12916-015-0470-9
80. Saperstein AM, Fuller RL, Avila MT, Adami H, McMahon RP, Thaker GK, et al. Spatial working memory as a cognitive endophenotype of schizophrenia: assessing risk for pathophysiological dysfunction. *Schizophr Bull*. 2006;32: 498–506. doi:10.1093/schbul/sbj072
81. Deary IJ. Cognitive epidemiology: Its rise, its current issues, and its challenges. *Personality and Individual Differences*. 2010;49: 337–343. doi:10.1016/j.paid.2009.11.012

Appendix D

Adapted from Kokkinakis, A. V., Cowling, P. I., Drachen, A., & Wade, A. R. (2017). Exploring the relationship between video game expertise and fluid intelligence. *PloS one*, 12(11), e0186621.

Here I provide the correlation values for subsets of the psychometric data that were incorporated into Study 1. The variable names are as follows:

ROTATION_ERRORS: Number of errors in the Rotation Task (Participant did not correctly rotate a letter presented to them).

MATH_CORRECT: Number of simple arithmetic operations that the participant got correct.

SYMMETRY_CORRECT: Number of Correct Answers in the first part of the Symmetry Span task (Participant had to identify whether an image was symmetric or not).

WASI_RAW_SCORE: Raw score in the Matrix Subtest.

LITERAL_RANKINGS: League of Legends Rank based on the players tier and division (Silver 5 is lower than Silver 4 etc, Silver 1 is lower than Gold 5 etc).

ROT_SPAN_ABSOLUTE: Participants were awarded points if they remembered a full sequence of arrows correctly.

SYM_SPAN_ABSOLUTE: Participants were awarded points if they remembered a full sequence of red squares appearing with the confines of matrix correctly. OS_SPAN_ABSOLUTE: Participants were awarded points if they remembered a full sequence of letters correctly.

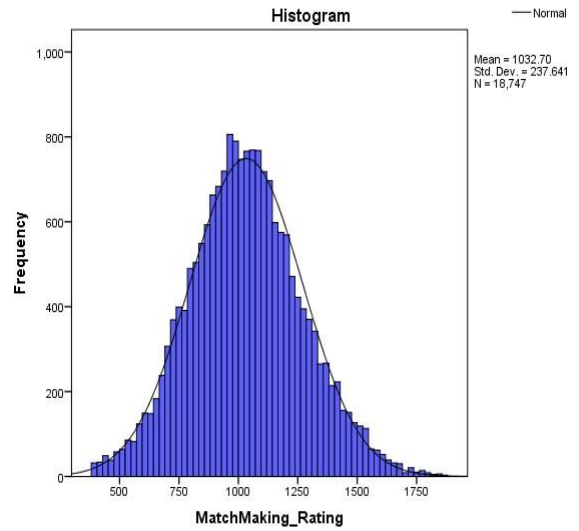
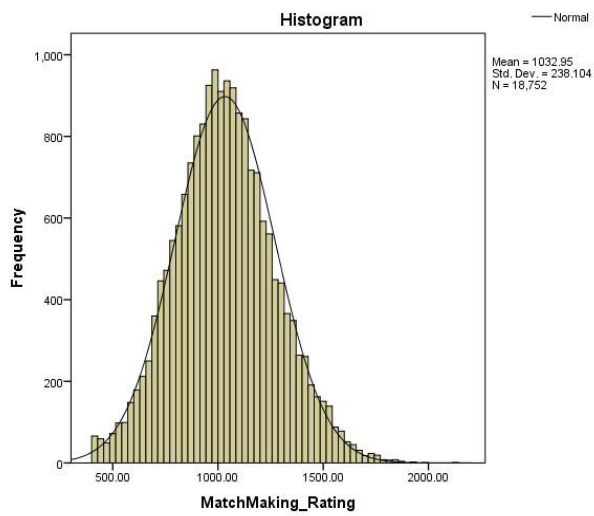
Correlations										
			ROTATION_ERRORS	MATH_CORRECT	SYMMETRY_CORRECT	WASLRW_SCORE	Literal_Ranking	ROT_SPAN_ABSOLUTE	SYM_SPAN_ABSOLUTE	OS_PAN_ABSOLUTE
Spearman's rho	ROTATION_ERRORS	Correlation Coefficient	1.000	-.371**	-.225*	-0.201	0.104	-.463**	-0.215	-0.140
		Sig. (1tailed)		0.003	0.049	0.071	0.225	0.000	0.061	0.154
		N	55	55	55	55	55	51	53	55
	MATH_CORRECT	Correlation Coefficient	-.371**	1.000	.255*	0.174	0.091	.254*	0.106	0.094
		Sig. (1tailed)	0.003		0.030	0.102	0.253	0.036	0.225	0.247
		N	55	55	55	55	55	51	53	55
	SYMMETRY_CORRECT	Correlation Coefficient	-.225*	.255*	1.000	0.000	0.115	.303*	0.124	.253*
		Sig. (1tailed)	0.049	0.030		0.499	0.201	0.015	0.188	0.031
		N	55	55	55	55	55	51	53	55
	WASLRW_SCORE	Correlation Coefficient	-0.201	0.174	0.000	1.000	.440**	.303*	.247*	0.071
		Sig. (1tailed)	0.071	0.102	0.499		0.000	0.014	0.036	0.302
		N	55	55	55	56	56	52	54	56
	Literal_Ranking	Correlation Coefficient	0.104	0.091	0.115	.440**	1.000	.260*	0.117	0.025
		Sig. (1tailed)	0.225	0.253	0.201	0.000		0.031	0.199	0.427
		N	55	55	55	56	56	52	54	56
	ROT_SPAN_ABSOLUTE	Correlation Coefficient	-.463**	.254*	.303*	.303*	.260*	1.000	.440**	0.198
		Sig. (1tailed)	0.000	0.036	0.015	0.014	0.031		0.001	0.080
		N	51	51	51	52	52	52	50	52
	SYM_SPAN_ABSOLUTE	Correlation Coefficient	-0.215	0.106	0.124	.247*	0.117	.440**	1.000	0.197
		Sig. (1tailed)	0.061	0.225	0.188	0.036	0.199	0.001		0.077
		N	53	53	53	54	54	50	54	54
OS_PAN_ABSOLUTE	Correlation Coefficient	-0.140	0.094	.253*	0.071	0.025	0.198	0.197	1.000	
	Sig. (1tailed)	0.154	0.247	0.031	0.302	0.427	0.080	0.077		
	N	55	55	55	56	56	52	54	56	

** . Correlation is significant at the 0.01 level (1-tailed).

* . Correlation is significant at the 0.05 level (1-tailed).

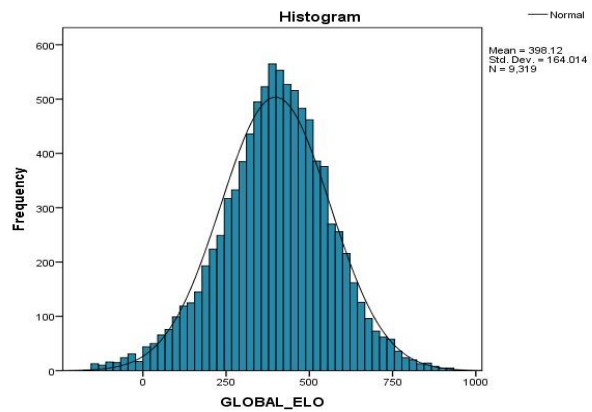
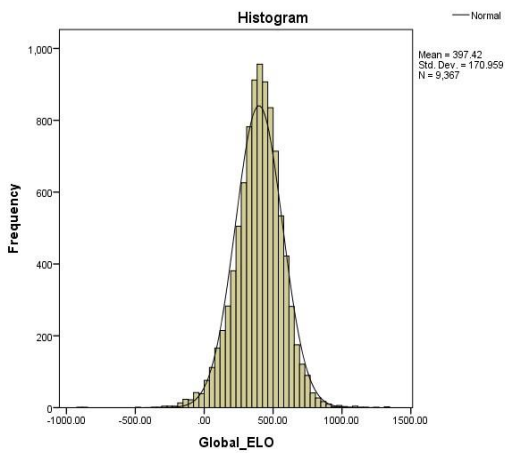
Appendix E

League of Legends



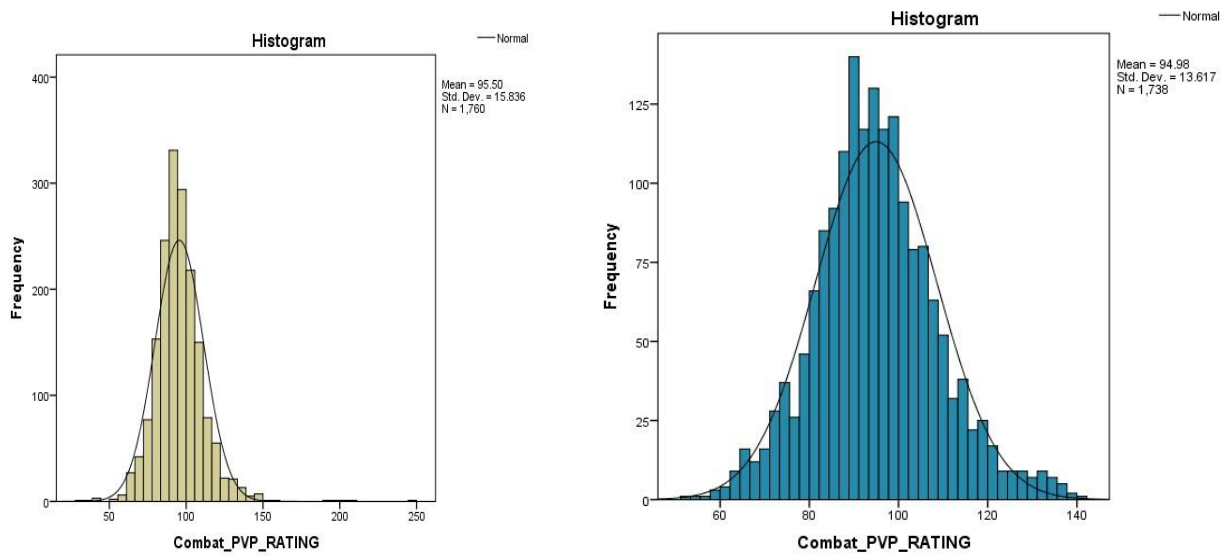
Histograms of League of Legends MMRs before (Left) and after (right) outlier rejection

Battlefield 3



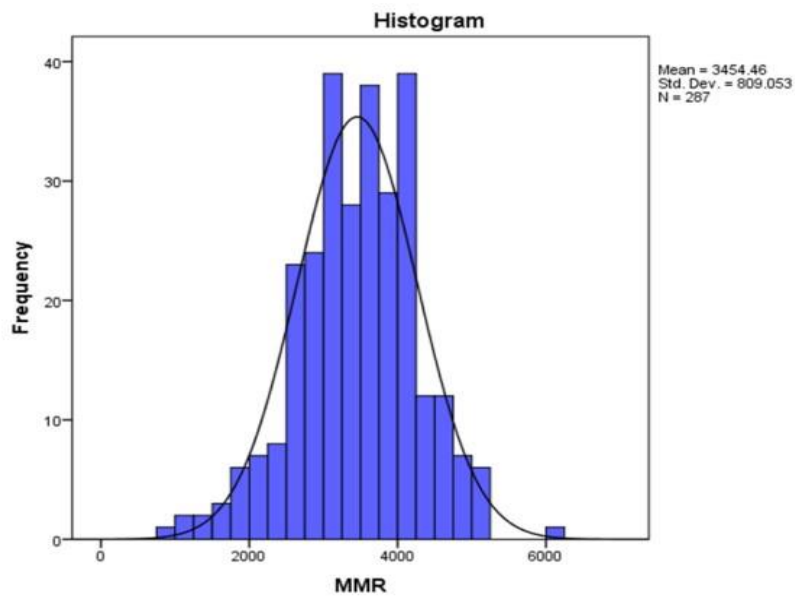
Histograms of Battlefield3 ELOs before (Left) and after (right) outlier rejection.

