Introverted Elves & Conscientious Gnomes: The Expression of Personality in World of Warcraft

Nick Yee[1], Nicolas Ducheneaut[1], Les Nelson[1], Peter Likarish[2]

[1]Palo Alto Research Center
3333 Coyote Hill Road, Palo Alto, CA
[nyee, nicolas, lnelson]@parc.com

[2]University of Iowa
Iowa City, IA
peter-likarish@uiowa.edu

ABSTRACT
Personality inference can be used for dynamic personalization of content or system customization. In this study, we examined whether and how personality is expressed in Virtual Worlds (VWs). Survey data from 1,040 *World of Warcraft* players containing demographic and personality variables was paired with their VW behavioral metrics over a four-month period. Many behavioral cues in VWs were found to be related to personality. For example, Extraverts prefer group activities over solo activities. We also found that these behavioral indicators can be used to infer a player's personality.

Author Keywords
Virtual worlds, online games, personality, Big 5, inference.

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Human Factors

INTRODUCTION

Games can be character revealing. One of the author's fathers once noted that he enjoys playing golf with his business partners because it lets him see which of them cheats on the golf course. The underlying implication, of course, is that how someone behaves on a golf course says something about how they may behave during a business transaction. And online gamers who have developed romantic relationships in virtual worlds [34] often say something similar:

"The game WAS the reason we fell in love. Going through all the adventures and quests together really built our relationship. We found out how the other person is when they are mad, tired, sad, happy, excited, annoyed, etc." [City of Villains, Female, Age 25]

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.

The unique affordances of virtual worlds offer an unparalleled platform for examining the intersections between personality and behaviors in virtual environments. On the other hand, unlike personality expression in physical settings, online games allow, or even encourage, users to behave in a manner inconsistent with their everyday identities. Thus, in this study, we ask:

- Is it true that a person's personality can be inferred from how they behave in a virtual world?
- And if so, what specific virtual cues are highly indicative of a person's introversion or conscientiousness (for example)?

Being able to infer a user's personality from online cues has direct relevance to HCI research, given the field's long-standing interest in interface personalization and system customization [e.g., 16, 24]. Indeed, knowing more about a user's personality could help design systems more responsive to users' needs in areas as diverse as e-commerce, social software, and recommender systems, to name a few.

In this paper, we use data from the widely popular massively multiplayer online game (MMOG), *World of Warcraft* (WoW), to answer the two questions above. We then use our results to discuss how personality data could be used in the design of future online systems, being mindful of some important limitations and potential pitfalls also suggested by our research.

The Expression of Personality

Studies in personality psychology have repeatedly shown that judgments of personality at zero acquaintance (i.e., by strangers) are moderately accurate. More importantly, the specific cues used to infer different personality traits are consensually shared. In other words, personality is readily expressed in specific cues in everyday life. This has been shown to be true for brief face-to-face encounters [9, 15]. For example, in an earlier study involving video-taped face-to-face conversations [9], Extraverted individuals spoke louder, with more enthusiasm and energy, and were more expressive with gestures.

Other studies have researched personality inference by examining an individual's bedroom or office [12], or their

music collection [23]. For example, in the study of personal spaces, Conscientious individuals had well-lit, neat, and well-organized bedrooms. And individuals who scored high on Openness to Experience had more varied books and magazines.

This line of research has also extended to computer-mediated communication (CMC). In particular, studies have shown that moderately accurate personality impressions can be formed based on an individual's personal website [18, 28], Facebook profile [3], email content [10], blog content [32], and even an individual's email address—the thinnest slice of CMC possible [2]. For example, in terms of linguistic output on blogs, Agreeable individuals were more likely to use the first person singular, words related to family, and words related to positive emotions (e.g., happy, joy). Conscientious individuals were more likely to use words related to achievement.

These studies illustrate that we leave behind personality traces in both the physical and digital spaces that we inhabit. Given that the average online gamer spends over 20 hours a week in a virtual world [31, 33], it is not difficult to imagine that some amount of personality traces could be gleaned from logs of their virtual interactions as well.

Limits to Personality Expression?

On the other hand, there are also reasons to believe that personality may not be readily expressed in virtual worlds. First of all, previous studies have largely focused on personality expression in everyday settings or linguistic output online. It is unclear how or whether personality is expressed via non-human bodies doing non-human things in a fantasy world (e.g., gnomish priests resurrecting the dead with magical light rays).

Related to this point, some scholars, like Turkle, have suggested that VWs allow us to constantly reinvent ourselves [27]. If the strongest interpretation of this notion were true, it would imply that there might be a clean break between personality and how a person behaves in a VW. In other words, people could express or reinvent themselves idiosyncratically in VWs, such that shared cues of personality would not exist. And finally, there is also evidence that users do alter their behaviors in online games. For example, studies have shown that role-players tend to be more imaginative and thus willing to experiment with their online personas [5, 26]. And studies of online dating [13] and online gaming [4] have shown that users in both settings tend to idealize their online personas to some degree. In particular, some studies have revealed that tendencies to idealize self-representation online are moderated by poor self-esteem [4, 7]. Thus, identity experimentation and individual variations in that experimentation may suppress stable personality expression cues in virtual worlds.

