The Use of Social Science Methods to Predict Player Characteristics from Avatar Observations

Carl Symborski, Gary M. Jackson, Meg Barton, Geoffrey Cranmer, Byron Raines and Mary Magee Quinn

Abstract The purpose of this study was to investigate the extent to which real world characteristics of massively multiplayer online role-playing game (MMORPG) players can be predicted based on the characteristics and behavior of their avatars. Ground truth on participants' real world characteristics was obtained through the administration of validated measures of personality and authoritarian ideology, as well as a demographics form. A team of trained assessors used quantitative assessment instruments to evaluate avatar characteristics, behavior, and personality from a recorded session of the participant's typical gameplay. The statistical technique of discriminant analysis was then applied to create predictive models for players' real world characteristics such as gender, approximate age, and education level, using the variables generated through observational assessment of the avatar.

This work was supported by the Air Force Research Laboratory (AFRL). The US Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of AFRL or the US Government.

C. Symborski (✉) · G. M. Jackson · M. Barton · G. Cranmer · B. Raines · M. M. Quinn
Leidos, 4001 N, Fairfax Dr., Arlington, VA 22203, USA
e-mail: carl.w.symborski@saic.com

G. M. Jackson
e-mail: gary.m.jackson@leidos.com

M. Barton
e-mail: marguerite.r.barton@leidos.com

G. Cranmer
e-mail: geoffrey.cranmer@leidos.com

B. Raines
e-mail: byron.a.raines@leidos.com

M. M. Quinn
e-mail: mary.m.quinn@leidos.com

M. A. Ahmad et al. (eds.), *Predicting Real World Behaviors from Virtual World Data*, Springer Proceedings in Complexity, DOI 10.1007/978-3-319-07142-8_2,

1 Introduction

The prediction of an individual's real world (RW) characteristics, purely based on observations of his or her avatar's characteristics and behaviors in a massively multiplayer online role-playing game (MMORPG), is a challenging yet fascinating proposition. The complex relationship between players and their avatars is not yet fully understood. While some players prefer to approximate their RW personas and appearances in virtual worlds (VWs), or even present more attractive, idealized versions of themselves [1], [2], others might opt to present a role-playing identity in-world [3]. This makes the determination of relationships between the avatar and his/her operator a unique challenge. A player may be presenting physical or behavioral characteristics consistent with his/her RW persona, or his/her avatar appearance and behaviors may display characteristics entirely different from that persona.

With this in mind, a study was designed to investigate whether it is possible to predict RW characteristics of MMORPG players from the characteristics and behavior of their avatars. Other studies, such as [4], have sought to accomplish the same end by parsing the massive amount of player data logs stored in MMORPG databases and tying the results back to measured RW player characteristics. This study sought to use more traditional modes of social science research and statistical techniques to develop predictive models, with avatar characteristics and behavior as the predictor variables and RW player characteristics as the criterion, or dependent, variables. This facilitated the creation of relatively simple algebraic equations that could be used by a trained observer to make predictions about an individual after a period of observing his/her avatar.

Using quantitative predictor/independent variables (IVs) generated through the observational collection of avatar data and criterion/dependent variables (DVs) collected through the administration of several ground truth instruments filled out by participants, the statistical technique of discriminant analysis (DA) was then employed to produce predictive models of RW characteristics using VW observations. In this study, the quantitative variables gathered were used to generate statistical models for the prediction of RW gender, age, education level, extraversion level, and submissive ideology.

2 Method

2.1 Participants

Participants were recruited using online advertisements, flyers, and online game forums. All participants were required to be at least 18 years of age and to have a minimum of 50 h of experience with either *Guild Wars*® [5] or *Aion*® [6], two popular MMORPGs. A total of 114 participants completed the study. Recruitment occurred primarily within the Washington DC region, USA.