Destiny



Histograms of Destiny PVPs before (Left) and after (right) outlier rejection.

DOTA 2



Histograms of DOTA2 MMRs. No outliers were detected in this dataset.

Appendix F

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
Honesty_Humility	56	1.88	4.56	3.35	.69
Emotionality	56	1.75	4.75	2.9	.62
eXtraversion	56	1.63	4.56	3.4	.66
Agreeableness	56	1.50	4.63	3.17	.68
Conscientiousness	56	1.88	4.63	3.33	.58
Openness_to_experience	56	1.75	4.44	3.45	.64
Valid N (listwise)	56				

The HEXACO-100 descriptives.

HEXACO-100 Intercorrelations. Correlations significant at $p < .05$ are in yellow.

		Correlations					
Spearman's rho		Honesty_Humility	Emotionality	eXtraversion	Agreeableness	Conscientiousness	Openness_to_experience
	Honesty_Humility	1.000	.115	-.042	.414*	.163	.151
			.400	.737	.002	.231	.266
	Emotionality	.115	1.000	-.182	.033	.170	.195
		.400	.178	.178	.812	.210	.151
	eXtraversion	.56	.56	1.000	.254	.126	.054
		.757	.178	.178	.059	.353	.695
	Agreeableness	.56	.56	.56	1.000	.56	.56
		.414*	.033	.254	.109	-.009	.109
	Conscientiousness	.56	.56	.56	.56	1.000	.56
		.163	.170	.126	-.009	.328	.073
		.231	.210	.353	.550	1.000	.328
	Openness_to_experience	.56	.56	.56	.56	.56	1.000
		.266	.151	.054	.109	.328	.328
		.151	.151	.695	.426	.073	.073
		.56	.56	.56	.56	.56	.56

**. Correlation is significant at the 0.01 level (2-tailed).
 *. Correlation is significant at the 0.05 level (2-tailed).

References

- /dev: Matchmaking Real Talk – Nexus. (n.d.). Retrieved October 22, 2018, from <https://nexus.leagueoflegends.com/en-us/2018/02/dev-matchmaking-real-talk/>
- About Us. (n.d.). Retrieved May 15, 2019, from Arctic Shores website: <https://www.arcticshores.com/about-us/>
- Achim, A. M., Guitton, M., Jackson, P. L., Boutin, A., & Monetta, L. (2013). On what ground do we mentalize? Characteristics of current tasks and sources of information that contribute to mentalizing judgments. *Psychological Assessment, 25*(1), 117–126. <http://doi.org/10.1037/a0029137>
- Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin, 102*(1), 3–27. <https://doi.org/10.1037/0033-2909.102.1.3>
- Ackerman, P. L. (2014a). Facts are stubborn things. *Intelligence, 45*, 104–106. <https://doi.org/10.1016/j.intell.2014.01.002>
- Ackerman, P. L. (2014b). Nonsense, common sense, and science of expert performance: Talent and individual differences. *Intelligence, 45*, 6–17. <https://doi.org/10.1016/j.intell.2013.04.009>
- Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: the same or different constructs? *Psychological Bulletin, 131*(1), 30–60. <https://doi.org/10.1037/0033-2909.131.1.30>
- Adachi, P. J. C., & Willoughby, T. (2013). More than just fun and games: the longitudinal relationships between strategic video games, self-reported problem solving skills, and

academic grades. *Journal of Youth and Adolescence*, 42(7), 1041–1052.

<http://doi.org/10.1007/s10964-013-9913-9>

Adams, D., & Mayer, R. (2012). Examining the connection between dynamic and static spatial skills and video game performance. In *Proceedings of the Cognitive Science Society* (Vol. 34).

Allen, R. D. (2004). Working memory and the visual tracking of multiple independently moving targets (Ph.D.). University of Aberdeen. Retrieved from http://digitool.abdn.ac.uk/R/HD92H7Y45DXA1NNDIY5EI3EUAQI5YMR765FLX2QD72UP7248K9-05479?func=collections-result&collection_id=1455

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology*, 106(1), 20–29. <https://doi.org/10.1016/j.jecp.2009.11.003>

API Documentation - Riot Games API. (n.d.). Retrieved September 26, 2016, from <https://developer.riotgames.com/api/methods>

Appelbaum, L. G., Cain, M. S., Darling, E. F., & Mitroff, S. R. (2013). Action video game playing is associated with improved visual sensitivity, but not alterations in visual sensory memory. *Attention, Perception, & Psychophysics*, 75(6), 1161–1167. Retrieved from <http://link.springer.com/article/10.3758/s13414-013-0472-7>

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc*, 11(2), 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(4), 340–345. <https://doi.org/10.1080/00223890902935878>
- Ashton, M. C., Lee, K., Visser, B. A., & Pozzebon, J. A. (2008). Phobic tendency within the Five-Factor and HEXACO models of personality structure. *Journal of Research in Personality*, 42(3), 734–746. <https://doi.org/10.1016/j.jrp.2007.10.001>
- Aung, M. T., Bonometti, V., Drachen, A., Cowling, P. I., Kokkinakis, A. V., Yoder, C., & Wade, A. R. P. (2018). Predicting skill learning outcomes in a large, longitudinal MOBA dataset [Proceedings Paper]. Retrieved October 22, 2018, from <http://eprints.whiterose.ac.uk/132541/>
- Babaev, A. (2019). Retrieved from <https://twitter.com/Arteezy/status/1123595644587925504>
- Back, M. D., Schmukle, S. C., & Egloff, B. (2008). How extraverted is honey.bunny77@hotmail.de? Inferring personality from e-mail addresses. *Journal of Research in Personality*, 42(4), 1116–1122. <http://doi.org/10.1016/j.jrp.2008.02.001>
- Baker, C. A., Peterson, E., Pulos, S., & Kirkland, R. A. (2014). Eyes and IQ: A meta-analysis of the relationship between intelligence and “Reading the Mind in the Eyes.” *Intelligence*, 44, 78–92. <https://doi.org/10.1016/j.intell.2014.03.001>
- Banerjee, S., & Iglewicz, B. (2007). A Simple Univariate Outlier Identification Procedure Designed for Large Samples. *Communications in Statistics - Simulation and Computation*, 36(2), 249–263. <http://doi.org/10.1080/03610910601161264>
- Baniqued, P. L., Lee, H., Voss, M. W., Basak, C., Cosman, J. D., Desouza, S., ... Kramer, A. F. (2013). Selling points: What cognitive abilities are tapped by casual video games? *Acta Psychologica*, 142(1), 74–86. <https://doi.org/10.1016/j.actpsy.2012.11.009>

- Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The “Reading the Mind in the Eyes” Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 42(2), 241–251.
- Baron-Cohen, Simon. (1997). Are children with autism superior at folk physics? *New Directions for Child and Adolescent Development*, 1997(75), 45–54.
<https://doi.org/10.1002/cd.23219977504>
- Barrett, L. (2002). Do chess and GO need ‘g’? *Trends in Cognitive Sciences*, 6(12), 499.
- Bartle, R. (1996). Hearts, clubs, diamonds, spades: Players who suit MUDs. *Journal of MUD research*, 1(1), 19.
- Bashiri, H., Barahmand, U., Akabri, Z. S., Hossein Ghamari, G., & Vusugi, A. (2011). A Study Of The Psychometric Properties And The Standardization Of HEXACO Personality Inventory. *Procedia - Social and Behavioral Sciences*, 30, 1173–1176.
<https://doi.org/10.1016/j.sbspro.2011.10.228>
- Bates, T. C., & Eysenck, H. J. (1993). Intelligence, inspection time, and decision time. *Intelligence*, 17(4), 523–531. [https://doi.org/10.1016/0160-2896\(93\)90016-X](https://doi.org/10.1016/0160-2896(93)90016-X)
- Bavelier, D., Levi, D. M., Li, R. W., Dan, Y., & Hensch, T. K. (2010). Removing brakes on adult brain plasticity: from molecular to behavioral interventions. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 30(45), 14964–14971. <http://doi.org/10.1523/JNEUROSCI.4812-10.2010>
- Berg, J. M., Lilienfeld, S. O., & Waldman, I. D. (2013). Bargaining with the devil: Using economic decision-making tasks to examine the heterogeneity of psychopathic traits.

Journal of Research in Personality, 47(5), 472–482.

<http://doi.org/10.1016/j.jrp.2013.04.003>

Berger, A., Jones, L., Rothbart, M. K., & Posner, M. I. (2000). Computerized games to study the development of attention in childhood. *Behavior Research Methods, Instruments, & Computers*, 32(2), 297–303. <http://doi.org/10.3758/BF03207798>

Bilalić, M., McLeod, P., & Gobet, F. (2007). Does chess need intelligence? — A study with young chess players. *Intelligence*, 35(5), 457–470.

<https://doi.org/10.1016/j.intell.2006.09.005>

Bittner, A.C., Jones, M.B., Carter R.C., Shannon, R.C., Douglas, C., & Kennedy, R.S. (1981). STATISTICAL ISSUES IN PERFORMANCE TESTING: COLLECTED PAPERS. *Unclassified Military Report*.

Blackburn, J., & Kwak, H. (2014). STFU NOOB!: predicting crowdsourced decisions on toxic behavior in online games (pp. 877–888). *ACM Press*.