The Collection of Personality Data

Previous studies of personality expression have tended to rely on linguistic output or behavioral traces. These traces are often artifacts of behaviors over time. For example, a person who is low on Conscientiousness may often forget to water their plant. A withered plant in a bedroom is an example of a behavioral trace. Of course, as some researchers have suggested [20], we should also study actual behaviors as they occur. These researchers argue that observations of individuals in their natural settings and "humdrum lives" (pg. 862) may yield a better understanding of the link between personality and behavior.

The problem is that the recording of behaviors in natural settings and the subsequent coding are daunting tasks using traditional tools. Shadowing and video recording individuals is a laborious method that significantly constrains sample size. Recent technology has begun to offset the daunting nature of behavioral data collection, however. For example, in a study of how personality is related to everyday linguistic output [20], researchers used an electronically activated recorder which was programmed to record a participant's acoustic space for 30 seconds every 12 minutes. A dictionary-based software tool was then used to generate quantitative linguistic metrics of these recordings.

Behavioral Data Collection in Virtual Worlds

Virtual Worlds (VWs) offer unique affordances for studying the link between personality and behavior. For the purposes of this paper, we define VWs as graphical environments that enable geographically-distant individuals to interact via avatars (i.e., digital representations of users). It is also important to note that VWs are no longer academic prototypes or niche cultures, but have become mainstream interaction platforms. For example, *WoW* has over 11 million active monthly subscribers [30], and the *Facebook* game *FarmVille* has over 80 million active monthly users.

VWs offer three unique features in terms of collection of natural behavioral data. First, unlike the physical world where it would be unfeasible to follow everyone around with video cameras, VWs come inherently instrumented. The computer systems running the VWs already track the movement and behavior of every avatar to make interactions possible (e.g., orienting avatars so that they can look at each other). Second, these high-precision sensors operate at all times. Thus, it is possible to generate not only snapshot data, but longitudinal behavior profiles for every user in a particular VW (e.g., see [8]). And finally, all these observations can be performed unobtrusively, thereby significantly reducing the observer effect [29]—participants cannot react to the camera if the camera is invisible.

Indeed, a recent study has illustrated that there are connections between personality and virtual behaviors in the VW *Second Life* [35]. In that study, 76 students were asked to participate in *Second Life* for six weeks while

"wearing" a scripted virtual tracking device that captured some of their behaviors and linguistic output. The findings revealed some interesting correlations. For example, high Conscientiousness was positively correlated with geographical movement.

There were several weaknesses in that study, however. First and foremost, it is difficult to capture natural behavior by assigning users to participate in a VW not of their own choosing. Being able to observe actual users would likely yield more reliable data. Second, only data from one VW was collected. Given that much of SL resembles suburban America [1], it would be helpful to gather data from additional VWs (and in particular fantasy-based online games) to see if the results generalize. Third, participants in that study were only asked to spend six hours each week in *Second Life*. On the other hand, we know that players of other VWs spend on average 20 hours a week (without being asked to) in games like WoW [33]. In other words, the participant sample may not be representative of VW users in either demographics or usage patterns. And finally, many of the correlations found in that study did not align with trait definitions of the personality variables used—e.g., virtual behaviors that correlated with Agreeableness were not related in obvious ways to Agreeableness. Thus, a replication in a different VW with existing users may help clarify whether the results are an artifact of the nature of *Second Life* or how people behave in VWs in general.

Research Questions

We focus on two research questions in this paper. While previous studies have examined personality expression in everyday settings, we were interested in examining whether and how personality is expressed in online games. To clarify and expand upon previous findings of personality correlates in VWs, we focus on the online game *World of Warcraft* in this paper with a sample of active players. Our first research question is thus:

RQ1. What are the behavioral correlates of personality in an online game?

If indeed personality is expressed in consistent cues in VWs, a pertinent question is whether these cues can be used specifically for personality inference. Thus, our second research question is:

RQ2. How well can we infer someone's personality from only observing their virtual behaviors?

METHOD

Given our focus on the online game WoW, we will begin by first briefly describing the game context to lay the foundation for understanding the variables we use as behavioral indicators.

World of Warcraft

WoW is currently one of the more popular online games available commercially [30]. Unlike *Second Life* (SL) where users create most of the in-world content (including buildings, clothing, hair styles, and avatar bodies), content in WoW is almost entirely created and designed by the company running the game. And unlike the open sandbox nature of SL, WoW uses a typical "leveling up" formula seen in computer role-playing games. Specifically, players start at level 1 and kill monsters to become higher level and acquire better weapons and armor in order to kill bigger monsters. Along the way, the game encourages players in different ways to collaborate with other players. Users can also create characters with different skill sets that complement each other. For example, heavily-armored *tank* classes shield the group from enemy attacks while lightly-armored damage dealing *DPS* (damage per second) classes deal damage to enemies and *healing* classes restore health lost in combat. In short, WoW is a collaborative virtual environment [22].

WoW draws from an established lore from the Warcraft franchise. Briefly, players must choose to belong to one of two primary factions—the Alliance or the Horde. Each faction has five distinct races, e.g., Night Elves or Trolls. A variety of rules dictate where and when players may attack and kill each other. Thus, a distinction is made between PvP (player-vs-player) activities and PvE (player-vs-environment) activities. PvP activites can range from one-to-one duels to large 40 vs. 40 battlegrounds (BGs). And in general, it is a player's choice as to how much PvP activity they want to engage in.

Players in WoW communicate via typed chat and might also use VoIP tools to communicate via speech. The game also provides a modest set of emotes (e.g., /hug). Players are also able to specialize in crafting professions and convert collected raw ingredients into finished goods, such as in tailoring or cooking.

