Proceedings of the Tenth Annual AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE 2014)

# Developing Social Identity Models of Players from Game Telemetry Data

**Chong-U Lim** and **D. Fox Harrell**
Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA
{culim, fox.harrell}@mit.edu

**Abstract**

In this paper, we present an approach to modeling aspects of the identities of videogame players by data mining game telemetry information on in-game player performance and customization preferences. Our model demonstrates that such data can be used to reveal aspects of the identities players express by their social networking profile information. We tested our model on players of the multiplayer first-person shooter videogame *Team Fortress 2*. It was able to significantly explain the variances of the players' number of friends (35.1%), number of uploaded screenshots (49.6%), and number of uploaded videos (39.2%) of their profiles on the gaming social network *Steam*. Our results revealed several findings, such as criteria indicating how players customized avatars differently according to notions of aesthetics and practicality, and how these notions contributed to predicting their number of friends on their social networking profiles. Responses evaluated from a conducted survey reaffirmed several of these findings.

## Introduction

As developers of systems such as videogames and social networks, it is important to understand more about who the users of our systems are. In videogames, *player modeling* enables designers to gain insight into how players are performing in videogames. This includes exhibited behaviors within the game and interactions that occur outside the game. Artificial Intelligence (AI) approaches are often used by researchers to develop computational player models and a common approach now is the use of *game telemetry*, which collects data on fine-grained statistics of players actions.

Collecting data via *game data mining* provides direct access to player information for analyzing large populations of player demographics, behaviors, and usage patterns (Bauckhage et al. 2012; Mellon 2009; Drachen et al. 2013). Examples include data on players' overall play time, highest number of points obtained per round, or number total number of enemies killed. Game AI researchers Christian Bauckhage et al. note that game data mining adds explanatory power and overcomes shortcomings of traditional tools for player research, such as surveys that are harder to evaluate and prone to biases (Bauckhage et al. 2012).



We seek to use such data mining approaches in developing models of players, considering both their virtual identities within games, as well as identities expressed outside of the game, such as on other platforms and even in everyday life. While game telemetry data provides virtual in-game identity information, we believe players express other aspects of their identities through their social networking profiles, which include information such as the number of friends a player has, how often the player interacts with others, or even attitudes toward publicly sharing information and media with others. Our approach is facilitated by the emergence of videogames that integrate social networking information, often for leaderboard creation and the sharing of media with friends such as in-game screenshots. We seek to use both game telemetry data and social networking profile information to construct what Harrell here terms **cross-platform identity** models, taking the perspective of Harrell's work on "blended identities" (Harrell 2010). Under this view, most digital self-representations are projections of some aspects of a real player (e.g., preferences, control, appearance, aspirations, etc.) onto the actual implemented (virtual) representation. While social networking profiles are "virtual identities" (Yee and Bailenson 2007), we refer to aspects of the players real-world identity as reflected in social networking profiles as the subset of a players real world preferences, affiliations, etc. that may be conveyed through their profiles.

## A Data-Driven Cross-Platform Identity Model

Specifically in this paper, we seek to construct a computational model of players using a data-driven approach by game telemetry data mining of their in-game behavior (gameplay achievements and avatar customization preferences) and aspects of their real-world social identity (social networking profile behavior). To test this computational identity model, we used players' in-game behavioral data to predict their real-world social identity data. Previous work has demonstrated that linear models perform effectively in mapping in-game behaviors to real-world demographics in other games like *World of Warcraft* (Yee et al. 2012) and *Battlefield 3* (Tekofsky, Spronck, and Broersen 2013). Our approach differs in several ways. We emphasize the need for avatar customization preference data in addition to gameplay statistics when data mining game telemetry. As social networking profiles often are used to express aspects of players' real-world identities, we are able to test our model against more than one outcome variable. To reaffirm our findings, we conducted a player survey with questions on gameplay behavior, customization preferences, and identity.

## Related Work

In this section, we discuss previous work related to player modeling, game telemetry data mining, how player identities are represented computationally in different applications and systems, and how they relate to this paper's research.