<http://doi.org/10.1145/2566486.2567987>

Blake, J. (2013, July 15). Video gamers given “athlete” status. *BBC*. Retrieved from <http://www.bbc.co.uk/newsbeat/23321595>

Blakemore, S.J. (2008). The social brain in adolescence. *Nature Reviews Neuroscience*, 9(4), 267–277. <http://doi.org/10.1038/nrn2353>

Blanch, A., Aluja, A., & Cornadó, M.-P. (2015). Sex differences in chess performance: Analyzing participation rates, age, and practice in chess tournaments. *Personality and Individual Differences*, 86, 117–121. <https://doi.org/10.1016/j.paid.2015.06.004>

- Blashki, K., & Nichol, S. (2005). Game geek's goss: linguistic creativity in young males within an online university forum (94/3 933k'5 9055oneone). *Australian Journal of Emerging Technologies and Society*, 3(2), 71–80.
- Bliss, J., Kennedy, R. S., Turnage, J. J., & Dunlap, W. P. (1991). Communalities of videogame performances with tracking tasks. *Perceptual and Motor Skills*, 73(1), 23–30.
<http://doi.org/10.2466/pms.1991.73.1.23>
- Bonny, J. W., & Castaneda, L. M. (2017). Number processing ability is connected to longitudinal changes in multiplayer online battle arena skill. *Computers in Human Behavior*, 66, 377–387. <https://doi.org/10.1016/j.chb.2016.10.005>
- Bonny, J. W., Castaneda, L. M., & Swanson, T. (2016). Using an International Gaming Tournament to Study Individual Differences in MOBA Expertise and Cognitive Skills. *In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 3473–3484). New York, NY, USA: ACM.
<https://doi.org/10.1145/2858036.2858190>
- Boot, W. R. (2015). Video games as tools to achieve insight into cognitive processes. *Frontiers in Psychology*, 6. <http://doi.org/10.3389/fpsyg.2015.00003>
- Boot, W. R., Kramer, A. F., Simons, D. J., Fabiani, M., & Gratton, G. (2008). The effects of video game playing on attention, memory, and executive control. *Acta Psychologica*, 129(3), 387–398. <https://doi.org/10.1016/j.actpsy.2008.09.005>
- Bošnjak-Nadž, K., Mejaški-Bošnjak, V., Popović Miočinović, L., Gverić Ahmetašević, S., Đaković, I., & Čikara Mladin, M. (2011). EARLY DETECTION OF CHILDREN AT NEURORISK AND ENROLMENT IN HABILITATION PROGRAMS. *Paediatrica Croatica*, 55(2). Retrieved from
https://hrcak.srce.hr/index.php?show=clanak&id_clanak_jezik=110276

- Brand, C. R., Freshwater, S., & Dockrell, W. B. (1989). Has There Been a ‘Massive’ Rise in IQ Levels in the West? Evidence from Scottish Children. *The Irish Journal of Psychology, 10*(3), 388–393. <https://doi.org/10.1080/03033910.1989.10557756>
- Breuer, J., Scharnow, M., & Quandt, T. (2015). Sore losers? A reexamination of the frustration–aggression hypothesis for colocated video game play. *Psychology of Popular Media Culture, 4*(2), 126–137. <http://doi.org/10.1037/ppm0000020>
- Brown, R. M., Brown, N. L., & Reid, K. (1992). Evidence for a player’s position advantage in a video game. *Perceptual and Motor Skills, 74*(2), 547–554. <http://doi.org/10.2466/pms.1992.74.2.547>
- Bucik, V., & Neubauer, A. C. (1996). Bimodality in the Berlin model of intelligence structure (BIS): A replication study. *Personality and Individual Differences, 21*(6), 987–1005. [https://doi.org/10.1016/S0191-8869\(96\)00129-8](https://doi.org/10.1016/S0191-8869(96)00129-8)
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon’s Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science, 6*(1), 3–5. <https://doi.org/10.1177/1745691610393980>
- Burgoyne, A. P., Sala, G., Gobet, F., Macnamara, B. N., Campitelli, G., & Hambrick, D. Z. (2016). The relationship between cognitive ability and chess skill: A comprehensive meta-analysis. *Intelligence, 59*, 72–83. <https://doi.org/10.1016/j.intell.2016.08.002>
- Burns, N. R., Nettelbeck, T., & Cooper, C. J. (1999). Inspection time correlates with general speed of processing but not with fluid ability. *Intelligence, 27*(1), 37–44. [http://doi.org/10.1016/S0160-2896\(99\)00013-6](http://doi.org/10.1016/S0160-2896(99)00013-6)

- Byrnes, M. M., & Spitz, H. H. (1977). Performance of retarded adolescents and nonretarded children on the Tower of Hanoi Problem. *American Journal of Mental Deficiency*, 81(6), 561–569.
- Cain, M. S., Landau, A. N., & Shimamura, A. P. (2012). Action video game experience reduces the cost of switching tasks. *Attention, Perception & Psychophysics*, 74(4), 641–647. <https://doi.org/10.3758/s13414-012-0284-1>
- Campbell, W. K., Bush, C. P., Brunell, A. B., & Shelton, J. (2005). Understanding the Social Costs of Narcissism: The Case of the Tragedy of the Commons. *Personality and Social Psychology Bulletin*, 31(10), 1358–1368.
<http://doi.org/10.1177/0146167205274855>
- Caron, M.-J., Mottron, L., Rainville, C., & Chouinard, S. (2004). Do high functioning persons with autism present superior spatial abilities? *Neuropsychologia*, 42(4), 467–481.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press.
- Casey, B. J., Jones, R. M., & Hare, T. A. (2008). The Adolescent Brain. *Annals of the New York Academy of Sciences*, 1124, 111–126. <http://doi.org/10.1196/annals.1440.010>
- Chory, R. M., & Goodboy, A. K. (2011). Is basic personality related to violent and non-violent video game play and preferences? *Cyberpsychology, Behavior and Social Networking*, 14(4), 191–198. <http://doi.org/10.1089/cyber.2010.0076>
- Chuderski, A. (2015). The broad factor of working memory is virtually isomorphic to fluid intelligence tested under time pressure. *Personality and Individual Differences*, 85, 98–104. <https://doi.org/10.1016/j.paid.2015.04.046>

- Colzato, L. S., van den Wildenberg, W. P. M., Zmigrod, S., & Hommel, B. (2013). Action video gaming and cognitive control: playing first person shooter games is associated with improvement in working memory but not action inhibition. *Psychological Research*, 77(2), 234–239. <https://doi.org/10.1007/s00426-012-0415-2>
- Colzato, L. S., van Leeuwen, P. J. A., van den Wildenberg, W. P. M., & Hommel, B. (2010). DOOM'd to Switch: Superior Cognitive Flexibility in Players of First Person Shooter Games. *Frontiers in Psychology*, 1. <https://doi.org/10.3389/fpsyg.2010.00008>
- Conway, A. R. A., & Kovacs, K. (2013). Chapter Seven - Individual Differences in Intelligence and Working Memory: A Review of Latent Variable Models. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 58, pp. 233–270). Academic Press. Retrieved from <http://www.sciencedirect.com/science/article/pii/B9780124072374000074>
- Craik, F. I. M., & Bialystok, E. (2006). Cognition through the lifespan: mechanisms of change. *Trends in Cognitive Sciences*, 10(3), 131–138. <https://doi.org/10.1016/j.tics.2006.01.007>
- Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a Tool for Experimental Behavioral Research. *PLOS ONE*, 8(3), e57410. <https://doi.org/10.1371/journal.pone.0057410>
- Dager, P. (2019). Retrieved from https://www.reddit.com/r/DotA2/comments/bj7bib/capitalist_on_the_state_of_pro_dota_in_na/em7j591/

- Davis, R. N. (1999). Web-based administration of a personality questionnaire: comparison with traditional methods. *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc*, 31(4), 572–577.
- de Bruin, A. B. H., Kok, E. M., Leppink, J., & Camp, G. (2014). Practice, intelligence, and enjoyment in novice chess players: A prospective study at the earliest stage of a chess career. *Intelligence*, 45, 18–25. <https://doi.org/10.1016/j.intell.2013.07.004>
- de Visser, L., van der Knaap, L. J., van de Loo, A. J. A. E., van der Weerd, C. M. M., Ohl, F., & van den Bos, R. (2010). Trait anxiety affects decision-making differently in healthy men and women: Towards gender-specific endophenotypes of anxiety. *Neuropsychologia*, 48(6), 1598–1606. <https://doi.org/10.1016/j.neuropsychologia.2010.01.027>
- Deary, I. J. (2010). Cognitive epidemiology: Its rise, its current issues, and its challenges. *Personality and Individual Differences*, 49(4), 337–343. <https://doi.org/10.1016/j.paid.2009.11.012>
- Deary, I. J., Penke, L., & Johnson, W. (2010). The neuroscience of human intelligence differences. *Nature Reviews Neuroscience*, 11(3), 201–211. <https://doi.org/10.1038/nrn2793>
- Deary, I. J., Whiteman, M. C., Starr, J. M., Whalley, L. J., & Fox, H. C. (2004). The impact of childhood intelligence on later life: following up the Scottish mental surveys of 1932 and 1947. *Journal of Personality and Social Psychology*, 86(1), 130–147. <https://doi.org/10.1037/0022-3514.86.1.130>
- DeWall, C. N., Buffardi, L. E., Bonser, I., & Keith Campbell, W. (2011). Narcissism and implicit attention seeking: Evidence from linguistic analyses of social networking and

- online presentation. *Personality and Individual Differences*, 51(1), 57–62.
<http://doi.org/10.1016/j.paid.2011.03.011>
- Detterman, D. K. (2014). Introduction to the intelligence special issue on the development of expertise: is ability necessary? *Intelligence*, 45, 1-5,
<https://doi.org/10.1016/j.intell.2014.02.004>.
- Digman, J. M. (1997). Higher-order factors of the Big Five. *Journal of Personality and Social Psychology*, 73(6), 1246–1256.
- Dodonova, Y. A., & Dodonov, Y. S. (2012). Processing speed and intelligence as predictors of school achievement: Mediation or unique contribution? *Intelligence*, 40(2), 163–171. <http://doi.org/10.1016/j.intell.2012.01.003>
- Dodonova, Y. A., & Dodonov, Y. S. (2013). Is there any evidence of historical slowing of reaction time? No, unless we compare apples and oranges. *Intelligence*, 41(5), 674–687. <http://doi.org/10.1016/j.intell.2013.09.001>
- Donchin, E. (1989). The learning strategies project: Introductory remarks. *Acta Psychologica*, 71(1–3), 1–15. [http://doi.org/10.1016/0001-6918\(89\)90002-4](http://doi.org/10.1016/0001-6918(89)90002-4)
- Donchin, E. (1995). Video games as research tools: The Space Fortress game. *Behavior Research Methods, Instruments, & Computers*, 27(2), 217–223.
<https://doi.org/10.3758/BF03204735>
- Donchin, E., Fabiani, M., & Sanders, A. (1989). The Learning Strategies Program: An Examination of the Strategies in Skill Acquisition. *Acta Psychologica*, 71. Retrieved from http://scholarcommons.usf.edu/psy_facpub/169
- Dota 2 - Steam Charts. (n.d.). Retrieved April 20, 2017, from <http://steamcharts.com/app/570>

- Drachen, A., Sifa, R., & Thureau, C. (2014). The name in the game: Patterns in character names and gamer tags. *Entertainment Computing*, 5(1), 21–32.
<http://doi.org/10.1016/j.entcom.2014.02.001>
- Dreyfuss, M., Caudle, K., Drysdale, A. T., Johnston, N. E., Cohen, A. O., Somerville, L. H., ... Casey, B. J. (2014). Teens impulsively react rather than retreat from threat. *Developmental Neuroscience*, 36(3-4), 220–227. <http://doi.org/10.1159/000357755>
- Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the short grit scale (grit-s). *Journal of Personality Assessment*, 91(2), 166–174.
<https://doi.org/10.1080/00223890802634290>
- Duckworth, A., Peterson, C., D Matthews, M., & R Kelly, D. (2007). Grit: Perseverance and Passion for Long-Term Goals. *Journal of Personality and Social Psychology*, 92, 1087–1101. <https://doi.org/10.1037/0022-3514.92.6.1087>
- Dumontheil, I., Apperly, I. A., & Blakemore, S.-J. (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental Science*, 13(2), 331–338.
<http://doi.org/10.1111/j.1467-7687.2009.00888.x>
- Elo, A. E. (2008). *The rating of chessplayers, past and present*. Bronx, N.Y.: Ishi Press International.
- Engel, D., Woolley, A. W., Jing, L. X., Chabris, C. F., & Malone, T. W. (2014). Reading the Mind in the Eyes or Reading between the Lines? Theory of Mind Predicts Collective Intelligence Equally Well Online and Face-To-Face. *PLoS ONE*, 9(12), e115212.
<https://doi.org/10.1371/journal.pone.0115212>