There is also a system of Achievements that keeps track of a wide variety of combat and non-combat based objectives. There are Achievements for zones explored, for dungeons completed, for number of hugs given, and for cooking proficiency. These Achievement scores provide a good sense of how a player chooses to spend their time in WoW.

Thus, overall, WoW offers a wide and varied set of rich behavioral cues to draw from. From class choice to amount of PvP activity, from number of emotes used to amount of world exploration, the game context offers a range of measurable behaviors. This is also a point of differentiation from SL. Due to the open nature of SL, most higher-level conceptual behaviors are not defined in the environment and it is up to individual users to define their creations. Thus, there is no overarching set of metrics beyond fairly low level behaviors, whereas in WoW, the game keeps track of many behaviors and activities using a standardized lexicon.

The World of Warcraft Armory

Indeed, the standardized lexicon and data format inherently lends itself to automated data collection. Blizzard, the developer of WoW, is unique in that they have provided public access to much of their internally-collected data at a

website known as the Armory. In short, by searching for a character's name, anyone can view details about their past activities, including how many hugs they have given, the quality of their equipment, the class they prefer to play, etc. More importantly, these metrics have been tracked since the character was first created. With a few clicks, we can gather a character profile that has cumulative data over many months of game play. It bears emphasizing the tremendous social science research opportunities that are made possible by this publicly-available database of longitudinal behavioral metrics. It is from the Armory that we gathered the behavioral metrics for this study.

Participants

1,040 WoW players were included in the study. We recruited participants from forums dedicated to WoW, publicity on popular gaming sites (e.g., WoW.com), word-of-mouth on social media like Twitter, and mailing lists from previous studies of online gamers. We note that due to human subjects regulations, minors were excluded from participating in the study. Nevertheless, we were still able to gather data from a very wide age range (18-65). The average age of our sample was 27.03 (SD = 8.21). 26% of participants were women.

Procedure

Participants began by completing a web-based survey that gathered their demographic and personality information. Participants were also asked to list up to 6 WoW characters they were actively playing. Once these characters were in our database, an automated data collection system was activated. The system launches a web scraper that gathers character profiles (large XML files) from the WoW Armory. The Armory updates itself once per day (in the early morning) if a character has been active the previous day. Thus, our script follows this schedule with a daily interval and collects any updated profiles. For the results, we analyzed data from a contiguous 4-month period in the spring and summer of 2010.

Personality Measures

In personality psychology, the Big-5 model is the gold standard. The model measures five traits: Extraversion, Agreeableness, Conscientiousness, Emotional Stability, and Openness to Experience.

For comparability, we also used an inventory that measured these 5 factors. A 20-item scale measuring the Big-Five Factor structure was drawn from the International Personality Item Pool [11]. Participants rated themselves on the inventory items using a scale that ranged from 1 (Very Inaccurate) to 5 (Very Accurate).

Behavioral Measures in WoW

There are two main complexities we encountered when dealing with the Armory data. First, Armory profiles consist of hundreds of variables, oftentimes in a hierarchy. For example, there is a system of Achievements in WoW that tracks progress in a variety of defined goals, such as Exploration Achievements and Dungeon Achievements. Under Exploration Achievements, there is a category for each continent. Under each continent, there is a listing for each zone. To avoid being inundated by low-level variables or including overlapping variables, we adopted an analytic strategy of looking at or generating high level variables where possible. This in turn produces more stable variables that map to psychologically meaningful concepts. For example, a notion of geographical exploration would seem to be better tracked by the overall count of zones explored rather than looking at any one particular zone.

A second complexity is that most players have multiple active characters at the same time and it is not at first clear how to combine metrics across characters to derive participant-level aggregates. For example, a level 80 character can do much more damage than a level 60 character (and the function is non-linear). Thus, there is no way to easily combine damage done across characters. While these metrics needed to be normalized, there wasn't one single variable they could all be normalized against.

We therefore adopted the following normalization and variable generation strategies:

1) Static character attributes were normalized against total number of characters. E.g., ratio of male characters = male characters / total characters.

2) Variable character attributes were normalized against overall time played. E.g., for combat roles, we calculated how often each character was a tank/healer/DPS, and then calculated a participant-level ratio for each of those roles. A 0.24 tank ratio meant that across all of a participant's characters, they spent 0.24 of their total playing time as a tank.

3) Metrics that could be normalized against another variable were normalized accordingly. E.g., the score of Exploration Achievements could be divided by the score of All Achievements to generate an Exploration ratio. This thus filtered out the raw difference between someone with many and someone with few achievements, and focused instead on how they focus their game-play.

4) For metrics that could not be normalized and were highly dependent on character level, we extracted the maximum. E.g., it is very different having one character that has 80 vanity pets compared with having 4 characters with 20 vanity pets each. In these cases, we found the maximum number of vanity pets across a participant's characters.

5) For metrics that could not be normalized and were not dependent on character level, we calculated the sum. E.g., any level character can emote /hug as often as they'd like. In these cases, we summed up the count of hugs across all of their characters.

It is important to mention that we are not claiming to have extracted all possible variables for analysis in this paper, but rather, that we have extracted a meaningful and manageable

subset of higher-level variables that covers a wide range of behaviors in WoW. A description of each derived variable, along with its mean and standard deviation are presented in Table 1 below. Note that we excluded outliers that were more than 2 standard deviations away from the mean when deriving these metrics. For brevity we will only describe high-level trends in the text, but for ease of reference, we will include the table row index in round brackets after each mentioned correlate.