### Player Modeling & Game Telemetry Data Mining

Previous research has shown how players' gameplay behaviors are closely related to aspects of their "real-world" identities, such as their ages (Tekofsky et al. 2013), personality (Tekofsky, Spronck, and Broersen 2013), motivations (Canossa, Martinez, and Togelius 2013), behavior (Yee et al. 2011), and customization preferences (Lim and Harrell 2013). Here, we use the phrase *real-world* to describe aspects of player's social identity that exists outside of the game world, and the distinction between real-world and virtual-world identities is of interest here as people might create digital identities that are distinct from the ones they use in everyday life (Yee and Bailenson 2007). A common approach is the use of surveys or standardized tests that use a predefined set of questions, which map player responses to respective identity model categories. Here, we model aspects of a players' social identities by data mining social network profile information to augment our collected survey data.

### Computational Identities

Harrell describes how *shared technical underpinnings* (Harrell 2010) exist between computational identity representation systems such as character avatars, and social networking profiles. Thus, we believe that both virtual identity information (e.g., in-game behaviors) and real-world information (e.g., social network profiles) result in identity models that are directly comparable and often inter-operable at a technical level. Social scientist Nick Yee (Yee 2006) defines three broad categories of player types based on their motivations. They are "achievements" (e.g., overall numerical competence, achieved status), "social" (e.g., meaningful dealings, friendships) and "immersion" (e.g., avatar customization, discovery, role-playing). As Yee's immersion category is broadly defined across sub-categories that differs from other definitions of immersion, we focus on "customization" as our third category of interest. While these categories are similar to games researcher Richard Bartle's player types (e.g., achievers, socializers) of Multi-User Dungeon (MUD) players, Yee argues against Bartle's assumption that players' preference for one type suppresses preference for another. Instead, player representations and motivations are present across all the categories and this helps to "emphasize player behavior" (Manninen and Kujanpää 2007). Hence, our model is constructed from data collected across the achievement, customization, and social categories.

## Test Case: Steam & Team Fortress 2

As an initial test case application of this model, we applied it to a gaming social network (Steam ) and an online multiplayer first-person-shooter videogame (Team Fortress 2 ).

**Steam** The social network used in our test case is *Steam*, an integrated game distribution platform and social networking site. *Steam* allows users to manage their collections of games purchased through it. *Steam* requires users to sign up for a *Steam* account with a unique identifier called a "Steam Id" in order to create individual *Steam Profiles*. The games available include both *Valve*-published and third-party titles. For social networking, players connect to one another through their friends lists and can send messages, view others profiles, or find others to play with. Players may also create, manage and join "groups," which are communities of players with similar interests. Players are able to see whenever their friends are online and what games they are playing, which facilitates playing together. *Steam* also allows users to connect to other social networking applications, such as *Facebook*. In 2011, there were approximately 82.2 million friendship edges, 1824 games and 1.98 million groups (Becker et al. 2012). At the time of writing, the number of players concurrently active on Steam is between 2–6 million (Steam Stats 2013). Its network size, gamer-centric demographic, and tight integration with games makes it an interesting domain in which to research the relationships between social network behavior and gameplay.

**Team Fortress 2** The videogame used in our test case is *Team Fortress 2* (TF2), which was released in 2007. In 2011, it became the top-played game on *Steam* by player count, and is third-highest at the time of writing. The high number of players, constant active development, and close integration with *Steam* makes it an interesting and relevant videogame to study. There are nine character classes, each with a unique visual 3-dimensional (3D) model, abilities, and weapons. In traditional gameplay, two teams of players compete against each other, with each player character as a member of one of several different classes that teams need to balance out in order to be more effective at accomplishing their goals. The game characters and classes all have a base visual appearance and set of attributes associated per class. However, players are able to customize their characters through a *loadout* menu. There are 7 customizable *slots* and classes have no headwear or accessories by default and each possess a default set of weapons. Players are able to customize their characters with items which provide functional benefits (i.e., weapons that deal more damage) and items which modify the visual appearance of the character (i.e. a whimsical hat that gives rabbit ears). Avatar customization through such modular 3D visual graphics is a key component in digital identity representation (Harrell 2010) and an important motivation for play (Yee 2006).