- Eysenck, M. W., & Calvo, M. G. (1992). Anxiety and Performance: The Processing Efficiency Theory. *Cognition and Emotion*, 6(6), 409–434.
<https://doi.org/10.1080/02699939208409696>
- Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: attentional control theory. *Emotion* (Washington, D.C.), 7(2), 336–353.
<https://doi.org/10.1037/1528-3542.7.2.336>
- Farrow, T. F. ., & Woodruff, P. W. . (2005). Neuroimaging of Forgivability. In *Handbook of Forgiveness* (pp. 259–273).
- Fite, P. J., Colder, C. R., Lochman, J. E., & Wells, K. C. (2008). Developmental trajectories of proactive and reactive aggression from fifth to ninth grade. *Journal of Clinical Child and Adolescent Psychology: The Official Journal for the Society of Clinical Child and Adolescent Psychology*, American Psychological Association, Division 53, 37(2), 412–421. <http://doi.org/10.1080/15374410801955920>
- Flanagan, D. P. (2008). The Cattell-Horn-Carroll Theory of Cognitive Abilities. In *Encyclopedia of Special Education* (pp. 368–382). American Cancer Society.
<https://doi.org/10.1002/9780470373699.speced0381>
- Flanagan, D. P., & Dixon, S. G. (2014). The Cattell-Horn-Carroll Theory of Cognitive Abilities. In *Encyclopedia of Special Education*. American Cancer Society.
<https://doi.org/10.1002/9781118660584.ese0431>
- Flynn, J. R. (2007). *What Is Intelligence?: Beyond the Flynn Effect*. Cambridge University Press.
- Flynn, J. R. (2013). The “Flynn Effect” and Flynn’s paradox. *Intelligence*, 41(6), 851–857.
<https://doi.org/10.1016/j.intell.2013.06.014>

- Foroughi, C. K., Serraino, C., Parasuraman, R., & Boehm-Davis, D. A. (2016). Can we create a measure of fluid intelligence using Puzzle Creator within Portal 2? *Intelligence*, 56, 58–64. <https://doi.org/10.1016/j.intell.2016.02.011>
- Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., & Engle, R. W. (2014). Shortened complex span tasks can reliably measure working memory capacity. *Memory & Cognition*, 43(2), 226–236. <https://doi.org/10.3758/s13421-014-0461-7>
- Foundry 10 Homepage. (2016, June 20). Retrieved August 20, 2017, from <http://foundry10.org/>
- Fozard, J. L., Vercryssen, M., Reynolds, S. L., Hancock, P. A., & Quilter, R. E. (1994). Age differences and changes in reaction time: the Baltimore Longitudinal Study of Aging. *Journal of Gerontology*, 49(4), P179-189.
- Gagnon, D. (1985). Videogames and spatial skills: An exploratory study. *ECTJ*, 33(4), 263–275. <http://doi.org/10.1007/BF02769363>
- Galvan, A., Hare, T. A., Parra, C. E., Penn, J., Voss, H., Glover, G., & Casey, B. J. (2006). Earlier Development of the Accumbens Relative to Orbitofrontal Cortex Might Underlie Risk-Taking Behavior in Adolescents. *The Journal of Neuroscience*, 26(25), 6885–6892. <http://doi.org/10.1523/JNEUROSCI.1062-06.2006>
- Game stats by champion - League of Legends. (n.d.). Retrieved October 22, 2018, from <http://na.op.gg/statistics/champion/>
- Gardner, M., & Steinberg, L. (2005). Peer influence on risk taking, risk preference, and risky decision making in adolescence and adulthood: an experimental study. *Developmental Psychology*, 41(4), 625–635. <http://doi.org/10.1037/0012-1649.41.4.625>

- Gaschler, R., Progscha, J., Smallbone, K., Ram, N., & Bilalić, M. (2014). Playing off the curve - testing quantitative predictions of skill acquisition theories in development of chess performance. *Frontiers in Psychology, 5*.
<https://doi.org/10.3389/fpsyg.2014.00923>
- Gaudiosi, J. (2012). Riot Games' League Of Legends Officially Becomes Most Played PC Game In The World. Retrieved from
<http://www.forbes.com/sites/johngaudiosi/2012/07/11/riot-games-league-of-legends-officially-becomes-most-played-pc-game-in-the-world/>
- Gazzard, A. (2013). *Mazes in Videogames: Meaning, Metaphor and Design*. Jefferson, North Carolina: McFarland.
- Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., ... Rapoport, J. L. (1999). Brain development during childhood and adolescence: a longitudinal MRI study. *Nature Neuroscience, 2*(10), 861–863.
<http://doi.org/10.1038/13158>
- Gignac, G. E. (2014). Fluid intelligence shares closer to 60% of its variance with working memory capacity and is a better indicator of general intelligence. *Intelligence, 47*, 122–133. <https://doi.org/10.1016/j.intell.2014.09.004>
- Gignac, G. E. (2015). Raven's is not a pure measure of general intelligence: Implications for g factor theory and the brief measurement of g. *Intelligence, 52*, 71–79.
<https://doi.org/10.1016/j.intell.2015.07.006>
- Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should We Trust Web-Based Studies? A Comparative Analysis of Six Preconceptions About Internet Questionnaires. *American Psychologist, 59*(2), 93–104. <https://doi.org/10.1037/0003-066X.59.2.93>

- Gottfredson, L. S. (1997). Mainstream science on intelligence: An editorial with 52 signatories, history and bibliography. *Intelligence*, 24(1), 13–23.
[https://doi.org/10.1016/S0160-2896\(97\)90011-8](https://doi.org/10.1016/S0160-2896(97)90011-8)
- Gottfredson, L. S. (1998). The general intelligence factor. Retrieved from
http://www.hucamabenelux.com/uploads/1/6/5/0/16501994/g-factor_intelligence_1998.pdf
- Grabner, R. H. (2014). The role of intelligence for performance in the prototypical expertise domain of chess. *Intelligence*, 45, 26–33. <https://doi.org/10.1016/j.intell.2013.07.023>
- Grabner, R. H., Stern, E., & Neubauer, A. C. (2007). Individual differences in chess expertise: a psychometric investigation. *Acta Psychologica*, 124(3), 398–420.
<https://doi.org/10.1016/j.actpsy.2006.07.008>
- Graham, C. (1981). Task Validation for Studies on Fragmented Sleep and Cognitive Efficiency Under Stress. Unclassified Military Report.
- Graham, C., Cook, M. R., Cohen, H. D., Phelps, J. W., & Gerkovich, M. M. (1985). STAR: A unique embedded performance assessment technique. *Behavior Research Methods, Instruments, & Computers*, 17(6), 642–651. <https://doi.org/10.3758/BF03200976>
- Graham, J. W. (2009). Missing data analysis: making it work in the real world. *Annual Review of Psychology*, 60, 549–576.
<https://doi.org/10.1146/annurev.psych.58.110405.085530>
- Granic, I., Lobel, A., & Engels, R. C. M. E. (2014). The benefits of playing video games. *The American Psychologist*, 69(1), 66–78. <http://doi.org/10.1037/a0034857>

- Green, C. Shawn, Pouget, A., & Bavelier, D. (2010). Improved probabilistic inference as a general learning mechanism with action video games. *Current Biology: CB*, 20(17), 1573–1579. <https://doi.org/10.1016/j.cub.2010.07.040>
- Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature*, 423(6939), 534–537. <http://doi.org/10.1038/nature01647>
- Green, C. S., & Bavelier, D. (2008). Exercising your brain: a review of human brain plasticity and training-induced learning. *Psychology and Aging*, 23(4), 692–701. <https://doi.org/10.1037/a0014345>
- Green, C. Shawn, & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature*, 423(6939), 534–537. <https://doi.org/10.1038/nature01647>
- Greitemeyer, T., & Osswald, S. (2010). Effects of prosocial video games on prosocial behavior. *Journal of Personality and Social Psychology*, 98(2), 211–221. <http://doi.org/10.1037/a0016997>
- Griffith, J. L., Voloschin, P., Gibb, G. D., & Bailey, J. R. (1983). Differences in eye-hand motor coordination of video-game users and non-users. *Perceptual and Motor Skills*, 57(1), 155–158. <http://doi.org/10.2466/pms.1983.57.1.155>
- Gross, M. U. M. (2006). Exceptionally Gifted Children: Long-Term Outcomes of Academic Acceleration and Nonacceleration. *Journal for the Education of the Gifted*, 29(4), 404–429. <https://doi.org/10.4219/jeg-2006-247>
- Guillon, M. J. (2010). Cross-modal compensation between name and visual aspect in socially active avatars. *Computers in Human Behavior*, 26(6), 1772–1776. <http://doi.org/10.1016/j.chb.2010.07.004>

- Gunter, B. (1998). *The Effects of Video Games on Children: The Myth Unmasked*. A&C Black.
- Hadwin, J. A., Brogan, J., & Stevenson, J. (2005). State anxiety and working memory in children: A test of processing efficiency theory. *Educational Psychology, 25*(4), 379–393. <https://doi.org/10.1080/01443410500041607>
- Hartley, C. A., & Phelps, E. A. (2012). Anxiety and Decision-Making. *Biological Psychiatry, 72*(2), 113–118. <https://doi.org/10.1016/j.biopsych.2011.12.027>
- Hawkins, G. E., Rae, B., Nesbitt, K. V., & Brown, S. D. (2012). Gamelike features might not improve data. *Behavior Research Methods, 45*(2), 301–318. <http://doi.org/10.3758/s13428-012-0264-3>
- Hayashi, A., Abe, N., Ueno, A., Shigemune, Y., Mori, E., Tashiro, M., & Fujii, T. (2010). Neural correlates of forgiveness for moral transgressions involving deception. *Brain Research, 1332*, 90–99. <http://doi.org/10.1016/j.brainres.2010.03.045>
- Hegarty, M., & Waller, D. (2004). A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence, 32*(2), 175–191. <https://doi.org/10.1016/j.intell.2003.12.001>
- Hern, A. (2018, May 6). Cambridge Analytica: how did it turn clicks into votes? *The Guardian*. Retrieved from <https://www.theguardian.com/news/2018/may/06/cambridge-analytica-how-turn-clicks-into-votes-christopher-wylie>

- Hoaglin, D. C., Iglewicz, B., & Tukey, J. W. (1986). Performance of Some Resistant Rules for Outlier Labelling. *Journal of the American Statistical Association*, 81(396), 991–999. <http://doi.org/10.1080/01621459.1986.10478363>
- Hoff, M. (2015). Retrieved from <http://www.twingalaxies.com/member.php/26500-1500points>.
- Holtzman, N. S., Vazire, S., & Mehl, M. R. (2010). Sounds like a Narcissist: Behavioral Manifestations of Narcissism in Everyday Life. *Journal of Research in Personality*, 44(4), 478–484. <http://doi.org/10.1016/j.jrp.2010.06.001>
- Horn, J. L., & Cattell, R. B. (1966). Age differences in primary mental ability factors. *Journal of Gerontology*, 21(2), 210–220.
- Howard, R. W. (2005). Objective evidence of rising population ability: a detailed examination of longitudinal chess data. *Personality and Individual Differences*, 38(2), 347–363. <https://doi.org/10.1016/j.paid.2004.04.013>
- Howard, R. W. (2009). Individual differences in expertise development over decades in a complex intellectual domain. *Memory & Cognition*, 37(2), 194–209. <https://doi.org/10.3758/MC.37.2.194>
- Howard, R. W. (2014). Learning curves in highly skilled chess players: A test of the generality of the power law of practice. *Acta Psychologica*, 151, 16–23. <https://doi.org/10.1016/j.actpsy.2014.05.013>
- Hunt, E., & Pellegrino, J. (1985). Using interactive computing to expand intelligence testing: A critique and prospectus. *Intelligence*, 9(3), 207–236. [http://doi.org/10.1016/0160-2896\(85\)90025-X](http://doi.org/10.1016/0160-2896(85)90025-X)

Jackson, D. N., Vernon, P. A., & Jackson, D. N. (1993). Dynamic spatial performance and general intelligence. *Intelligence*, 17(4), 451–460. [http://doi.org/10.1016/0160-2896\(93\)90012-T](http://doi.org/10.1016/0160-2896(93)90012-T)

Jensen, A. R. (1998). *The g Factor: The Science of Mental Ability* (1st Edition edition). Westport, Conn: Praeger.