#	Variable	Description	M (SD)	E	A	C	ES	O
1	Ratio of Alliance Characters	= Alliance Chars / Total Chars	0.53 (0.47)	0.00	0.05	**0.07**	0.04	0.02
2	Ratio of Opposite Gender Characters	= Opposite Gender Chars / Total Chars	0.27 (0.36)	**-0.07**	**-0.14**	-0.03	**0.07**	0.00
3	Total Character Count	Count of all active characters reported by participant	2.79 (1.51)	**-0.12**	0.03	**0.07**	0.02	**0.10**
4	Number of Days Played Since Start of Study	Count of unique active days since start of study	65.47 (34.89)	-0.04	0.00	0.01	-0.03	-0.01
5	Total Realm Count	Count of realms participant has active characters on	1.11 (0.31)	-0.05	0.06	0.01	-0.03	**0.09**
6	Max of Guild Changes	Highest number of guild change events	.78 (1.05)	**0.07**	-0.01	0.00	-0.05	0.03
7	Sum of Kills	Includes both kills against computer monsters and other players	162353.84 (108633.20)	-0.03	-0.03	0.00	-0.00	**-0.07**
8	Sum of Kills in BGs	Number of kills in battlegrounds	2705.70 (3589.28)	-0.01	**-0.06**	0.00	0.04	0.00
9	Sum of PvP Kills	Number of all PvP-related kills	10437.22 (12026.80)	-0.04	**-0.08**	0.05	**0.09**	-0.05
10	Sum of Deaths	Total number of deaths from any cause	1849.12 (1440.63)	0.05	**-0.07**	0.00	**0.05**	-0.04
11	Sum of Deaths in Raid Dungeons	Number of deaths in dungeons	1018.84 (899.94)	**0.06**	**-0.07**	0.01	0.02	**-0.08**
12	Sum of Deaths from Falling	Number of deaths from falling from high places	32.64 (69.10)	0.02	-0.05	**-0.07**	-0.02	0.00
13	Sum of Hugs	Number of /hug emote	38.57 (69.10)	-0.02	**0.11**	**0.10**	-0.03	**0.09**
14	Sum of LOLs	Number of /lol emote	63.73 (147.57)	0.01	0.01	-0.03	-0.02	0.05
15	Sum of Cheers	Number of /cheer emote	47.05 (90.40)	**-0.09**	**0.13**	**0.07**	0.04	**0.13**
16	Sum of Waves	Number of /wave emote	79.77 (140.21)	-0.06	**0.10**	**0.09**	**0.08**	**0.14**
17	Max Number of Mounts	Mounts increase travel speed and are both functional and collectible	32.08 (29.62)	-0.05	0.03	0.05	-0.02	0.01
18	Max Number of Vanity Pets	Vanity pets are small non-functional and largely decorative companions	39.45 (31.80)	**-0.07**	**0.07**	**0.08**	-0.05	**0.07**
19	Ratio of Need Rolls	= Need Rolls / Total Rolls	0.17 (0.11)	**0.10**	**-0.14**	**-0.08**	-0.06	**-0.09**
20	Max Equipment Score	Sum of all equipment item levels	3867.90 (813.20)	0.02	**-0.10**	-0.04	0.01	-0.06
21	Sum of Count of Respecs	Number of times player has changed skill specializations	27.02 (28.06)	0.03	**-0.09**	-0.02	0.03	-0.05
22	Max of Achievement Score	Total Achievement score	413.06 (195.44)	-0.01	-0.04	0.02	0.00	-0.03
23	Ratio of Quest Achievements	= Quest Achs / Total Achs (based on Sums)	.07 (.02)	**-0.10**	**0.07**	0.02	0.01	-0.01