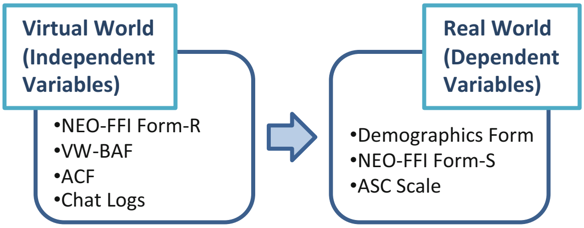

Fig. 1 Sources of real world dependent variables (DVs) and virtual world independent variables (IVs)

More *Guild Wars*® (73, or 64 %) than *Aion*® players (41, or 36 %) and more males (71, or 62 %) than females (43, or 38 %) participated in the study. The mean participant age was 31.7 years' old, with a standard deviation of 9.5 years. A slight majority of participants had less than a 4-year degree (51, or 45 %), while 46 participants held a Bachelor's degree (40 %), and 17 participants (15 %) held a graduate degree.

2.2 Instruments

An important facet of the research design was the distinction between the DVs to be predicted, or the selected RW characteristics—gender, age, education level, extraversion level, and submissive ideology—and the IVs to be used as predictors for the selected DVs, or the characteristics and behavior of the players' avatars. Several instruments were used to capture these data (see Fig. 1).

2.2.1 Data Collection Instruments—RW Characteristics

The RW characteristics of the participants, or the DVs to be predicted, were recorded using three forms. The **Demographics Form** was used to collect basic demographic information on the participants, such as gender, age, and highest education level achieved. Each participant also completed the validated **Aggression-Submission-Conventionalism Scale (ASC scale)**, which has three subscales—submissive, aggressive, and conventional—as a measure of ideology [7].

Finally, each participant completed the **NEO Five-Factor Inventory (NEO-FFI),** a standardized and validated personality assessment instrument constructed from the five-factor model of personality [8]. The instrument measures the "Big Five" traits of neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness. The NEO-FFI was selected for use because it has two forms: the NEO-FFI Form-S (NEO-S), a self-report form, and the NEO-FFI Form-R (NEO-R), an observer rating form; the two forms are identical, except that the NEO-S is written from a first person perspective and the NEO-R is written from a third person perspective [8]. This enabled the use of the NEO-S to gather ground truth data on the participants' RW personalities, and the use of the NEO-R to gather objective assessments of the avatar's personality, made by an independent rater (see below).

2.2.2 Data Processing Instruments—Avatar Characteristics and Behavior

Multiple mechanisms for collecting IV data on the characteristics and behaviors of the participants' avatars were employed. The NEO-R was used to gather information on avatar personality, the Virtual Worlds Behavior Analysis Form (VW-BAF) and Avatar Characteristics Form (ACF) were developed to collect data on avatar behaviors and characteristics, and chat logs were analyzed for linguistic data.

While the NEO-S was completed by the participants, the **NEO-R** (observer rating form) [8] was used to assess avatar personality. The NEO-R was completed by trained assessors observing the participant's avatar, facilitated by a "bridge" developed to resolve inconsistencies or conflicts of any NEO-R item that caused rating difficulty within the VW, considering that the original form had been developed to rate humans in the RW.

The **VW-BAF** was created to measure avatar activities and behavior in MMORPGs. The form has two pages. The first allows a rater to tally occurrences of behaviors on a minute-by-minute basis (e.g., number of emotes, number of times avatar dies, number of occasions avatar helps another avatar). The second page allows a rater to record presence or absence of major activity categories (e.g., quest activity, social activity) during the interval. The second page of the form also measures how much time an avatar spends in a group versus solo, and whether the mode was Player versus Player (PvP) or Player versus Environment (PvE).

The **ACF** was developed to capture the following general characteristics of avatars in the VW: gender, physical characteristics, class/role, armor/weapons, style of play (e.g., member of a guild), pets, and general play style (e.g., social player, strategic player).

A central feature of the instrument development was that all forms—the NEO Bridge, the VW-BAF, and the ACF—were designed to be generalizable across all MMORPGs. This is important because MMORPGS have a wide variety of different naming conventions for similar things (e.g., a healer class may be called a "monk" in one game and a "chanter" in another). Thus, the adoption of a standard vernacular was an essential component of the form development process. For example, regardless of game nomenclature, all character classes were assigned to one of the following four categories based on their primary characteristics: warrior, scout, priest, and mage.

