Revealing Social Identity Phenomena in Videogames with Archetypal Analysis

Chong-U Lim[1] and **D. Fox Harrell**[2]

Abstract. In this paper, we present a novel approach toward revealing social identity phenomena in videogames using archetypal analysis (AA). Conventionally used as a dimensionality reduction technique for multivariate data, we demonstrate how AA can reveal social phenomena and inequity such as gender/race-related steretoyping and marginalization in videogame designs. We analyze characters and default attribute distributions of two critically acclaimed and commercially successful videogames (*The Elder Scrolls IV: Oblivion* and *Ultima IV*) together with 190 characters created by players in a user-study using a third system of our own design. We show that AA can computationally 1) reveal implicit categorization of characters in videogames (e.g., base player roles and hybrid roles), 2) model real world racial stereotypes and stigma using character attributes (e.g., physically dominant attributes for *Oblivion*'s ostensibly African-American "Redguard" race) and 3) model gender marginalization and bias (e.g., males characterized as more archetypal representations of each race than females across attributes.) We highlight how AA is an effective approach for computationally modeling identity representations and how it provides a systematic way for the critical assessment of social identity phenomena in videogames.

1 INTRODUCTION

Videogames often construct virtual environments and worlds that are populated with virtual characters. Both these worlds and characters may be represented in a multitude of ways. Graphical 2-dimensional (2D) or 3-dimensional (3D) assets grant visual appearances, textual descriptions provide intriguing narrative, backstories, and characteristics, while numerical statistical attributes provide quantifiable measurements defining character skills and capabilities for a variety of interactions, from dealing damage against a mighty adversary, to charming a non-playable character into handing over an elusive item.

Though often considered to be purely virtual, these representations are in fact blended real/virtual identities that are both affected by, and capable of influencing, aspects of real world identities. Even in the case of a fairly rudimentary character such as Pac-Man, in action we have a blend of a real users control with a 2D animated sprite. Recent studies have shown how representations of race and gender within videogames have deep social implications [8]. In the commercially successful and critically acclaimed role-playing game (RPG) *The Elder Scrolls IV: Oblivion*, some character designs "implement and amplify many disempowering social identity constructions" [9]. "Females of some races are more intelligent than their male counterparts and individuals of the ostensibly French 'race' (Bretons) are twenty points more intelligent than their ostensibly Norwegian (Nords) counterparts, regardless of gender" [9]. It highlights the *importance of the underlying implementations and data structures used to construct these representations*. If developed without due consideration, undesirable social implications related to identity such as marginalization and stereotyping may be further perpetuated. Research has shown that peoples' performances are impacted by stereotypes [20] and behaviors in the physical worlds are altered by their avatar use [22].

However, it is important to recognize that these issues are not simply technical in nature. We adopt a *critical computing* [9] approach, using algorithmic processing and data structuring for critically assessing and providing commentary about the real world and related social phenomena. In this paper, we demonstrate an how archetypal analysis can be used as an Artificial Intelligence (AI) tool for such critical assessments of computational identity-related social phenomena in two commercial videogames, as well as a character creation system of our own design. The upshot is that we found that AA is a robust method for computationally modeling underlying social identity phenomena grounded in cognitive science. We use AA to model social phenomena within games, such as male characters being favored over female characters based on statistical attribute distributions or in-game races having real world stereotypes imparted upon them (e.g., ostensibly African-American "Redguard" characters in *Oblivion* given better physical but lower mental stats.) To the best of our knowledge, this application of cognitive science (apart from some notable exceptions like Santa Ana's work on discrimination and racism [17] and Lakoff's work on political affiliations [12]) and AI has not often been applied to analyze nuances of identity. Most computational systems, like videogames, are built with classical models and categories explicitly built into software. Hence, they provide a good venue to critically assess such cognitively grounded AI approaches to studying digital identity.

2 BACKGROUND

In this section, we present the theoretical framework for our work and provide an overview of the videogames used for our analysis.