Table Continued

#	Variable	Description	M (SD)	E	A	C	ES	O
24	Ratio of Exploration Achievements	= Exploration Achs / Total Achs (based on Sums)	.10 (.05)	-0.04	**0.09**	0.06	0.02	**0.13**
25	Ratio of PvP Achievements	= PvP Achs / Total Achs (based on Sums)	.10 (.05)	0.00	**-0.12**	-0.03	**0.07**	-0.01
26	Ratio of Dungeon Achievements	= Dungeons Achs / Total Achs (based on Sums)	.36 (.12)	**0.12**	**-0.17**	**-0.12**	0.01	**-0.17**
27	Ratio of Profession Achievements	= Profession Achs / Total Achs (based on Sums)	.10 (.06)	-0.04	**0.13**	**0.07**	-0.02	**0.12**
28	Ratio of Reputation Achievements	= Reputation Achs / Total Achs (based on Sums)	.03 (.01)	-0.03	-0.02	-0.03	-0.01	**-0.12**
29	Ratio of World Event Achievements	= World Achs / Total Achs (based on Sums)	.13 (.07)	**-0.08**	**0.16**	**0.10**	-0.04	**0.11**
30	Max of Cooking Achievements	Highest cooking score	6.33 (4.85)	**-0.07**	**0.07**	**0.07**	-0.01	0.05
31	Max of Fishing Achievements	Highest fishing score	7.26 (6.02)	**-0.06**	**0.07**	**0.07**	0.01	0.05
32	Sum of End Game 10-man Raids Done	Total 10-man end-game raids completed	16.78 (17.83)	0.06	**-0.11**	-0.05	0.00	**-0.13**
33	Sum of End Game 25-man Raids Done	Total 25-man end-game raids completed	18.13 (22.99)	**0.08**	**-0.09**	-0.05	0.00	**-0.12**
34	Ratio of Healing Done	= Healing Done / Damage Done (based on Sums)	.32 (.46)	0.00	0.00	-0.03	-0.02	0.01
35	Sum of Arenas Played	Number of Arenas entered	55.57 (155.31)	-0.01	**-0.09**	0.01	0.06	0.01
36	Sum of BGs Played	Number of BGs entered	98.36 (147.11)	**-0.07**	**-0.07**	0.02	0.05	0.04
37	Sum of Duels Played	Number of Duels entered	52.80 (94.73)	**0.11**	**-0.07**	-0.04	-0.05	-0.03
38	Ratio of Arena Wins	= Arena Wins / Arenas Entered	.33 (.18)	**-0.10**	**-0.12**	0.03	**0.08**	-0.01
39	Ratio of BG Wins	= BG Wins / BGs Entered	.48 (.18)	-0.02	**-0.06**	-0.01	-0.01	0.01
40	Ratio of Duel Wins	= Duel Wins / Duels Entered	.46 (.21)	**-0.06**	0.02	**0.07**	0.04	-0.01
41	Sum of Flight Paths Taken	Flight paths are used to fly from one fixed location to another	1424.42 (1117.06)	**-0.08**	**0.07**	0.05	-0.02	-0.01
42	Sum of Hearths	Hearthstones allow a character to teleport to a pre-determined location	454.08 (310.49)	0.00	0.02	-0.01	-0.03	-0.03
43	Ratio of Melee DPS Role	Ratio of time spent in hand-to-hand DPS role (e.g., fury warriors, rogues)	.30 (.30)	**-0.08**	-0.01	0.03	0.02	0.05
44	Ratio of Ranged DPS Role	Ratio of time spent in ranged DPs role (e.g, hunters, mages)	.38 (.32)	0.06	0.05	0.01	**-0.08**	0.04
45	Ratio of Healing Role	Ratio of time spent in healing role (e.g., holy priests, restoration druids)	.20 (.24)	0.04	-0.05	-0.05	-0.01	-0.05
46	Ratio of Tank Role	Ratio of time spent in tanking role (e.g., protection warrior, protection paladin)	.13 (.20)	-0.01	-0.01	-0.01	**0.11**	-0.07

Table 1. Means, standard deviations, and correlation coefficients of VW behavioral measures. Correlation coefficients in bold are p < .05.

RESULTS

To analyze how personality is expressed in VWs, we examined the correlations between the virtual behaviors and the personality factors. Given the increased risk of experiment-wise error in large correlation tables with 46 variables against the five personality factors, we used an analytic method developed by Sherman and Funder [25] to address this specific issue. The method employs a Monte Carlo simulation of repeatedly randomized data within each participant. Thus, the method preserves the statistical properties of the data gathered. The method creates 1,000 instances of these randomized data sets and tabulates the

number of observed significant correlations (at alpha of .05). The probability of the actual number of significant correlations is then calculated based on where it lies on the distribution of these 1,000 randomizations. In other words, this technique answers whether we found a significantly higher number of significant correlations in our data set than would be expected by chance alone. In our case, using an alpha of .05, we had 83 observed significant correlations where only 11.50 would be expected by chance based on the simulations. According to this Monte Carlo method, the probability of this number of observed correlations is $p < .001$. This provides assurance that the observed correlations, as a whole, are non-random.

We will now describe each of the Big 5 personality factors and the virtual behaviors they were correlated with. We will not discuss every significant correlation, but instead try to find clusters of correlations that trace out the bigger picture.

Extraversion

According to the trait definition, individuals who score high on Extraversion tend to be outgoing, gregarious, and energetic, while those who score low on Extraversion tend to be reserved, shy, and quiet.

In terms of behavioral indicators in VWs, individuals who score high on Extraversion tend to prefer group activities. They have a higher ratio of Dungeon Achievements (26), which requires collaboration with other players. They have also completed a higher number of end-game 25-man raid dungeons (33). Their higher number of guild changes also implies social promiscuity (6).

On the other hand, players who score low on Extraversion prefer solo activities, such as questing (23), cooking (30), and fishing (33). They also are more likely to have more vanity pets (18), which are silent pet-like companions.

We also see that players who score low on Extraversion have a preference and higher win ratios for some PvP activities (36, 37, 38, & 40), but it is less obvious what the connection is. The same is true for the higher ratio of opposite gender characters (2) among those who score low on Extraversion.

Agreeableness

According to the trait definition, individuals who score high on Agreeableness tend to be friendly, caring, and cooperative, while those who score low on Agreeableness tend to be suspicious, antagonistic, and competitive.

In terms of behavioral indicators in VWs, individuals who score high on Agreeableness give out more positive emotes (13. 15, 16), i.e., hugs, cheers, and waves, and prefer non-combat activities such as exploration (24), crafting (13), world events (29), cooking (30), and fishing (31).

On the other hand, players who score low on Agreeableness prefer the more competitive and antagonistic aspects of game-play. They enjoy killing other players (8 & 9). They also have more deaths (10), focus more on getting better equipment (20), and have engaged in more PvP activities (25), including BGs (35), Arenas (36), and duels (37). Their competitive edge also translates to a higher winning ratio in Arenas (38) and BGs (39).