Jensen, A. R. (2006). *Clocking the mind: Mental chronometry and individual differences*. Elsevier.

Johnson, J. H., Jones, M. B., & Kennedy, R. S. (1984). Cognitive Predictors of Tank Commander Performance. Retrieved from <http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA146843>

Johnson, W., & Bouchard, T. J. (2005). The structure of human intelligence: It is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. *Intelligence*, 33(4), 393–416. <https://doi.org/10.1016/j.intell.2004.12.002>

Jones, M. B. (1984). Video Games as Psychological Tests. *Simulation & Gaming*, 15(2), 131–157. <https://doi.org/10.1177/0037550084152001>

Jones, M. B., Dunlap, W. P., & Bilodeau, I. M. (1984). Factors appearing late in practice. *Organizational Behavior and Human Performance*, 33(2), 153–173. [http://doi.org/10.1016/0030-5073\(84\)90018-7](http://doi.org/10.1016/0030-5073(84)90018-7)

Jones, M. B., Kennedy, R. S., & Bittner, A. C., Jr. (1981). A Video Game for Performance Testing. *The American Journal of Psychology*, 94(1), 143–152. <http://doi.org/10.2307/1422349>

- Jones, M. B., Kennedy, R. S., & Bittner, A. C. (1981). Video games and convergence or divergence with practice. *US Naval Biodynamics Laboratory*, NBDL-81R010, 1-5.
- Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: a latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology. General*, 133(2), 189–217. <http://doi.org/10.1037/0096-3445.133.2.189>
- Kaufman, A. S. (2001). WAIS-III IQs, Horn's theory, and generational changes from young adulthood to old age. *Intelligence*, 29(2), 131–167. [https://doi.org/10.1016/S0160-2896\(00\)00046-5](https://doi.org/10.1016/S0160-2896(00)00046-5)
- Kaufman, A. S., & Lichtenberger, E. O. (2005). *Assessing Adolescent and Adult Intelligence*. John Wiley & Sons.
- Kawamura, M., Kobayashi, Y., & Morioka, S. (2012). Effects of Differences in Working Memory Capacity on Patterns of Word Generation (Vol. 2). Retrieved from <http://eric.ed.gov/?id=ED537410>
- Kensinger, E. A., Shearer, D. K., Locascio, J. J., Growdon, J. H., & Corkin, S. (2003). Working memory in mild Alzheimer's disease and early Parkinson's disease. *Neuropsychology*, 17(2), 230–239.
- Kim, Y. J., Engel, D., Woolley, A. W., Lin, J., McArthur, N., & Malone, T. W. (2016). What Makes a Strong Team? Collective Intelligence Predicts Team Performance in League of Legends. *Academy of Management Proceedings*, 2016(1), 13564. <https://doi.org/10.5465/AMBPP.2016.13564abstract>
- King, D. L., Delfabbro, P. H., & Griffiths, M. D. (2013). Trajectories of problem video gaming among adult regular gamers: an 18-month longitudinal study.

Cyberpsychology, Behavior and Social Networking, 16(1), 72–76.

<http://doi.org/10.1089/cyber.2012.0062>

Klimstra, T. A., Hale, W. W., Raaijmakers, Q. A. W., Branje, S. J. T., & Meeus, W. H. J.

(2009). Maturation of personality in adolescence. *Journal of Personality and Social Psychology*, 96(4), 898–912. <http://doi.org/10.1037/a0014746>

Kokkinakis, A. V., Lin, J., Pavlas, D., & Wade, A. R. (2016). What's in a name? Ages and names predict the valence of social interactions in a massive online game. *Computers in Human Behavior*, 55, Part B, 605–613. <https://doi.org/10.1016/j.chb.2015.09.034>

Kraut, R., Olson, J., Banaji, M., Bruckman, A., Cohen, J., & Couper, M. (2004).

Psychological research online: report of Board of Scientific Affairs' Advisory Group on the Conduct of Research on the Internet. *American Psychologist*, 59(2), 105.

Ku, J., Cho, W., Kim, J.-J., Peled, A., Wiederhold, B. K., Wiederhold, M. D., ... Kim, S. I.

(2003). A virtual environment for investigating schizophrenic patients' characteristics: assessment of cognitive and navigation ability. *Cyberpsychology & Behavior: The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society*, 6(4), 397–404. <http://doi.org/10.1089/109493103322278781>

Kwak, H., Blackburn, J., & Han, S. (2015). Exploring Cyberbullying and Other Toxic

Behavior in Team Competition Online Games (pp. 3739–3748). ACM Press.

<http://doi.org/10.1145/2702123.2702529>

Laakasuo, M., Palomäki, J., & Salmela, M. (2014). Experienced Poker Players Are

Emotionally Stable. *Cyberpsychology, Behavior, and Social Networking*, 17(10), 668–671. <https://doi.org/10.1089/cyber.2014.0147>

- Laakasuo, M., Palomäki, J., & Salmela, M. (2015). Emotional and Social Factors influence Poker Decision Making Accuracy. *Journal of Gambling Studies*, 31(3), 933–947.
<https://doi.org/10.1007/s10899-014-9454-5>
- Larson, M. K., Walker, E. F., & Compton, M. T. (2010). Early signs, diagnosis and therapeutics of the prodromal phase of schizophrenia and related psychotic disorders. *Expert Review of Neurotherapeutics*, 10(8), 1347–1359.
<https://doi.org/10.1586/ern.10.93>
- Latham, A. J., Patston, L. L. M., & Tippett, L. J. (2013). Just how expert are “expert” video-game players? Assessing the experience and expertise of video-game players across “action” video-game genres. *Frontiers in Psychology*, 4.
<https://doi.org/10.3389/fpsyg.2013.00941>
- Law, D. J., Morrin, K. A., & Pellegrino, J. W. (1995). Training effects and working memory contributions to skill acquisition in a complex coordination task. *Learning and Individual Differences*, 7(3), 207–234. [http://doi.org/10.1016/1041-6080\(95\)90011-X](http://doi.org/10.1016/1041-6080(95)90011-X)
- League of Legends. (2013, October 3). In Wikipedia, the free encyclopedia. Retrieved from http://en.wikipedia.org/w/index.php?title=League_of_Legends&oldid=575467374
- League of Legends. (2018). /dev diary: New Tiers & Placements in Ranked 2019. Retrieved from <https://www.youtube.com/watch?v=Jxt2Z4hEjwE>
- Lee, H., Voss, M. W., Prakash, R. S., Boot, W. R., Vo, L. T. K., Basak, C., ... Kramer, A. F. (2012). Videogame training strategy-induced change in brain function during a

- complex visuomotor task. *Behavioural Brain Research*, 232(2), 348–357.
<http://doi.org/10.1016/j.bbr.2012.03.043>
- Lee, K., & Ashton, M. C. (2016). Psychometric Properties of the HEXACO-100. *Assessment*.
<https://doi.org/10.1177/1073191116659134>
- Lee, K., Ogunfowora, B., & Ashton, M. C. (2005). Personality Traits Beyond the Big Five: Are They Within the HEXACO Space? *Journal of Personality*, 73(5), 1437–1463.
<https://doi.org/10.1111/j.1467-6494.2005.00354.x>
- Lerner, J. S., & Keltner, D. (2001). Fear, anger, and risk. *Journal of Personality and Social Psychology*, 81(1), 146–159.
- Levy, L. J., Astur, R. S., & Frick, K. M. (2005). Men and women differ in object memory but not performance of a virtual radial maze. *Behavioral Neuroscience*, 119(4), 853–862.
<http://doi.org/10.1037/0735-7044.119.4.853>
- Li, R., Polat, U., Makous, W., & Bavelier, D. (2009). Enhancing the contrast sensitivity function through action video game training. *Nature Neuroscience*, 12(5), 549–551.
<http://doi.org/10.1038/nn.2296>
- Linden, D. E. J. (2007). What, when, where in the brain? Exploring mental chronometry with brain imaging and electrophysiology. *Reviews in the Neurosciences*, 18(2), 159–171.
- Lintern, G. (1989). The learning strategies program: Concluding remarks. *Acta Psychologica*, 71(1–3), 301–309. [http://doi.org/10.1016/0001-6918\(89\)90013-9](http://doi.org/10.1016/0001-6918(89)90013-9)
- Logie, R., Baddeley, A., Mané, A., Donchin, E., & Sheptak, R. (1989). Working memory in the acquisition of complex cognitive skills. *Acta Psychologica*, 71(1–3), 53–87.
[https://doi.org/10.1016/0001-6918\(89\)90005-X](https://doi.org/10.1016/0001-6918(89)90005-X)

- Lohman, D. F. (1996). Spatial ability and g. In *Human abilities: Their nature and measurement* (pp. 97–116). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.
- LoL Stats, Record Replay, Database, Guide, MMR - OP.GG. (n.d.). Retrieved October 22, 2018, from <http://euw.op.gg/>
- Lowrie, T., Logan, T., & Ramful, A. (2017). Visuospatial training improves elementary students' mathematics performance. *British Journal of Educational Psychology*, 87(2), 170–186. <https://doi.org/10.1111/bjep.12142>
- Maccow, G. (2011). Overview of WASI-II. Pearson Corporation.
- Mané, A., & Donchin, E. (1989). The space fortress game. *Acta Psychologica*, 71(1–3), 17–22. [https://doi.org/10.1016/0001-6918\(89\)90003-6](https://doi.org/10.1016/0001-6918(89)90003-6)
- Markham, A., & Buchanan, E. (2012). Ethical decision-making and Internet research (version 2.0): Recommendations from the AoIR Ethics Working Committee. Available Online: [Aoir.org/reports/ethics2](http://aoir.org/reports/ethics2). Pdf.
- Martin, C. F., Bhui, R., Bossaerts, P., Matsuzawa, T., & Camerer, C. (2014a). Chimpanzee choice rates in competitive games match equilibrium game theory predictions. *Scientific Reports*, 4. <http://doi.org/10.1038/srep05182>
- Martinovic, D., Burgess, G. H., Pomerleau, C. M., & Marin, C. (2015). Comparison of children's gaming scores to NEPSY-II scores: Validation of computer games as cognitive tools. *Computers in Human Behavior*, 49, 487–498. <http://doi.org/10.1016/j.chb.2015.03.039>
- Mason, O., Linney, Y., & Claridge, G. (2005). Short scales for measuring schizotypy. *Schizophrenia Research*, 78(2-3), 293–296. <http://doi.org/10.1016/j.schres.2005.06.020>

Matchmaking | Dota 2. (n.d.). Retrieved March 28, 2017, from

<http://blog.dota2.com/2013/12/matchmaking/>

Matchmaking Guide. (n.d.). Retrieved 28 March 2017, from

<http://support.riotgames.com/hc/en-us/articles/201752954-Matchmaking-Guide>

McDermott, A. F., Bavelier, D., & Green, C. S. (2014). Memory abilities in action video game players. *Computers in Human Behavior*, 34, 69–78.