## Data Collection

An online survey containing questions about the players' identity, gameplay behavior, and customization preferences was posted on the *Team Fortress 2* boards on social news and discussion site *Reddit*. Each player provided their *Steam ID* to enable us to data-mine public game telemetry from their *Team Fortress 2* accounts. We obtained a total of 219 responses. We removed a duplicate entry and another 9 responses due to privacy settings set on their game profiles preventing us from retrieving game telemetry data. Thus, we ended up with a final total of 209 valid responses. Apart from 3 players who chose not to respond, respondents were mainly males ($91.4\%$). The gender distribution for users of *Steam* and *Reddit* is likely to be more balanced than that of our responders. Players were mostly in the age groups of $18-24$ ( $85\%$) and $25-34$ ($\sim 13\%$). These demographics represented actively playing, intermediate-to-advanced TF2 players and likely a majority of users of *Steam*. As the TF2 *sub-reddit* was used, players who responded are not representative of *Reddit* users in general and are likely younger.

Our data mining system collected public game telemetry data using each player's *Steam ID*. We collected **gameplay statistics** of each player (e.g., damage dealt, number of kills, buildings destroyed) and their **customization data** for each loadout slot (e.g., primary weapon, secondary weapon, accessories). Gameplay statistics have two types: **overall performance** (i.e., collected over the lifetime of playing) and **best-performance** (i.e., attained during single round at any point of time) data. As the *Steam API* provides gameplay statistics per class, values across the nine playable classes were averaged into single value per gameplay statistic. As in (Lim and Harrell 2013), we obtained community-derived price lists from a 3rd party site in order to assign monetary values to the customization items as a measure of customization preferences. Thus, we had a total of 43 data variables (36 gameplay + 7 customization). For modeling the *Social* characteristics of players, we collected **social networking profile** data variables from each player's public *Steam Community Profile Page*, resulting in 9 variables for commonly used features on *Steam* (e.g., number of friends, uploaded screenshots) used as outcome variables model validation.

## Methods

In this section, we outline the steps undertaken for constructing our player identity model using gameplay statistics and avatar customization data. We also describe how we validated our model by using it to predict the social networking profile information of players. Also, we outline the steps undertaken in collecting responses for our player survey.

### Model Construction

1. In order to differentiate between gameplay statistics and avatar customization, we grouped our in-game telemetry data variables into to the following three sets.
   (a) Accumulated Gameplay Stats
   (b) Best Performing Gameplay Stats
   (c) Customization Slots
2. To reduce the dimensionality (number of data variables), Principal Component Analysis (PCA) was performed on each set of data variables.
   (a) Varimax rotation was used to ensure that the coefficients of each data variable loads maximally onto a single resultant principal component (loadings).
   (b) This enables us to identify what each principal component describes based on high loading data variables.
3. We selected principal components with inclusion criteria:
   (a) The principal component has an eigen-value $>= 1$.
   (b) There are $>= 2$ variables in the principal component with loadings of an absolute value $>= 0.7$.

   This inclusion criteria matches those used by (Tekofsky et al. 2013) to ensure selected principal components were *semantically coherent*, though ours has a stricter loading threshold of $0.7$ (versus $0.5$) for increased robustness.

### Model Validation

1. In order to model aspects of a player's real-world social identity, we selected outcome variables that were obtained from a players' *Steam* social networking profile page.
2. We obtained a total of 9 outcome variables. The number of each player's 1) artworks 2) badges obtained, 3) friends, 4) groups, 5) guides, 6) reviews, 7) screenshots uploaded, 8) videos uploaded, and 9) *Steam Workshop* items.
3. To validate our model against these outcome variables, Multiple Linear Regression (MLR) was performed for each of them. Backward-selection and t-tests were used to select the factors that contribute most to the model.
4. We identified models for outcome variables that had an explained variance of $>= 30\%$. It indicates that a moderate amount of the variance of the distribution of the outcome variable was accounted for by our model.
5. We compared our model against two control experiments by repeating step (3) to construct two control models:
   (a) C1: Model with the 36 gameplay-related data variables.
   (b) C2: Model using all 43 data variables.