The negative correlation with ratio of need rolls (19) is also telling. Valuable equipment drops from monsters are given to players according to dice rolls. Players select to roll based on "Need" or "Greed", of which the former is given higher priority. We found that players who are low on Agreeableness often insist on being given higher priority over others by rolling "Need". While this is tolerated in some cases, abusing Need rolls is often seen as anti-social (there is even a specific epithet used by the community to describe these players: ninja looters).

Conscientiousness

According to the trait definition, individuals who score high on Conscientiousness are organized, self-disciplined, and dutiful, while those who score low on Conscientiousness are careless, spontaneous, and easy-going.

In terms of behavioral indicators in VWs, individuals who score high on Conscientiousness seem to enjoy disciplined collections in non-combat settings. This is reflected in having a large number of vanity pets (18) which must be collected one at a time, and having high cooking (30) and fishing scores (31) which reflect self-discipline in collecting unique recipes and visiting unique fishing locations (as well as patiently staying put for significant amounts of time in these locations, since fishing in the game is surprisingly close to its real-world equivalent: catches can be few and far between). The same is true for world event achievements (29) which often require disciplined collections of items and visiting a set of locations around the world.

On the other hand, individuals who score low on Conscientiousness seem to be more careless and are more likely to die from falling from high places (12).

Emotional Stability

According to the trait definition, individuals who score high on Emotional Stability are calm, secure, and confident, while those who score low on Emotional Stability are nervous, sensitive, and vulnerable.

While there were significant correlations between behavioral metrics and this personality trait, these correlations were more difficult to interpret as a whole. Individuals who score low on Emotional Stability prefer PvP related activities, including having a higher PvP achievement score (25) and higher wins in the Arena (38). Individuals who score higher on Emotional Stability are more likely to have characters of the opposite gender (2).

It is worth noting that previous studies have also had difficulty identifying meaningful behavioral correlates for

Emotional Stability [12, 17], so our findings here may reflect an overall weaker behavioral expression of this trait.

Openness to Experience

According to the trait definition, individuals who score high on Openness to Experience are abstract thinkers, imaginative, and intellectually curious, while those who score low on Openness to Experience are down-to-earth, conventional, and traditional.

In terms of behavioral indicators in VWs, we see a cluster of correlates that reflect exploration and curiosity. For example, individuals who score higher on Openness have more characters (3). They also have characters on more realms (5), i.e., game servers or parallel worlds that each character resides on. And they spend more of their play-time exploring the world (reflected by the higher exploration achievement ratio, 24). They also spend more time participating in non-combat activities, such as crafting professions (27) and world events (29).

On the other hand, individuals who score low on Openness prefer the more traditional, combat-oriented aspects of game-play, spending more time in dungeons and raids (26, 32, & 33).

Personality Inference from Behavioral Metrics

To examine how well personality can be inferred from virtual behavioral metrics alone, we conducted a series of multiple regressions on each of the personality factors using the respective ten highest behavioral correlates. We note that this method is imperfect and creates a "double-dipping" concern, but provides a rough sense of how well personality can be inferred. The results are shown in Table 2.

All of the multiple regressions were significant at $p < .05$; four were significant at $p < .001$. This suggests that virtual behavioral metrics can be used to provide statistically significant models of a player's personality. According to Cohen [6], an R of .30 is a medium effect size, while an R of .10 is a small effect size. Thus, many of our regression models had around medium effect sizes.

Variable	R	R^2	Adj. R^2	STE	F	p
Extrav.	0.30	0.09	0.07	0.93	4.73	< .001
Agreeable.	0.30	0.09	0.07	0.67	4.67	< .001
Conscient.	0.20	0.04	0.03	0.79	4.86	< .001
Emo. Sta.	0.21	0.04	0.02	0.79	2.13	0.03
Openness	0.26	0.07	0.06	0.75	4.93	< .001

Table 2. Multiple regressions on each of the personality factors.

DISCUSSION

The availability of fine-grained virtual behavioral metrics in the WoW Armory allowed us to gather longitudinal profiles of actual VW users. While studies in the past have examined links between personality and linguistic output online (in emails or blogs), our study is the first to examine the links between personality and virtual behavior in an online game. Our findings reveal that our personalities are expressed in VWs via consistent cues, and that most of these cues reflect trait definitions of standard personality factors. For example, players who score high on Extraversion prefer group-oriented activities. And players who score high on Agreeableness use more positive emotes and prefer non-combat activities. More importantly, our multiple regressions reveal that behavioral cues in VWs can be used to infer an individual's personality. These findings suggest that while some degree of identity experimentation is occurring in virtual worlds, basic personality is still being readily expressed.

While an earlier study of personality expression in VWs [35] had trouble finding trait-aligned behavioral correlates in *Second Life*, we were able to find much more coherent behavioral clusters that were consistent with personality trait definitions in our study. Findings in the earlier study may have been impacted by participants with no prior experience with the VW. Also, it bears pointing out that the WoW Armory allowed us to gather a set of more conceptually meaningful variables. Due to constraints in the scripting language and sandbox nature of *Second Life*, there is no standardized set of high-level behavioral variables that are shared. Thus the earlier study relied on lower-level variables such as distance walked or ratio of time sitting down, which may be less powerful in capturing personality expression, as opposed to behaviors such as hugging someone.