<http://doi.org/10.1016/j.chb.2014.01.018>

McGrew, K. S. (2005). The Cattell-Horn-Carroll Theory of Cognitive Abilities: Past, Present, and Future. In *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (pp. 136–181). New York, NY, US: Guilford Press.

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, 37(1), 1–10. <https://doi.org/10.1016/j.intell.2008.08.004>

McPherson, J. L. (2008). Development and assessment of computer-game-like tests of human cognitive abilities. (Thesis). Retrieved from

<https://digital.library.adelaide.edu.au/dspace/handle/2440/50739>

McPherson, J., & Burns, N. R. (2007). Gs Invaders: Assessing a computer game-like test of processing speed. *Behavior Research Methods*, 39(4), 876–883.

<https://doi.org/10.3758/BF03192982>

McPherson, J., & Burns, N. R. (2007). Gs Invaders: Assessing a computer game-like test of processing speed. *Behavior Research Methods*, 39(4), 876–883.

<http://doi.org/10.3758/BF03192982>

- McPherson, J., & Burns, N. R. (2008). Assessing the validity of computer-game-like tests of processing speed and working memory. *Behavior Research Methods*, 40(4), 969–981. <http://doi.org/10.3758/BRM.40.4.969>
- McPherson, J., & Burns, N. R. (2005). A speeded coding task using a computer-based mouse response. *Behavior Research Methods*, 37, 538–544.
- Mesholam-Gately, R. I., Giuliano, A. J., Goff, K. P., Faraone, S. V., & Seidman, L. J. (2009). Neurocognition in first-episode schizophrenia: a meta-analytic review. *Neuropsychology*, 23(3), 315–336. <https://doi.org/10.1037/a0014708>
- Meta - OpenDota - Dota 2 Statistics. (n.d.). Retrieved October 23, 2018, from <https://www.opendota.com/meta?>
- Metastrategy dictionary definition | metastrategy defined. (n.d.). Retrieved October 23, 2018, from <http://www.yourdictionary.com/metastrategy>
- Mincic, A. M. (2015). Neuroanatomical correlates of negative emotionality-related traits: A systematic review and meta-analysis. *Neuropsychologia*, 77, 97–118. <https://doi.org/10.1016/j.neuropsychologia.2015.08.007>
- Mineski Interview: “KR solo queue is so competitive...I have Plat 1 friends here. When they went to the Philippines, they all hit Challenger.”* (2018). Retrieved from <https://www.invglobal.com/articles/5898/mineski-interview-kr-solo-queue-is-so-competitive-i-have-plat-1-friends-here-when-they-went-to-the-philippines-they-all-hit-challenger>
- Miu, A. C., Heilman, R. M., & Houser, D. (2008). Anxiety impairs decision-making: Psychophysiological evidence from an Iowa Gambling Task. *Biological Psychology*, 77(3), 353–358. <https://doi.org/10.1016/j.biopsycho.2007.11.010>

- Mix, K. S., & Cheng, Y.-L. (2012). The relation between space and math: developmental and educational implications. *Advances in Child Development and Behavior*, 42, 197–243.
- Moinester, M., & Gottfried, R. (2014). Sample size Estimation for correlations with prespecified confidence interval. *The Quantitative Methods for Psychology*, 10(2), 124-130.
- Morton, N., & Morris, R. G. (1995). Image transformation dissociated from visuospatial working memory. *Cognitive Neuropsychology*, 12(7), 767–791.
<https://doi.org/10.1080/02643299508251401>
- Moul, C. C., & Nye, J. V. C. (2009). Did the Soviets collude? A statistical analysis of championship chess 1940–1978. *Journal of Economic Behavior & Organization*, 70(1), 10–21. <https://doi.org/10.1016/j.jebo.2009.01.009>
- Moutafi, J., Furnham, A., & Tsaousis, I. (2006). Is the relationship between intelligence and trait Neuroticism mediated by test anxiety? *Personality and Individual Differences*, 40(3), 587–597. <https://doi.org/10.1016/j.paid.2005.08.004>
- Nacke, L. E., Bateman, C., & Mandryk, R. L. (2011, October). BrainHex: preliminary results from a neurobiological gamer typology survey. In *International Conference on Entertainment Computing* (pp. 288-293). Springer, Berlin, Heidelberg.
- Nuechterlein, K. H., Barch, D. M., Gold, J. M., Goldberg, T. E., Green, M. F., & Heaton, R. K. (2004). Identification of separable cognitive factors in schizophrenia. *Schizophrenia Research*, 72(1), 29–39. <https://doi.org/10.1016/j.schres.2004.09.007>
- Nyhus, E., & Barceló, F. (2009). The Wisconsin Card Sorting Test and the cognitive assessment of prefrontal executive functions: A critical update. *Brain and Cognition*, 71(3), 437–451. <http://doi.org/10.1016/j.bandc.2009.03.005>

- Okagaki, L., & Frensch, P. A. (1994). Effects of video game playing on measures of spatial performance: Gender effects in late adolescence. *Journal of Applied Developmental Psychology*, 15(1), 33–58. [http://doi.org/10.1016/0193-3973\(94\)90005-1](http://doi.org/10.1016/0193-3973(94)90005-1)
- Owen, A., & Highfield, R. (2010). The 12 pillars of wisdom: Understanding intelligence. *New Scientist*, 208(2784), 38–43. [http://doi.org/10.1016/S0262-4079\(10\)62699-1](http://doi.org/10.1016/S0262-4079(10)62699-1)
- Palau, M., Marron, E. M., Viejo-Sobera, R., & Redolar-Ripoll, D. (2017). Neural Basis of Video Gaming: A Systematic Review. *Frontiers in Human Neuroscience*, 11, 248. <https://doi.org/10.3389/fnhum.2017.00248>
- Palomäki, J., Laakasuo, M., & Salmela, M. (2013). ‘This is just so unfair!’: A qualitative analysis of loss-induced emotions and tilting in on-line poker. *International Gambling Studies*, 13(2), 255–270. <https://doi.org/10.1080/14459795.2013.780631>
- Park, J., Song, Y., & Teng, C.-I. (2011). Exploring the links between personality traits and motivations to play online games. *Cyberpsychology, Behavior and Social Networking*, 14(12), 747–751. <http://doi.org/10.1089/cyber.2010.0502>
- Park, S., & Gooding, D. C. (2014). Working memory impairment as an endophenotypic marker of a schizophrenia diathesis. *Schizophrenia Research: Cognition*, 1(3), 127–136. <http://doi.org/10.1016/j.scog.2014.09.005>
- Parker, S. (2014). Were the Victorians cleverer than us? Maybe, maybe not. *Intelligence*, 47, 1–2. <http://doi.org/10.1016/j.intell.2014.08.002>
- Patch 4.4 Notes | League of Legends. (n.d.). Retrieved October 23, 2018, from <https://na.leagueoflegends.com/en/news/game-updates/patch/patch-44-notes>
- Paulhus, D. L. (1998). Interpersonal and intrapsychic adaptiveness of trait self-enhancement: a mixed blessing? *Journal of Personality and Social Psychology*, 74(5), 1197–1208.

- Pellegrino, J. W., Hunt, E. B., Abate, R., & Farr, S. (1987). A computer-based test battery for the assessment of static and dynamic spatial reasoning abilities. *Behavior Research Methods, Instruments, & Computers*, 19(2), 231–236.
<http://doi.org/10.3758/BF03203790>
- Petzoldt, T., Bellem, H., & Krems, J.F. (2014). The Critical Tracking Task: A potentially useful method to assess driver distraction? *Human Factors*, 56(4), 784–803.
[doi:10.1177/0018720813501864](https://doi.org/10.1177/0018720813501864)
- Pihlstrøm, L., Wiethoff, S., & Houlden, H. (2017). Genetics of neurodegenerative diseases: an overview. *Handbook of Clinical Neurology*, 145, 309–323.
<https://doi.org/10.1016/B978-0-12-802395-2.00022-5>
- Ployhart, R. E., & Vandenberg, R. J. (2010). Longitudinal Research: The Theory, Design, and Analysis of Change. *Journal of Management*, 36(1), 94–120.
<https://doi.org/10.1177/0149206309352110>
- Pudge - Melee, Disabler, Durable, Initiator, Nuker. (n.d.). Retrieved October 22, 2018, from <https://www.dotabuff.com/heroes/pudge>
- Quiroga, M. A., Escorial, S., Román, F. J., Morillo, D., Jarabo, A., Privado, J., ... Colom, R. (2015). Can we reliably measure the general factor of intelligence (g) through commercial video games? Yes, we can! *Intelligence*, 53, 1–7.
<https://doi.org/10.1016/j.intell.2015.08.004>
- Quiroga, M. A., Herranz, M., Gómez-Abad, M., Kebir, M., Ruiz, J., & Colom, R. (2009). Video-games: Do they require general intelligence? *Computers & Education*, 53(2), 414–418. <https://doi.org/10.1016/j.compedu.2009.02.017>

r/leagueoflegends - What's In A Name? - Innapropriate Name Change Audit FAQ. (n.d.).

Retrieved September 3, 2018, from

https://www.reddit.com/r/leagueoflegends/comments/3ooizi/whats_in_a_name_innapropriate_name_change_audit_faq/

Rabbitt, P., Banerji, N., & Szymanski, A. (1989). Space fortress as an IQ test? Predictions of learning and of practised performance in a complex interactive video-game. *Acta Psychologica*, 71(1–3), 243–257. [https://doi.org/10.1016/0001-6918\(89\)90011-5](https://doi.org/10.1016/0001-6918(89)90011-5)

Ranked Play FAQ. (n.d.). Retrieved March 28, 2017, from

<http://support.riotgames.com/hc/en-us/articles/204010760-Ranked-Play-FAQ>

Ranked. (n.d.). Retrieved 9 October 2017, from <https://hearthstone.gamepedia.com/Ranked>

Raven's Progressive Matrices | Pearson Assessment. (n.d.). Retrieved May 15, 2019, from

<https://www.pearsonclinical.co.uk/Psychology/AdultCognitionNeuropsychologyandLanguage/AdultGeneralAbilities/Ravens-Progressive-Matrices/Ravens-Progressive-Matrices.aspx>

Raven, J. (1989). The Raven Progressive Matrices: A Review of National Norming Studies and Ethnic and Socioeconomic Variation within the United States. *Journal of Educational Measurement*, 26(1), 1–16. Retrieved from JSTOR.