Additionally, the research team transcribed participant **chat logs**. Care was taken only to record chat by the target participant/avatar. These chat logs were run through a parser, which extracted variables such as number of words and average word length.

2.3 Design and Procedure

The basic research design included three primary phases which will be described here: the initial laboratory session with the participant to establish ground truth, the gathering/processing of quantitative data, and the development of predictive statistical models using DA.

2.3.1 Laboratory/Home Session

Upon arriving at the laboratory, each participant viewed a standard introductory video describing the study and completed the informed consent process, if he/she desired to participate. The participant then completed the demographics form, the NEO-S, and the ASC scale. These collected data formed the basis for ground truth on the participant's RW characteristics.

At the participant's leisure after returning home, the participant recorded a 1-h gameplay session of the selected game using a screen capture program. Participants were requested to choose one avatar and to use only that avatar for the entire 1-h home session, and to play the game just as they normally would.

Once the home session was completed, the participant saved the recorded session on a USB flash drive and mailed it back to the laboratory. Participants were asked to complete the home session within 2 weeks of the laboratory session. Once the flash drive was received, participants were compensated for their contribution to the study.

2.3.2 Quantitative Data Processing

Upon receipt of the flash drive by the laboratory, trained assessors rated that session using the NEO-R, the VW-BAF, and the ACF. Separately, chat was transcribed. A minimum of 80 % inter-rater agreement was obtained for all rating forms.

For completion of the NEO-R, raters observed the entire hour of the recorded session and then provided avatar ratings using the NEO Bridge to help answer specific items, such as "He often feels tense and jittery."* For this item, the NEO Bridge specified that, if the avatar exhibited spastic movement, swiveled the camera frequently, ran in circles or rapidly manipulated the user interface, the item was coded as "agree." However, if the avatar typically appeared relaxed and calm, the item was coded as "disagree."

The VW-BAF rating occurred across three 10-min samples, which were standardized in terms of start times across the full 1-h session to ensure a representative sample of avatar behavior. Observations were recorded once a minute. As described previously, the VW-BAF has two pages: the first is a series of counts for how many times particular behaviors occurred each minute, and the second recorded whether or not certain behaviors occurred within each minute. Both kinds of variable were summed over each 10-min sample, and then averaged across the three 10-min samples for each participant before being used for analysis. Hence, each VW-BAF variable either corresponds to how many times a behavior occurred in an average minute, or to how many minutes out of 10 included that behavior on average.

At the same time that the VW-BAF ratings were recorded, the ACF was completed. The ACF items consist of presence or absence variables; for example, presence or absence of a costume or of a ranged damage-per-second combat role.

2.3.3 Statistical Technique: DA

DA was used to generate predictive models for participants' RW characteristics based on their avatars' characteristics and behavior. The purpose of DA is to predict group membership, based on RW DVs such as gender or extraversion, from a linear combination of VW IVs, such as avatar gender or type of armor worn. DA begins with a data set containing many cases (participants), where both the values of the IVs and the group membership (DVs) are known. The end result is an equation or set of equations that predict(s) group membership for new cases where only the values of the IVs are known [9].

Specifically, DA using *backward* stepwise reduction was conducted. This form of stepwise reduction begins with a given set of variables and reduces the set by eliminating the variables that are associated with the DV to a lesser degree than remaining variables [10]. In this way, only the variables that are most predictive of the DV remain as part of the predictive model when the analysis is complete. As an additional quality control, leaving-one-out cross-validation was used for all stepwise reduction DA runs. This form of validation accuracy is the process by which a model is trained on all cases but one and tested on the one case that was withheld. The process repeats until all cases have been withheld and tested blindly, eliminating any chance of predicting a case based on extracted knowledge of that case [10].