Knowing the specific behavioral correlates for personality expression in virtual worlds is also important for several reasons. First, it helps researchers triage the large number of behavioral variables gathered in future studies, and helps prioritize where to start looking. Second, it helps psychologists understand whether certain personality traits are more easily predicted via behavioral indicators. And finally, comparing the findings across these studies will help us understand whether these behavioral correlates are consistent or idiosyncratic among different virtual worlds.

Implications for CHI

Personalized interfaces and system customization have long been of interest to the HCI community [16, 24]. It is reasonable to assume that information needs vary based on a user's personality – for instance, extroverts using an online shopping website might be more interested in other customers' reviews, while introverts might prefer seeing mostly technical data about the product instead. Our paper points at the possibility of inferring users' personalities based on their activity traces (which need not come from online games) and customizing their experience based on the results.

Another possibility directly applicable to online games but also other forms of social software would be to use inferred personality information to assist in the formation of groups,

perhaps by recommending compatible partners based on the task to be accomplished. For instance, groups requiring a diversity of opinions might benefit from the inclusion of a wide range of personality types [14]. In other contexts, a more homogeneous mix could be beneficial. And it is worth pointing out that we are not suggesting an automated system that would kick some players out of groups because they are low on Agreeableness. After all, the competitive nature of these players can be an asset in PvP settings, and an assertive nature can also be an asset for raid leaders. In a related fashion, personality data could also be used in recommender systems: recommendations from other users with similar personality profiles could be given more or less weight, depending on the user's desire for more homogeneity or diversity in the options they are presented with [19].

Limitations and Future Research Directions

There were several limitations to our study. First, we only collected data from one VW. It is unclear whether the behavioral cues we identified generalize to other similar online games. Moreover, it is difficult to say how our indicators translate to VWs that do not employ the dragon-slaying role-playing paradigm. Nevertheless, our findings hint at potential metrics to collect and analyze in future studies. For example, emotes (for Agreeableness) or geographical movement (for Openness) have analogous metrics across many types of VWs.

A related limitation to generalizability is that WoW users are highly-engaged users who spend on average 20 weeks producing behaviorally-rich metrics. This usage profile is likely atypical of normal website or mobile app usage. Whether the more typical casual engagement with websites and mobile apps would allow personality inference is certainly an avenue for future research.

Third, while the correlation coefficients appear to be quite low (ranging from .06-.17), a similar large-scale study (i.e., >500 participants) of linguistic output among bloggers yielded similar effect sizes [32]. Given the larger variances in demographics (with an age range of 16-85 in our sample) and unavoidable noise among natural setting samples, these smaller effect sizes are probably not surprising in hindsight.

And finally, we relied on the set of variables that Blizzard shares publicly via the Armory. It is possible that other unshared variables, such as logged chat, may be even more predictive of personality. Given the existing work on linguistic predictors of personality [21, 32], it would be interesting to be able to directly compare the predictive power of linguistic and behavioral cues.

Overall, it is important to continue exploring how personality is expressed across a range of VWs (using a variety of metrics) to understand how generalizable these findings are.

Ending Thoughts

VWs provide a novel research platform with unique affordances and challenges. The automated longitudinal data collection across a wide range of behaviors is impossible to mirror using traditional data collection techniques, and similar techniques could also be used to study other social phenomena, such as the emergence of group norms or leadership.

On the other hand, VWs come with unique challenges as well. Above all, the ability to create tracking systems that essentially shadow a user wherever they go in a VW raises privacy concerns. In our study, the consent process spelled out the data collection scripts to participants, but given that VWs like WoW are a kind of pseudonymous public space, data collection studies (without a survey component like ours) largely fall into the exempt category for human subjects Institutional Review Board (IRB) review. The gray area arises due to the fact that the public space of WoW is unlike any physical public space we know--with microphones and video cameras that could follow every user unobtrusively.

This becomes even more complicated when the game developer makes public what would otherwise be private data. Such is the case with the WoW Armory. After all, before the WoW Armory, players could make the case that they had a reasonable expectation of privacy in WoW (with regard to IRB review). This expectation is no longer reasonable with the release of the Armory. In short, VWs create new research platforms, but at the same time, force us to address our role as researchers in the face of such powerful data collection tools.

It is easy to imagine that VWs allow us to become whatever we want to be, but our findings show that our personalities remain even when we don virtual bodies. These findings of personality expression in VWs suggest that our first lives still play an important role even when we are in *Second Life*. And our personalities are readily expressed even when we are Elves and Gnomes.

ACKNOWLEDGMENTS

This research is sponsored by the Air Force Research Laboratory.

REFERENCES

1. Au, W. Linden Suburban Home Owners More Likely To Treat Their Place As Extension of the Real Life Self, Academic Suggests. *New World Notes*. http://nwn.blogs.com/nwn/2010/04/linden-homes-study.html
2. Back, M., Schukle, S. and Egloff, B. How extraverted is honey.bunny77@hotmail.de? Inferring personality from e-mail addresses. *Journal of Research in Personality*, *42* (2008), 1116-1122.
3. Back, M., Stopfer, J., Vazire, S., Gaddis, S., Schmukle, S., Egloff, B. and Gosling, S. Facebook profiles reflect