### Survey Evaluation

1. We selected questions that related to aspects of our model.
2. We analyzed the responses and compared them against findings from our model construction and validation.

## Results & Analysis

In this section, we present the results and analysis from both the model construction and model validation steps.

### Model Construction

Recall that our aim in model construction was to use data variables for in-game achievement (accumulated and best-performing stats) and customization preferences to predict a player's social networking profile information. Originally, the total number of data variables were 18 for Accumulated Gameplay Stats, 18 for Best-Performing Gameplay Stats, and 7 for Customization Slot Preferences for a total of 43 variables. Using PCA, we were able to reduce the number of variables by $79\%$ to $4$ for Accumulated Gameplay Stats, 3 for Best-Performing Gameplay Stats, and 2 for Customization Slot Preferences for a total of 9 principal components as variables. We next describe how each principal component was obtained. For brevity, we only go into the numerical details of Accumulated Gameplay Stats by referring to the PCA results in Table 2 of the Appendix, since the process of selecting the principal components based on our selection criteria is similar for all three sets of data variables.

**Accumulated Gameplay Stats** Our analysis is based on using the loadings for each variable in Table 2, to assign semantically coherent names to each component. We obtained 4 principal components that matched our inclusion criteria. PC1 described **Overall Combat** (high kill, high defense) players, PC2 described **Overall Support** (high healing, buildings built, provided teleports) players, PC3 described **Overall Stealth** players (high backstabs and leached health are traits of the Spy class), and PC4 described **Overall Capturer** (capturing points are traits of the Scout class) players. Thus, the model of Accumulated Gameplay Stats has the formula: $A = a_1.Overall.Combat + a_2.Overall.Support + a_3.Overall.Stealth + a_4.Overall.Capturer$. This implies that our model distinguishes between players based on their exhibited competency across the major gameplay roles over the entire period of time that they've been playing the game.

**Best-Performing Gameplay Stats** From the results of PCA, we obtained 3 principal components that matched our inclusion criteria. PC1 described **Best-Performing Artillery** (high kills, kill assists) players, PC2 described **Best-Performing Stealth** (destroying sentries via "sapping" is characteristic of the Spy class), and PC3 described **Best-Performing Capturer** players. Thus, the model of Best-Performing Gameplay Stats has the formula: $B = b_1.Best.Artillery + b_2.Best.Stealth + b_3.Overall.Capturer$. Principal components that included the defense and support-related data variables (e.g., iPointDefenses, iBuildingsBuilt, iNumTeleports) did not match our inclusion criteria and were omitted. This implies that our model distinguishes between players based on their exhibited competency across offensive-driven roles based on players' best-ever performances within the game.

**Customization** From the results of PCA, we obtained 2 principal components that matched our inclusion criteria. PC1 described players with **Practicality-driven** (high-value weapons equipped) customization behaviors, while PC2 described players with **Aesthetic-driven** (high-value accessories equipped) customization behaviors. Thus, the model of Player Customization has the formula: $C = c_1.Practicality + c_2.Aesthetics$. This implies that our model distinguishes between players based on their preference for practicality (e.g., higher attack) or aesthetic in customizing their avatars based on their monetary values.

## Model Validation

We validated our identity model by predicting the values of the 9 outcome variables of a players' social networking profile. Given player $i$ and outcome variable $j$, the model has the formula: $O_i^j = A_i^j + B_i^j + C_i^j + k^j$. For each outcome $j$, different coefficients $(a_1^j, a_2^j, a_3^j, a_4^j, b_1^j, b_2^j, b_3^j, b_4^j)$ are obtained using MLR. Our model achieved an explained variance of $>= 30\%$ for the outcome variables: (1) number of friends, (2) number of uploaded screenshots, and (3) number of uploaded videos. Next, we analyze each outcome variable's model and discuss our findings.

## Model Validation Findings

Here, we have focused on concise exposition of main results. We have included the numerical details for the MLR results of the outcome variable Number of Friends in Table 3 of the Appendix as a reference for the analysis process.