Raven, J. (2000). The Raven's Progressive Matrices: Change and Stability over Culture and Time. *Cognitive Psychology*, 41(1), 1–48. <http://doi.org/10.1006/cogp.1999.0735>

re:Work - Riot Games: Assessing toxicity in the workplace. (n.d.). Retrieved September 3,

2018, from <https://rework.withgoogle.com/case-studies/riot-games-assessing-toxicity/>

- Robinson, L., Tang, E., & Taylor, J.-P. (2015). Dementia: timely diagnosis and early intervention. *The BMJ*, 350, h3029. <http://doi.org/10.1136/bmj.h3029>
- Röhlcke, S., Bäcklund, C., Sörman, D. E., & Jonsson, B. (2018). Time on task matters most in video game expertise. *PLOS ONE*, 13(10), e0206555. <https://doi.org/10.1371/journal.pone.0206555>
- Rosser, J. C., Lynch, P. J., Cuddihy, L., Gentile, D. A., Klonsky, J., & Merrell, R. (2007). The impact of video games on training surgeons in the 21st century. *Archives of Surgery* (Chicago, Ill.: 1960), 142(2), 181–186; discussion 186. <http://doi.org/10.1001/archsurg.142.2.181>
- Rumbaugh, D. M., Richardson, W. K., Washburn, D. A., Savage-Rumbaugh, E. S., & Hopkins, W. D. (1989). Rhesus monkeys (*Macaca mulatta*), video tasks, and implications for stimulus-response spatial contiguity. *Journal of Comparative Psychology* (Washington, D.C.: 1983), 103(1), 32–38.
- Ryan, E. (1986). Memory for goblins: A computer for assessing and training working memory skill. *Clinical Gerontologist*, 6, 64-67.
- Ryan, E. B. (1994). Memory for Goblins: A computer game for assessing and training working memory skill. *Computers in Human Services*, 213–217.
- Salthouse, T. A. (2009). When does age-related cognitive decline begin? *Neurobiology of Aging*, 30(4), 507–514. <https://doi.org/10.1016/j.neurobiolaging.2008.09.023>
- Salthouse, T. A. (2010). Selective review of cognitive aging. *Journal of the International Neuropsychological Society: JINS*, 16(5), 754–760. <https://doi.org/10.1017/S1355617710000706>

- Salthouse, T. A., Pink, J. E., & Tucker-Drob, E. M. (2008). Contextual analysis of fluid intelligence. *Intelligence*, 36(5), 464–486. <https://doi.org/10.1016/j.intell.2007.10.003>
- Schaie, K. W. (1993). The Seattle Longitudinal Studies of Adult Intelligence. *Current Directions in Psychological Science*, 2(6), 171–175. <https://doi.org/10.1111/1467-8721.ep10769721>
- Schaie, K. W., & Hofer, S. M. (2001). Longitudinal studies in aging research. In *Handbook of the psychology of aging, 5th ed* (pp. 53–77). San Diego, CA, US: Academic Press.
- Schaie, K. W., & Strother, C. R. (1968). A cross-sequential study of age changes in cognitive behavior. *Psychological Bulletin*, 70(6, Pt.1), 671–680. <https://doi.org/10.1037/h0026811>
- Sorby, S., Veurink, N., & Streiner, S. (2018). Does spatial skills instruction improve STEM outcomes? The answer is ‘yes.’ *Learning and Individual Differences*, 67, 209–222. <https://doi.org/10.1016/j.lindif.2018.09.001>
- Schei, T. H. (2014). Drone. Documentary.
- Schneider, D., Slaughter, V. P., & Dux, P. E. (2017). Current evidence for automatic Theory of Mind processing in adults. *Cognition*, 162, 27–31. <https://doi.org/10.1016/j.cognition.2017.01.018>
- Schneider, W. J., & Newman, D. A. (2015). Intelligence is multidimensional: Theoretical review and implications of specific cognitive abilities. *Human Resource Management Review*, 25(1), 12–27. <https://doi.org/10.1016/j.hrmr.2014.09.004>
- Schneider, W., & McGrew, K. (2012). The Cattell-Horn-Carroll model of intelligence. In *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 99–144).

- Schroeder, J.E., Dyer, F.N., Czerny, P., Youngling, E.W., & Gillotti, D.P. (1986).
VIDEODISC INTERPERSONAL SKILLS TRAINING AND ASSESSMENT
(VISTA): OVERVIEW AND FINDINGS, VOLUME 1. Retrieved from
<http://www.dtic.mil/dtic/tr/fulltext/u2/a168287.pdf>
- Seasonal Ranked Update | Dota 2. (n.d.). Retrieved October 22, 2018, from
<http://blog.dota2.com/2017/11/seasonal-ranked-update/>
- Shackman, A. J., Sarinopoulos, I., Maxwell, J. S., Pizzagalli, D. A., Lavric, A., & Davidson,
R. J. (2006). Anxiety selectively disrupts visuospatial working memory. *Emotion*
(Washington, D.C.), 6(1), 40–61. <https://doi.org/10.1037/1528-3542.6.1.40>
- Shah, A., & Frith, U. (1993). Why do autistic individuals show superior performance on the
block design task? *Journal of Child Psychology and Psychiatry, and Allied*
Disciplines, 34(8), 1351–1364.
- Shapiro, K. L., & Raymond, J. E. (1989). Training of efficient oculomotor strategies
enhances skill acquisition. *Acta Psychologica*, 71(1–3), 217–242.
[http://doi.org/10.1016/0001-6918\(89\)90010-3](http://doi.org/10.1016/0001-6918(89)90010-3)
- Shebilske, W. L., Volz, R. A., Gildea, K. M., Workman, J. W., Nanjanath, M., Cao, S., &
Whetzel, J. (2005). Revised Space Fortress: a validation study. *Behavior Research*
Methods, 37(4), 591–601.
- Sheppard, L. D., & Vernon, P. A. (2008). Intelligence and speed of information-processing:
A review of 50 years of research. *Personality and Individual Differences*, 44(3), 535–
551. <https://doi.org/10.1016/j.paid.2007.09.015>

- Shute, V. J., Ventura, M., & Ke, F. (2015). The power of play: The effects of Portal 2 and Lumosity on cognitive and noncognitive skills. *Computers & Education*, 80, 58–67. <https://doi.org/10.1016/j.compedu.2014.08.013>
- Siler, K. (2010). Social and Psychological Challenges of Poker. *Journal of Gambling Studies*, 26(3), 401–420. <https://doi.org/10.1007/s10899-009-9168-2>
- Sims, V. K., & Mayer, R. E. (2002). Domain specificity of spatial expertise: the case of video game players. *Applied Cognitive Psychology*, 16(1), 97–115. <http://doi.org/10.1002/acp.759>
- Smith, R. E., Smoll, F. L., Cumming, S. P., & Grossbard, J. R. (2006). Measurement of Multidimensional Sport Performance Anxiety in Children and Adults: The Sport Anxiety Scale-2. *Journal of Sport and Exercise Psychology*, 28(4), 479–501. <https://doi.org/10.1123/jsep.28.4.479>
- Sorkin, A., (2006). Virtual Reality for Diagnostic Assessment of Schizophrenia Deficits. Phd Thesis.
- Spearman, C. (1904). “General Intelligence,” Objectively Determined and Measured. *The American Journal of Psychology*, 15(2), 201–292. <https://doi.org/10.2307/1412107>
- Spitz, H. H. (1978). The universal nature of human intelligence: Evidence from games. *Intelligence*, 2(4), 371–379. [http://doi.org/10.1016/0160-2896\(78\)90026-0](http://doi.org/10.1016/0160-2896(78)90026-0)
- Spitz, H. H., & DeRISI, D. T. (1978). Porteus Maze Test performance of retarded young adults and nonretarded children. *American Journal of Mental Deficiency*, 83(1), 40–43.

- Spitz, Herman H., & Winters, E. A. (1977). Tic-Tac-Toe performance as a function of maturational level of retarded adolescents and nonretarded children. *Intelligence*, 1(1), 108–117. [https://doi.org/10.1016/0160-2896\(77\)90030-7](https://doi.org/10.1016/0160-2896(77)90030-7)
- Stafford, T., & Dewar, M. (2014). Tracing the Trajectory of Skill Learning With a Very Large Sample of Online Game Players. *Psychological Science*, 25(2), 511–518. <http://doi.org/10.1177/0956797613511466>
- Stafford, T., Devlin, S., Sifa, R., & Drachen, A. (2017). Exploration and Skill Acquisition in a Major Online Game [Proceedings Paper]. Retrieved May 14, 2019, from The 39th Annual Meeting of the Cognitive Science Society (CogSci) website: <http://eprints.whiterose.ac.uk/118051/>
- SteamID, Steam Account Names, Merging Accounts and Deleting Accounts - Managing Your Account Features - Knowledge Base - Steam Support. (n.d.). Retrieved November 26, 2018, from https://support.steampowered.com/kb_article.php?ref=1558-qyax-1965
- Steinberg, L. (2004). Risk taking in adolescence: what changes, and why? *Annals of the New York Academy of Sciences*, 1021, 51–58. <http://doi.org/10.1196/annals.1308.005>
- Steinberg, L., Graham, S., O'Brien, L., Woolard, J., Cauffman, E., & Banich, M. (2009). Age Differences in Future Orientation and Delay Discounting. *Child Development*, 80(1), 28–44. <http://doi.org/10.1111/j.1467-8624.2008.01244.x>
- Steger, D., Schroeders, U., & Wilhelm, O. (2019). On the dimensionality of crystallized intelligence: A smartphone-based assessment. *Intelligence*, 72, 76–85. <https://doi.org/10.1016/j.intell.2018.12.002>

- Stieff, M. (2013). Sex Differences in the Mental Rotation of Chemistry Representations. *Journal of Chemical Education*, 90(2), 165–170. <https://doi.org/10.1021/ed300499t>
- Stollery, B. T. (1996). Reaction time changes in workers exposed to lead. *Neurotoxicology and Teratology*, 18(4), 477–483.
- Strang, S., Utikal, V., Fischbacher, U., Weber, B., & Falk, A. (2014). Neural Correlates of Receiving an Apology and Active Forgiveness: An fMRI Study. *PLoS ONE*, 9(2), e87654. <http://doi.org/10.1371/journal.pone.0087654>
- Strohschneider, S., & Güss, D. (1999). The Fate of the Moros: A Cross-cultural Exploration of Strategies in Complex and Dynamic Decision Making. *International Journal of Psychology*, 34(4), 235–252. <https://doi.org/10.1080/002075999399873>
- Subrahmanyam, K., & Greenfield, P. M. (1994). Effect of video game practice on spatial skills in girls and boys. *Journal of Applied Developmental Psychology*, 15(1), 13–32. [http://doi.org/10.1016/0193-3973\(94\)90004-3](http://doi.org/10.1016/0193-3973(94)90004-3)
- Surrender at 15 available now | League of Legends. (n.d.). Retrieved October 22, 2018, from <https://oce.leagueoflegends.com/en/news/game-updates/features/surrender-15-available-now>
- Tam, Y. P., Wong, T. T.-Y., & Chan, W. W. L. (2019). The relation between spatial skills and mathematical abilities: The mediating role of mental number line representation. *Contemporary Educational Psychology*, 56, 14–24. <https://doi.org/10.1016/j.cedpsych.2018.10.007>