actual personality not self-idealization. *Psychological Science*, *21* (2010), 372-374.
4. Bessiere, K., Seay, A. and Kiesler, S. The Ideal Elf: Identity Exploration in World of Warcraft. *CyberPsychology and Behavior*, *10* (2007), 530-535.
5. Carroll, J. and Carolin, P. Relationship between game playing and personality. *Psychological Reports*, *64* (1989), 705-706.
6. Cohen, J. *Statistical Power Analysis*. Lawrence Erlbaum Associates, 1988.
7. Ducheneaut, N., Wen, M., Yee, N. and Wadley, G. Body and mind: a study of avatar personalization in three virtual worlds. *Proceedings of CHI*, *1* (2009), 1151-1160.
8. Ducheneaut, N., Yee, N., Nickell, E. and Moore, R. The life and death of online gaming communities: a look at guilds in World of Warcraft. *CHI 2007 Proceedings* (2007), 839-848.
9. Funder, D. and Sneed, C. Behavioral Manifestations of Personality: An Ecological Approach to Judgmental Accuracy. *Journal of Personality and Social Psychology*, *64* (1993), 479-490.
10. Gill, A., Oberlander, J. and Austin, E. Rating E-mail Personality at Zero Acquaintance. *Personality and Individual Differences*, *40* (2006), 497-507.
11. Goldberg, L. A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. in Mervielde, I., Deary, I., De Fruyt, F. and Ostendorf, F. eds. *Personality Psychology in Europe*, Tilburg University Press, Tilburg, The Netherlands, 1999, 7-28.
12. Gosling, S., Ko, S., Mannarelli, T. and Morris, M. A Room with a cue: Judgments of personality based on offices and bedrooms. *Journal of Personality and Social Psychology*, *82* (2002), 379-398.
13. Hancock, J., Toma, C. and Ellison, N. The truth about lying in online dating profiles. *Proceedings of CHI 2007*, *1* (2007), 449-452.
14. Harper, F., Frankowski, D., Drenner, S., Ren, Y.Q., Kiesler, S., Terveen, L., Kraut, R. and Riedl, J. Talk Amongst Yourselves: Inviting Users to Participate in Online Conversations. *IUI 2007* (2007), 62-71.
15. Kenny, D., Horner, C., Kashy, D. and Chu, L. Consensus at zero acquaintance: Replication, behavioral cues, and stability. *Journal of Personality and Social Psychology*, *62* (1992), 88-97.
16. Mackay, W. Triggers and Barriers to Customizing Software. *Proceedings of SIGCHI 1991*, *1* (1991), 153-160.
17. Mairesse, F. and Walker, M. Automatic Recognition of Personality in Conversation. *Proceedings of the Human Language Technology Conference*, *1* (2006), 85-88.
18. Marcus, B., Machilek, F. and Schutz, A. Personality in Cyberspace: Personal Web Sites as Media for Personality Expressions and Impressiosn. *Journal of Personality and Social Psychology*, *90* (2006), 1014-1031.
19. McNee, S., Riedl, J. and Konstan, J. Making Recommendations Better: An Analytic Model for Human-Recommender Interaction. *CHI 2006* (2006), 1103-1108.
20. Mehl, M., Gosling, S. and Pennebaker, J. Personality in Its Natural Habitat: Manifestations and Implicit Folk Theories of Personality in Daily Life. *Journal of Personality and Social Psychology*, *90* (2006), 862-877.
21. Mehl, M. and Pennebaker, J. The Sounds of Social Life: A Psychometric Analysis of Students' Daily Social Environment and Natural Conversations. *Journal of Personality and Social Psychology*, *84* (2003), 857-870.
22. Nardi, B. and Harris, J. Strangers and Friends: Collaborative Play in World of Warcraft. *CSCW 2006* (2006), 149-158.
23. Rentfrow, P. and Gosling, S. Message in a Ballad: The Role of Music Preferences in Interpersonal Perception. *Psychological Science*, *17* (2006), 236-242.
24. Riecken, D. Personalized Views of Personalization. *Communications of the ACM*, *43* (2000), 26-28.
25. Sherman, R. and Funder, D. Evaluating correlations in studies of personality and behavior: Beyond the number of significant findings to be expected by chance. *Journal of Research in Personality*, *43* (2009), 1053-1063.
26. Simon, A. Emotional stability pertaining to the game of Dungeons and Dragons. *Psychology in the Schools*, *24* (1987), 329-332.
27. Turkle, S. *Life on the Screen: Identity in the Age of the Internet*. New York: Simon and Schuster., 1995.
28. Vazire, S. and Gosling, S. e-Perceptions: Personality Impressions Based on Personal Websites. *Journal of Personality and Social Psychology*, *87* (2004), 123-132.
29. Webb, E., Campbell, D., Schwartz, R. and Sechrest, L. *Unobtrusive measures: non-reactive research in the social sciences*. Rand McNalley, Chicago, 1966.
30. White, P. MMOGData: Charts. *2009* (2008).
31. Williams, D., Yee, N. and Caplan, S. Who plays, how much, and why? Debunking the stereotypical gamer profile. *Journal of Computer-Mediated Communication*, *13* (2008), 993-1018.
32. Yarkoni, T. Personality in 100,000 Words: A large scale analysis of personality and word use among bloggers. *Journal of Research in Personality* (in press).
33. Yee, N. The demographics, motivations, and derived experiences of users of massively multi-user online graphical environments. *Presence: Teleoperators and Virtual Environments*, *15* (2006), 309-329.
34. Yee, N. The "Impossible" Romance. *The Daedalus Project*. http://www.nickyee.com/daedalus/archives/001534.php
35. Yee, N., Harris, H., Jabon, M. and Bailenson, J. The Expression of Personality in Virtual Worlds. *Social Psychological & Personality Science* (in press).