**1. Veteran players with high customization have higher number of friends** A player is viewed as a *veteran* through exhibiting high competency through high overall performance at playing the combative and support roles, and best performances with stealth roles (perceived to be difficult to play for beginners). From the significant factors identified, we observe that veteran players had larger friend network sizes. The model had an explained variance (Adj. $R^2$) of $35.1\%$ ($F[7, 201] = 17.1, p < .000$) and a standard error of 73.6. Table 3 of the Appendix shows these results.

**2. Offensive-driven players upload more screenshots** The significant factors identified are similar to those from the previous result. The main differences are 1) players demonstrated best performances with the artillery roles (high kills, assists, damage) and 2) customization preferences did not have any significant effects. A reason could be that artillery roles usually place players in more spectacular scenarios (e.g., explosions, kills, damage), thus increasing the likelihood of players capturing screenshots with them. The model had an explained variance (Adj. $R^2$) of $49.6\%$ ($F[5, 203] = 41.9, p < .000$) and a standard error of 201.2.

**3. Stealth or support-driven players upload more videos** The significant factors identified differ greatly from the previous two results. Only high overall support role and best performing stealth players contributed significantly to the model. A reason could be that videos require more effort to create and are used for scenarios involving action sequences (as opposed to a one-off screenshot), thus being more suited for supporting role or stealth players (e.g., showing how a Spy disguises as the enemy to backstab the opponent.) The model had an explained variance (Adj. $R^2$) of $39.2\%$ ($F[4, 204] = 34.5, p < .000$) and a standard error of 9.5.

## Model Comparison

In Table 1, we compare the MLR results of our identity model on the three outcome variables from the previous section against the two control experiments. We observed that:

**1. Customization preferences are significant factors in predicting number of friends** Recall control experiment #1's model consists of only gameplay-related variables. Despite having four times the number of data variables before backward-selection (36 vs. 9) and more than twice the number of significant ($p < .01$) variables after backward-selection (12 vs. 5) compared to our model, its explained variance was only marginally better (37.6% vs 35.1%.)

**2. The trade-off in having a lower explained variance is warranted by having less data variables that reveal more about the players** Due to the significantly larger number of data variables used in both control experiments, it is expected that they would yield higher explained variances. However, the maximum difference in explained variance is about 17.3%, and we feel that this is a good trade-off because of these reasons: 1) Our model reduced the required data variables by $\geq 50\%$ in all cases, 2) the principal components of our model allowed us to understand the relationship between the original set of data variables to better represent the roles of players within the game, and 3) in some cases (e.g., number of friends) our model had a comparable explained variance than a model with more data variables.

## Player Survey Evaluation

In this section, we analyze selected responses by the players from the survey, with the aim of gaining insight and seeking to understand more about our constructed models.

### Survey Findings

**1. Players consider both aesthetics and practicality when customizing avatars expression of identities** Players were asked to rate the importance of each of the 8 loadout customization slots for expressing their identity on a 5-point Likert scale. A majority of players selected the *Head* slot as the most important, followed by the *Primary Weapon* slot, the *Misc. Item* slots. We performed factor analysis (FA) on the data to discover underlying latent variables. The results are shown in Table 4 of the Appendix. By considering loading values $> 0.7$ and factors with eigen-values $>= 1.0$, we ended up with two latent factors F1 (*Head* and *Misc.* slots) and F2 (*Primary* and *Secondary*), corresponding closely to the principal components obtained via PCA with our model construction. The factors were significant ($\chi^2$[3, 209]=2.29, $p = .514$) with a cumulative variance of 70.5%.

Players were asked what helped them to determine what to equip in each slot. Their responses reaffirmed our findings above. **Aesthetics** was the highest-selected factor for the *Head* and *Misc.* slots while **Practicality** was the highest-selected factor for *Weapons*, which corresponds to our model's resultant principal components obtained for customization. Some responses were "*distinguishing hat/misc slots is a tad unimportant. They're all aesthetic slots, and don't give stats*" and "*change the question about changing customization when playing with different people from having hats and miscs as separate options to just having an option for cosmetics.*" It is worth noting that *Valve* recently introduced an update relabeling the *Head* slot as a *Misc.* slot.