3 Results and Discussion

3.1 The Discriminant Function

As described above, DA generates a predictive model in the form of a linear combination of independent, or predictor, variables. There is also a constant term for each equation, which is used as the linear offset in the discriminant functions. The general form of the discriminant function is as follows:

$$DV\ Group\ 0 = (b_1 \times x_1) + (b_2 \times x_2) + \ldots + (b_\mathrm{n} \times x_\mathrm{n}) + c,$$
$$DV\ Group\ 1 = (b_1 \times x_1) + (b_2 \times x_2) + \ldots + (b_\mathrm{n} \times x_\mathrm{n}) + c,$$

where,

$b_n =$ the Fisher coefficient for that variable,
$x_n =$ the value of the independent variable, and
$c =$ the value of the constant.

Once the values of the variables are substituted in the above equations for the names of the variables, whichever of the two equations evaluates to the greater number will be the prediction for that participant. In other words, whichever of the two Fisher's discriminant functions produces a higher value "wins," and the participant will be predicted as a member of the corresponding category [11].

Table 1 Overall accuracy for predicting RW characteristics from VW observations

RW characteristic	Overall accuracy (%)	Precision (%)	Recall (%)
Gender	83	93	79
Age	70	80	56
Education Level	66	68	71
Extraversion Level	68	64	63
Submissive Ideology	65	66	66

RW real world, *VW* virtual world

3.2 Accuracy Metrics

Several accuracy metrics are reported for each of the predictive models. The first is the leaving-one-out cross-validated overall accuracy for each model. It is the best representation of the DA model because it captures information about how well the model classifies cases into both the target and non-target groups. Precision, representing the proportion of those cases predicted to be in the target group that were actually in that group, and recall, denoting the proportion of cases in the target group that were correctly predicted to be in the target group, are also presented to provide information about the accuracy of the models.

3.3 Overall Results

Table 1 presents the accuracy results of the developed models. Individual results are presented in the following sections. These sections explain each model along with a description of how the DV was defined, an overview of the accuracy of the model, and a brief discussion of the IVs relevant to the prediction of the target DV. Some of these discussions of IVs have more obvious support in the literature and appear to be more clear-cut than others. For the cases in which the relationship between the predictor IV and the DV being predicted was not clear, the subject matter experts of the research team developed potential explanations for the relationship based on their knowledge of MMORPGs.

3.4 Gender

3.4.1 Definition of the DV

For the DV of gender, males were assigned as one group (71 participants) and females were assigned as the other group (43 participants).

Table 2 Accuracy of gender model

Overall accuracy (%)	Precision (%)	Recall (%)
83	93	79

3.4.2 Accuracy of Gender Model

The gender model obtained 83 % overall accuracy (see Table 2). Precision was high at 93 % and recall was 79 %.

3.4.3 Discriminant Function for Gender

In determining which group a new case should be classified into, the values for each of the IVs were plugged into both of the following two equations. Whichever equation yielded the largest value represents the group that the new case was classified into. The following are the discriminant functions derived for gender:

$$\begin{aligned} Male = {} & (4.800 \times MaleAV) + (4.924 \times MajRoleSupport) \\ & + (2.429 \times HairAccNA) - 2.646, \end{aligned} \tag{1}$$

$$\begin{aligned} Female = {} & (0.578 \times MaleAV) + (0.695 \times MajRoleSupport) \\ & + (0.081 \times HairAccNA) - 0.721, \end{aligned} \tag{2}$$

Table 3 presents descriptions of the predictor, or independent, variables in the discriminant functions.

Through interpretation of the discriminant functions above, the models can be described loosely as follows:

If the avatar is male, and/or has a majority combat role of support, and/or has covered hair, then it is likely that the player's RW gender is male.

Otherwise, it is likely that the RW gender is female.

NOTE: These written descriptions are not to be substituted for the above equations in making predictions; they are only intended to aid the reader in understanding the statistical models.

3.4.4 Discussion of IVs Relevant to the Prediction of Gender

Avatar Gender (ACF)

It is well-known that avatar gender is closely related to RW gender [12]; therefore, it is not surprising that avatar gender surfaced as a predictor variable in this analysis. In our study worlds, the appearance of a male avatar strongly predicted a RW gender of male. Female avatars required the support of additional IVs to distinguish between female avatars operated by RW females and female avatars operated by RW males, because "gender-bending" is more common among male players (who play female avatars) than among female players [12].