- Tassi, P. (n.d.). Riot's "League of Legends" Reveals Astonishing 27 Million Daily Players, 67 Million Monthly. Retrieved October 22, 2018, from <https://www.forbes.com/sites/insertcoin/2014/01/27/riots-league-of-legends-reveals-astonishing-27-million-daily-players-67-million-monthly/>
- Tekofsky, S., Spronck, P., Goudbeek, M., Plaat, A., & van den Herik, J. (2015). Past Our Prime: A Study of Age and Play Style Development in Battlefield 3. *IEEE Transactions on Computational Intelligence and AI in Games*, 7(3), 292–303. <https://doi.org/10.1109/TCIAIG.2015.2393433>
- Tekofsky, S., Spronck, P., Plaat, A., Van den Herik, J., & Broersen, J. (2013). Play style: Showing your age. In *2013 IEEE Conference on Computational Intelligence in Games (CIG)* (pp. 1–8). <https://doi.org/10.1109/CIG.2013.6633616>
- Teng, C.-I. (2008). Personality differences between online game players and nonplayers in a student sample. *Cyberpsychology & Behavior*, 11(2), 232–234. <http://doi.org/10.1089/cpb.2007.0064>
- The DOD Symposium USAF Academy, Colorado Springs.
- Thomas, P., Goudemand, M., & Rousseaux, M. (1999). Attentional resources in major depression. *European Archives of Psychiatry and Clinical Neuroscience*, 249(2), 79–85. <https://doi.org/10.1007/s004060050070>
- Thompson, J. J., Blair, M. R., & Henrey, A. J. (2014). Over the Hill at 24: Persistent Age-Related Cognitive-Motor Decline in Reaction Times in an Ecologically Valid Video Game Task Begins in Early Adulthood. *PLoS ONE*, 9(4), e94215. <https://doi.org/10.1371/journal.pone.0094215>

Towne, T. J. M., Ericsson, K. A. P., & Sumner, A. M. (2014). Uncovering mechanisms in video game research: suggestions from the expert-performance approach. *Cognition*, 5, 161. <http://doi.org/10.3389/fpsyg.2014.00161>

Tsuang, M. (2000). Schizophrenia: genes and environment. *Biological Psychiatry*, 47(3), 210–220. [https://doi.org/10.1016/S0006-3223\(99\)00289-9](https://doi.org/10.1016/S0006-3223(99)00289-9)

Twenge, J. M., & Foster, J. D. (2008). Mapping the scale of the narcissism epidemic: Increases in narcissism 2002–2007 within ethnic groups. *Journal of Research in Personality*, 42(6), 1619–1622.

Twenge, J. M., & Foster, J. D. (2010). Birth cohort increases in narcissistic personality traits among American college students, 1982–2009. *Social Psychological and Personality Science*, 1(1), 99–106. Retrieved from <http://spp.sagepub.com/content/1/1/99.short>

Twenge, J. M., Konrath, S., Foster, J. D., Campbell, W. K., & Bushman, B. J. (2008). Egos inflating over time: a cross-temporal meta-analysis of the Narcissistic Personality Inventory. *Journal of Personality*, 76(4), 875–902; discussion 903–928. <http://doi.org/10.1111/j.1467-6494.2008.00507.x>

Unsworth, N., Redick, T. S., McMillan, B. D., Hambrick, D. Z., Kane, M. J., & Engle, R. W. (2015). Is Playing Video Games Related to Cognitive Abilities? *Psychological Science*, 26(6), 759–774. <https://doi.org/10.1177/0956797615570367>

Unsworth, Nash, & Engle, R. W. (2006). Simple and complex memory spans and their relation to fluid abilities: Evidence from list-length effects. *Journal of Memory and Language*, 54(1), 68–80. <https://doi.org/10.1016/j.jml.2005.06.003>

- Vaci, N., Gula, B., & Bilalić, M. (2014). Restricting range restricts conclusions. *Frontiers in Psychology*, 5. <https://doi.org/10.3389/fpsyg.2014.00569>
- Valve. (n.d.). DOTA 2 UPDATE - JANUARY 14TH, 2016. Retrieved from Product Update website: <https://store.steampowered.com/news/19988/>
- Valve. (2016). 7.00 Update – The New Journey Begins. Retrieved from <http://blog.dota2.com/2016/12/7-00-update-the-new-journey-begins/>
- van Bokhoven, I., van Goozen, S. H. M., van Engeland, H., Schaal, B., Arseneault, L., Séguin, J. R., ... Tremblay, R. E. (2006). Salivary testosterone and aggression, delinquency, and social dominance in a population-based longitudinal study of adolescent males. *Hormones and Behavior*, 50(1), 118–125. <http://doi.org/10.1016/j.yhbeh.2006.02.002>
- Van den Akker, A. L., Deković, M., Asscher, J., & Prinzie, P. (2014). Mean-level personality development across childhood and adolescence: A temporary defiance of the maturity principle and bidirectional associations with parenting. *Journal of Personality and Social Psychology*, 107(4), 736–750. <http://doi.org/10.1037/a0037248>
- van Leeuwen, S., Müller, N. G., & Melloni, L. (2009). Age effects on attentional blink performance in meditation. *Consciousness and Cognition*, 18(3), 593–599. <https://doi.org/10.1016/j.concog.2009.05.001>
- Vine, S. J., & Wilson, M. R. (2011). The influence of quiet eye training and pressure on attention and visuo-motor control. *Acta Psychologica*, 136(3), 340–346. <https://doi.org/10.1016/j.actpsy.2010.12.008>

- Vries, R. E. de, Vries, A. de, & Feij, J. A. (2009). Sensation seeking, risk-taking, and the HEXACO model of personality. *Personality and Individual Differences*, 47(6), 536–540. <https://doi.org/10.1016/j.paid.2009.05.029>
- Wagenmakers, E.-J., & van der Maas, H. (2008). Clocking the mind: Mental chronometry and individual differences, A.R. Jensen. Elsevier, Amsterdam (2006). *Intelligence*, 36(5), 493–494. <http://doi.org/10.1016/j.intell.2007.09.001>
- Washburn, D. A. (2003). The games psychologists play (and the data they provide). *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc*, 35(2), 185–193.
- Wechsler Abbreviated Scale of Intelligence - Second Edition (WASI-II) | Pearson Assessment. (n.d.). Retrieved 5 May 2016, from <http://www.pearsonclinical.co.uk/Psychology/AdultCognitionNeuropsychologyandLanguage/AdultGeneralAbilities/WASI-II/WechslerAbbreviatedScaleofIntelligenceSecondEdition.aspx>
- Weller, J. A., & Thulin, E. W. (2012). Do honest people take fewer risks? Personality correlates of risk-taking to achieve gains and avoid losses in HEXACO space. *Personality and Individual Differences*, 53(7), 923–926. <https://doi.org/10.1016/j.paid.2012.06.010>
- Weller, J. A., & Tikir, A. (2011). Predicting domain-specific risk taking with the HEXACO personality structure. *Journal of Behavioral Decision Making*, 24(2), 180–201. <https://doi.org/10.1002/bdm.677>
- Weniger, G., & Irle, E. (2008). Allocentric memory impaired and egocentric memory intact as assessed by virtual reality in recent-onset schizophrenia. *Schizophrenia Research*, 101(1–3), 201–209. <http://doi.org/10.1016/j.schres.2008.01.011>

- Wetherell, J. L., Reynolds, C. A., Gatz, M., & Pedersen, N. L. (2002). Anxiety, Cognitive Performance, and Cognitive Decline in Normal Aging. *The Journals of Gerontology: Series B*, 57(3), P246–P255. <https://doi.org/10.1093/geronb/57.3.P246>
- What is the Metagame? (n.d.). Retrieved October 23, 2018, from <https://magic.wizards.com/en/articles/archive/magic-academy/what-metagame-2007-01-06>
- Williams, S. B., & Leavitt, H. J. (1947). Group opinion as a predictor of military leadership. *Journal of Consulting Psychology*, 11(6), 283–291.
- Wilson, M., Smith, N. C., Chattington, M., Ford, M., & Marple-Horvat, D. E. (2006). The role of effort in moderating the anxiety-performance relationship: Testing the prediction of processing efficiency theory in simulated rally driving. *Journal of Sports Sciences*, 24(11), 1223–1233. <https://doi.org/10.1080/02640410500497667>
- Woodley, M. A., Nijenhuis, J., & Murphy, R. (2013). Were the Victorians cleverer than us? The decline in general intelligence estimated from a meta-analysis of the slowing of simple reaction time. *Intelligence*, 41(6), 843–850. <http://doi.org/10.1016/j.intell.2013.04.006>
- Woolley, A. W., Aggarwal, I., & Malone, T. W. (2015). Collective Intelligence and Group Performance. *Current Directions in Psychological Science*, 24(6), 420–424. <https://doi.org/10.1177/0963721415599543>
- Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a Collective Intelligence Factor in the Performance of Human Groups. *Science*, 330(6004), 686–688. <https://doi.org/10.1126/science.1193147>

- Worth, N. C., & Book, A. S. (2014). Personality and behavior in a massively multiplayer online role-playing game. *Computers in Human Behavior*, 38, 322–330.
<http://doi.org/10.1016/j.chb.2014.06.009>
- Wu, H.-K., & Shah, P. (n.d.). Exploring visuospatial thinking in chemistry learning. *Science Education*, 88(3), 465–492. <https://doi.org/10.1002/sce.10126>
- Yee, N., Ducheneaut, N., Nelson, L., & Likarish, P. (2011). Introverted elves & conscientious gnomes: the expression of personality in world of warcraft. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 753–762). ACM.
Retrieved from <http://dl.acm.org/citation.cfm?id=1979052>
- Yoon, G., & Vargas, P. T. (2014). Know Thy Avatar The Unintended Effect of Virtual-Self Representation on Behavior. *Psychological Science*, 0956797613519271.
<http://doi.org/10.1177/0956797613519271>
- Zagal, J. P., Tomuro, N., & Shepitsen, A. (2012). Natural language processing in game studies research: An overview. *Simulation & Gaming*, 43(3), 356-373.
- Zawadzki, J. A., Girard, T. A., Fousias, G., Rodrigues, A., Siddiqui, I., Lerch, J. P., Remington, G., & Wong, A. H. (2013). Simulating real world functioning in schizophrenia using a naturalistic city environment and single-trial, goal-directed navigation. *Frontiers in Behavioral Neuroscience*, 7, 180.
<http://doi.org/10.3389/fnbeh.2013.00180>
- Zeigler-Hill, V., & Monica, S. (2015). The HEXACO model of personality and video game preferences. *Entertainment Computing*, 11, 21-26.

Zhang, Z. (2015). Missing data exploration: highlighting graphical presentation of missing pattern. *Annals of Translational Medicine*, 3(22), 356.

<https://doi.org/10.3978/j.issn.2305-5839.2015.12.28>

Zygmunt, C., & R. Smith, M. (2014). Robust factor analysis in the presence of normality violations, missing data, and outliers: Empirical questions and possible solutions. *The Quantitative Methods for Psychology*, 10, 40–55.

<https://doi.org/10.20982/tqmp.10.1.p040>