**2. Players differentiated between favorite and best-performing classes** Players were asked to select their *favorite* and *best-performing* classes. Only 52.7% of the players had the same class selected for both, indicating that a majority of the players did not necessarily equate a best-performing class with one that they possibly spent more time with. Assuming that favoring a class means spending more time playing it, this could explain why differentiating between accumulated gameplay and best-performing stats proved effective in our constructed identity models.

**3. Computational methods like PCA reveal more representative player roles than pre-determined categories** From the resulting principal components from PCA, our model demonstrated that gameplay statistics were categorized into factors that described more fine-grained player roles than those originally defined defined by *Valve*. This demonstrates that using PCA to discover latent factors helps to identify emergent player roles that better represent the game. Such AI approaches can thus be seen as a way to avoid the lack of nuance or errors in categorizing roles that might be caused by using predetermined categories, like "support" in this case. Player survey responses reflected this, as some remarked that the default roles defined by *Valve* did not apply well for the game. For example, one player responded that "*a person who likes Sniper (which you classify as a Support class) is unlikely to also like Medic (another Support class)*" and another responded that "*the class roles (offense, defense and support) really doesnt server any purpose. Demo, for example, is one of the best offensive classes even though hes listed as a defensive class.*"

## Discussion

In this section, we discuss several implications of our findings and cover ways that this work might be extended. First, we observed that using PCA on game telemetry data is effective in discovering underlying identity and behavioral patterns exhibited by the players. In addition to accumulated gameplay statistics, best-performing statistics were useful in constructing a more complete player model. This was shown by our models using both accumulated and best-performing principal components and from player survey responses. One notable aspect was that almost all significant principal components had coefficients with the same sign. We suspect that this is due to *Steam Web API*'s data variables not possessing negative characteristics (e.g., deaths, games lost). We plan to find other ways to obtain more data variables.

Second, we demonstrated that both practical and aesthetic factors are considered by players in customizing their characters. In predicting players' number of friends, both customization factors were significant contributing factors. We showed that monetary values of equipped customization items are effective computational measures of customization preferences. However, despite the similarities obtained by both PCA of monetary values and FA of survey responses in highlighting the duality of customization preferences, preliminary work into correlating the players' responses with their customization factors did not show significant effects. We aim to use follow-up surveys to investigate this further.

Third, we showed that aspects of a player's social identity can be modeled using in-game behavior and customization preferences. Players' social identity data variables were obtained from their *Steam Community Profile* page. A data-driven approach is beneficial over a self-reported survey as it minimizes occurrences of response bias. As *Steam* is a gaming-oriented social network, there might be effects of other biases (e.g., friends on the network are different from real-world friends). Extending this with information from another social network (e.g., *Facebook*) may provide greater insight into constructing the players social identity model.

Fourth, we highlight the need for designers and players to be aware of how data is collected, handled, and disseminated due to the close relationship that exists between different ypes of data (achievement, customization, social) that can computationally model a player's identity, behavior, and preferences. More applications are integrating both real-world social information with virtual systems, and data mining makes information easily obtainable. Such identity models enable designers to serve their players better by understanding how and what data should be accessible or restricted for protecting players' privacy.

## Conclusion

In this paper, we developed a computational identity model of players with game telemetry data. Using players' gameplay statistics and customization preferences, our model predicted aspects of players' identity expressed through social networking profiles. It significantly explained the variances of players' number of friends (35.1%), number of uploaded screenshots (49.6%), and number of uploaded videos (39.2%). We found that 1) veteran players with high customization have more friends, 2) offensive-driven players upload more screenshots, and 3) stealth/support-driven players upload more videos.