Table 3 Description of IVs relevant to prediction of gender

Variable	Description of variable
MaleAV	Avatar is male
MajRoleSupport	Majority combat role is support (e.g., healing, buffing)
HairAccNA	Hair accessories cannot be observed, likely because hair is covered (e.g., by a helmet or costume)

IVs independent variables

Majority Role of Support (ACF)

This variable indicated whether or not an avatar's majority combat role was "support," a combat style characterized by healing, providing enhancement buffs, or otherwise supporting others in battle (50 % of the time or greater). It is commonly assumed that females gravitate toward healing roles in MMORPGs, since women are considered to be more nurturing and supportive by nature [13]. Our research suggests the opposite; the presence of this theme was a predictor of RW male gender. Because male avatar gender is such a strong predictor of RW male gender, the remaining variables in the discriminant function above primarily indicate the remaining RW males who were "gender-bending." Yee et al. [14] conducted a study that offers an explanation as to why the theme of healing others in combination with the presence of a female avatar might predict RW males: they discovered that "[male] players enact this stereotype [of women as healers] when gender-bending." Hence, a male player creating an avatar for a healing role might tend to choose a female.

Covered Hair (ACF)

The variable "Hair Accessories N/A" indicated if an avatar's hair accessories could not be observed because the avatar's hair was covered, due to a helmet or costume that occluded the hair (e.g., a hood). The model indicated that RW male players playing female characters in the sample were less likely to expose their hair and hair accessories than RW female players were. While there does not appear to be a ready explanation for this phenomenon in the literature, the research team postulated that RW male players operating female avatars might be less concerned with displaying their hair than RW female players, instead being more focused on gameplay activities, while RW female players may have more interest in ensuring that their avatars have a feminine appearance, complete with displayed flowing locks.

3.5 Age

3.5.1 Definition of the DV

The DV of age was divided into two groups: under the age of 30 and 30 and over. This is consistent with demographic research [15] and qualitative observations indicating

Table 4 Accuracy of age model

Overall accuracy (%)	Precision (%)	Recall (%)
70	80	56

that MMORPG players tend to be younger than social game players. Furthermore, the age group of 18–29 years is one that is frequently cited in research studies, particularly in voting-related and medical contexts, and thus seemed a logical breakdown to use. Using this division, 59 participants were age 30 years or over and 55 participants were under age 30.

3.5.2 Accuracy of Age Model

The age model achieved 70 % overall accuracy (see Table 4). Though precision was high at 80 %, recall was low at 56 %, suggesting that this model may need to be further refined.

3.5.3 Discriminant Function for Age

The following are the two equations derived for age:

$$30\ \mathit{or\ Older} = (1.206 \times \mathit{BAF35}) + (2.789 \times \mathit{UnconvHair}) + (2.245 \times \mathit{NEOR21}) - 2.814, \tag{3}$$

$$\mathit{Under}\ 30 = (0.397 \times \mathit{BAF35}) + (0.815 \times \mathit{UnconvHair}) + (2.905 \times \mathit{NEOR21}) - 3.008. \tag{4}$$

Table 5 presents descriptions of the variables used.

Simply stated, this model becomes:

If the avatar spends more time stationary, and/or has unconventionally colored hair, and/or does not appear tense and jittery, then it is likely that the player's age is 30 or over.

Otherwise, it is likely that the player's age is under 30.

3.5.4 Discussion of IVs Relevant to the Prediction of Age

Avatar does not move for the full 60 s (VW-BAF)

This item measured how many full minutes an avatar spent stationary over the observation period. In the study sample, an avatar that spent more time stationary was more likely to be older (age 30 or over). Younger people tend to be more active than older people, and perhaps more prone to bouts of fidgeting and nervous activity. In

Springer
http://www.springer.com/978-3-319-07141-1

Predicting Real World Behaviors from Virtual World Data
Ahmad, M.A.; Shen, C.; Srivastava, J.; Contractor, N. (Eds.)
2014, XIV, 118 p. 40 illus., 27 illus. in color., Hardcover
ISBN: 978-3-319-07141-1