2.1 Cognitive Categorization and the Sociology of Classification

Our view of categorization is based upon cognitive scientist George Lakoff's work in cognitive categorization [11] termed *category gradience* and psychologist Eleanor Rosch's *protoypes* [16]. As opposed to outmoded classical or "folk" approaches, which character-

[1] Computer Science & Artificial Intelligence Laboratory, Massachusetts Institute of Technology, USA, email: culim@mit.edu

[2] Computer Science & Artificial Intelligence Laboratory, Comparative Media Studies Program, Massachusetts Institute of Technology, USA, email: fox.harrell@mit.edu

ize category membership to be defined by a fixed set of characteristics, centrality gradience recognizes that some members are typically deemed "better examples" of a category than others. Extending upon this, we use the following concepts from the sociology of classification by Geoffrey Bowker and Susan Leigh Star [3] for describing categorization-related social phenomena. **Membership** is the experience of encountering and interacting with objects within certain social groups, and increasingly engaging in naturalized relationships with them. **Naturalization** is the deepening familiarity of such interactions within a given social group. **Marginalization** is a result of enforced naturalization occurring where members of a marginal category exist outside of social groups, or are less prototypical members of communities. It is also characterized by exclusion from a social group or an individual having *multiple memberships* and often refers to *exclusion or difference from normative behaviors* (Stigma) [7, 9]. **Markedness** indicates that, unlike normative categories, marginal categories are demarcated visually and linguistically.

To reconcile these concepts with the systems in this paper, we use a cognitively-grounded model for critically assessing computing systems for social analysis [9]. It suggests that category gradience enables semantic relations to be structured or ranked according to how constitutive they are of the category. Naturalization may be assessed by *user actions and attributes* that reinforce category semantics, resulting in a higher degree of membership. Marginalization may be implemented through enabling *degrees of membership* and represented as being *further away from the prototypes*. Normative groups that are often unnamed and unmarked may possess *implicitly assumed* normative privileges that may be identified and modeled. This theoretical framework forms the basis for using archetypal analysis as an approach for social analysis and empowerment through critically assessing the statistical attributes of characters within videogames for revealing implicitly-derived social phenomena such as gender-related marginalization and stereotyping.

2.2 Archetypal Analysis

Archetypal Analysis (AA), introduced by Cutler and Breiman [5], is a method for reducing the dimensionality of multivariate data [1]. Given a set of multivariate data points, the aim of AA is to be able to represent each data point as a *convex combination* of a set of key data points called **archetypes**. For example, applying AA on a dataset of basketball players and their statistics [6] computationally revealed and represented the following four archetypes – "benchwarmer," "rebounder," "three-point shooter," and "offensive." Every individual player in the entire data set could then be represented as a hybrid mixture of these archetypes [18]. Formally, given a data set of points $\{x_1, x_2, ..., x_n\}$, AA seeks to find a set of archetypes $\{z_1, z_2, ..., z_k\}$, where $z_j = \sum_{i=1}^{n} \beta_{ij} x_i$, and enables each data point x_i to be represented in terms of the k archetypes as $x_i = \sum_{j=1}^{k} \alpha_{ji} z_j$. The objective function minimizes the residual sum of squares $RSS = ||x_i - \sum_{j=1}^{k} \beta_{ij} z_j||^2$ under the constraints that the weights $\sum \beta_{ij} = 1$ $\beta_{ij} \geq 0$ and coefficients $\sum \alpha ji = 1$ $\alpha_{ji} \geq 0$. These ensure the archetypes *meaningfully resemble* and are *convex mixtures* of the data. These archetypes are located on the data convex hull [5] and are represented are combinations of individual points, making them more easily interpretable [1], unlike other dimensionality reduction techniques like Principal Component Analysis [10] and Non-negative Matrix Factorization [13]. AA has been shown to be effective compared to other techniques for various AI-related problems. Compared to other recommender models (nearest neighbor, two popularity, random baseline) AA provided the highest recall rates for archetypal recommender systems [19] in games, demonstrating robustness for finding relevant recommendations. Here, AA is an appropriate approach given our aim to computationally model individuals that are more "prototypical" than others (archetypes) and being able to measure the "centrality gradience" of each individual with respect to these archetypes. As described in Section 2.1, we believe that such models would enable us to begin critically assess social phenomena such as marginalization and stereotyping computationally.