Player survey responses reaffirmed several of these findings: 1) Players distinguish between aesthetics and practicality in avatar customization, 2) players differentiated between favorite and best-performing classes, and 3) computational approaches can reveal more gameplay-representative player roles as compared to using pre-determined categories. Our model provides insight into how player identities are closely inter-related between different platforms (e.g., a social network and a game) despite each enabling different goals and motivations. It highlights that consideration of players' privacy concerns and needs that developers should take note of when integrating such platforms together. We aim to further extend our findings with further research studies and by applying our model to other videogames and social networks.

## Acknowledgment

This material is based upon work supported by the National Science Foundation under Grant No. 1064495.

## Appendix

| Outcome | Model | #Vars | # Sig. Vars | Adj. $R^2$ |
|---|---|---|---|---|
| Friends | Ours | 9 | 5 | 35.1% |
| | Control #1 | 36 | 11 | 37.6% |
| | Control #2 | 43 | 14 | 52.1% |
| Screenshots | Ours | 9 | 4 | 49.6% |
| | Control #1 | 36 | 12 | 64.9% |
| | Control #2 | 43 | 16 | 66.9% |
| Videos | Ours | 9 | 2 | 39.2% |
| | Control #1 | 36 | 5 | 48.2% |
| | Control #2 | 43 | 7 | 49.9% |

Table 1: Comparative MLR results of our constructed identity model against two control experiments. (All $p < 0.001$).

| | PC1 | PC2 | PC3 | PC4 |
|---|---|---|---|---|
| acc.iBackstabs | 0.401 | 0.056 | * 0.839 | -0.010 |
| acc.iBuildingsBuilt | 0.214 | * 0.803 | 0.227 | -0.007 |
| acc.iBuildingsDestroyed | 0.663 | 0.574 | 0.379 | -0.003 |
| acc.iDamageDealt | * 0.890 | 0.377 | 0.191 | -0.004 |
| acc.iDominations | * 0.813 | 0.382 | 0.252 | -0.008 |
| acc.iFireDamage | 0.485 | 0.381 | 0.288 | -0.032 |
| acc.iHeadshots | 0.657 | -0.129 | 0.188 | -0.002 |
| acc.iHPHealed | 0.204 | * 0.834 | -0.192 | -0.013 |
| acc.iHPLeached | 0.361 | 0.168 | * 0.835 | -0.009 |
| acc.iKillAssists | 0.600 | * 0.769 | 0.003 | -0.011 |
| acc.iNumInvulnerable | 0.189 | 0.341 | -0.343 | -0.012 |
| acc.iNumTeleports | 0.155 | * 0.906 | 0.129 | 0.008 |
| acc.iNumberOfKills | * 0.907 | 0.293 | 0.248 | -0.007 |
| acc.iPlayTime | * 0.807 | 0.437 | 0.148 | 0.008 |
| acc.iPointCaptures | -0.021 | -0.013 | -0.008 | * 1.000 |
| acc.iPointDefenses | * 0.738 | 0.466 | 0.292 | -0.000 |
| acc.iPointsScored | -0.002 | -0.002 | -0.003 | * 1.000 |
| acc.iRevenge | * 0.808 | 0.189 | -0.112 | -0.028 |
| Eigen Values | 9.220 | 2.028 | 1.999 | 1.159 |

Table 2: PCA on data set: **Accumulated Gameplay Stats** (* variable satisfies the inclusion criteria $>= 0.7$)

| | | $\beta$ | SE | $t$ |
|---|---|---|---|---|
| $a_1$ | Overall.Combat †** | 24.384 | 5.978 | 4.079 |
| $a_2$ | Overall.Support †** | 14.538 | 5.393 | 2.696 |
| $a_3$ | Overall.Stealth | -6.102 | 5.149 | -1.185 |
| $a_4$ | Overall.Capturer † | 100.868 | 63.652 | 1.585 |
| $b_1$ | Best.Artillery | 3.751 | 5.154 | 0.728 |
| $b_2$ | Best.Stealth †** | 18.155 | 5.153 | 3.523 |
| $b_3$ | Best.Capturer † | -102.146 | 63.754 | -1.602 |
| $c_1$ | Practicality †** | 17.040 | 5.336 | 3.193 |
| $c_2$ | Aesthetic †** | 30.035 | 5.189 | 5.788 |
| $k$ | (Intercept) | 90.014 | 5.090 | 17.683 |

Table 3: MLR results for outcome: **Number of Friends** († chosen after backward-selection, ** $p < .001$, * $p < .01$)

| | F1 | F2 | F3 |
|---|---|---|---|
| Primary.Weapon.Rating | -0.003 | * 0.814 | -0.045 |
| Secondary.Weapon.Rating | 0.009 | * 0.845 | 0.116 |
| Melee.Weapon.Rating | 0.093 | 0.521 | 0.266 |
| Head.Item.Rating | * 0.728 | 0.076 | 0.228 |
| Misc1.Item.Rating | * 0.988 | 0.030 | 0.137 |
| Misc2.Item.Rating | * 0.924 | 0.012 | 0.126 |
| Action.Item.Rating | 0.300 | 0.166 | * 0.790 |
| Eigen Values | 2.996 | 1.988 | 0.761 |
| Prop. of Variance | 0.35 | 0.24 | 0.11 |
| Cumulative Variance | 0.35 | 0.59 | 0.71 |

Table 4: FA on survey responses on the importance of customization slots for identity expression. (* loading $>= 0.7$)

## References


Bauckhage, C.; Kersting, K.; Sifa, R.; Thurau, C.; Drachen, A.; and Canossa, A. 2012. How players lose interest in playing a game: An empirical study based on distributions of total playing times. In *Computational Intelligence and Games (CIG), 2012 IEEE Conference on*, 139–146. IEEE.

Becker, R.; Chernihov, Y.; Shavitt, Y.; and Zilberman, N. 2012. An analysis of the steam community network evolution. *2012 IEEE 27th Convention of Electrical and Electronics Engineers in Israel* 1–5.

Canossa, A.; Martinez, J. B.; and Togelius, J. 2013. Give Me a Reason to Dig Minecraft and Psychology of Motivation. *In Proceedings of IEEE Conference on Computational Intelligence and Games*.

Drachen, A.; Sifa, R.; Thurau, C.; and Bauckhage, C. 2013. A comparison of methods for player clustering via behavioral telemetry. *In Proceedings of the 2013 Conference on Foundations in Digital Games*.

Harrell, D. F. 2010. Towards a theory of critical computing: The case of social identity representation in digital media applications. CTheory.

Lim, C.-U., and Harrell, D. F. 2013. Modeling player preferences in avatar customization using social network data. In *Proceedings of the IEEE Conference on Computational Intelligence and Games*, 153–160.

Manninen, T., and Kujanpää, T. 2007. The value of virtual assets–the role of game characters in mmogs. *International Journal of Business Science and Applied Management* 2(1):21–33.

Mellon, L. 2009. Applying metrics driven development to mmo costs and risks. *Versant Corporation*.

Steam Stats. 2013. http://store.steampowered.com/stats/.

Steam. http://www.valvesoftware.com.

Team Fortress 2. http://www.teamfortress.com.

Tekofsky, S.; Spronck, P.; Goudbeek, M.; and Broersen, J. 2013. Towards a player age model. In *Ninth Artificial Intelligence and Interactive Digital Entertainment Conference*.

Tekofsky, S.; Spronck, P.; and Broersen, J. 2013. PsyOps : Personality Assessment Through Gaming Behavior.

Yee, N., and Bailenson, J. 2007. The Proteus Effect: The Effect of Transformed Self-Representation on Behavior. *Human Communication Research* 33(3):271–290.

Yee, N.; Ducheneaut, N.; Nelson, L.; and Likarish, P. 2011. Introverted Elves & Conscientious Gnomes : The Expression of Personality in World of Warcraft. *In Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems* 753–762.

Yee, N.; Ducheneaut, N.; Shiao, H.-T.; and Nelson, L. 2012. Through the azerothian looking glass: mapping in-game preferences to real world demographics. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, 2811–2814. ACM.

Yee, N. 2006. Motivations for play in online games. *Cyberpsychology & behavior : the impact of the Internet, multimedia and virtual reality on behavior and society* 9(6).