2.3 Overview of Videogames

We provide an overview of the two commercially successful videogames used in this paper. Both are important open-world single-player RPGs with strong customization. *Ultima IV: Quest of the Avatar* is arguably the most influential game on the open world RPG genre and *The Elder Scrolls IV: Oblivion* is a stunning recent success with a strong customization system and diversity. Even in excellent games, there is the potential for implicit stereotypes and inequity. Our observations are meant to be useful for improvement in this regard.

The Elder Scrolls IV: Oblivion is the fourth installment of the popular *Elder Scrolls* computer role-playing game series, developed by *Bethesda*. In the lore of the game designed by game designers there are several races, each with their own fictional background stories and histories. Three basic player roles exist in the game – "Fighter", "Mage" and "Thief" [15], which are derived from common roles across most RPGs stemming from old table-top RPGs like Dungeons and Dragons. Each race is associated with the three basic roles in varying degrees (hybrid roles), which compliment the game's lore about its people and races. Players choose to play as one of the ten different races available, customizing characters over 7 basic attributes (strength, intelligence, willpower, agility, speed, endurance, and personality,) together with their height and weight.

Ultima IV: Quest of the Avatar is the fourth installment of the *Ultima* series of role-playing games, and the first in the "Age of Enlightenment" trilogy, *Ultima IV* was first released in 1985 by *Origin Systems*. The player is assigned one of eight classes to play and does not directly choose or assign values to attributes. Instead, the user is posed several questions embedded within the games narrative at the beginning, resulting in the players ranking of eight **virtues** in the game based on the game's three **principles** of Truth, Love, and Courage. There are seven companions that the player may choose to form a party with. Each character has a particular class, each associated with a virtue, and possesses seven numerical attributes (strength, dexterity, intelligence, hit points (HP), magic points (MP), level, and experience,) an armor type, a weapon type, and their gender.

3 APPROACH

1. Analyzing existing systems for designer-centered phenomena. In order to assess the kinds of categorization and social identity phenomena that arise as a result of designer choices (top-down), we applied archetypal analysis to the statistical attribute allocation for new characters in both *Oblivion* and *Ultima IV*. For *Oblivion*, the variables included the races, gender, and eight attributes. For *Ultima IV*, the variables included the character classes and seven attributes.

2. Analyzing emergent phenomena with a system of our own creation. For the purpose of assessing the kinds of categorization and social phenomena that may be *implicitly-derived* from players (bottom-up), we conducted a user-study with 190 players where they constructed avatars in an avatar constructor of our own creation. Players customized both their character's visual appearance and statistical attributes values of six commonly used videogame attributes (strength, endurance, dexterity, intelligence, charisma, and wisdom) on a 7-point Likert scale with a total of 27 allocatable points. The avatar constructor used our avatar game data-mining system called *AIRvatar* [14], that stores each created avatar, the statistical attribute allocations, and textual descriptions made by the players.

3. Determining the number of archetypes During AA, we varied the number of archetypes k in the range $1 \leq k \leq 10$. We adopt the convention of the Cattell scree test [4] for using the residual sum-of-squares (RSS) to determine the optimal number of archetypes by picking the value of k matching the first point of the "elbow" of a screeplot with corresponding to the biggest change in RSS. This balances the trade off between minimizing RSS and overfitting.

4 RESULTS

We present results describing the archetypes obtained from analyzing the statistical attributes of each system using archetypal analysis.

4.1 Oblivion

In *Oblivion* we found $k = 3$ to be optimal. Both Archetypes 2 and 3 were pure archetypes ($\alpha_j = 1$). The ternary plot in Figure 2(a) of the Appendix shows a visualization of the α coefficients of these archetypes. We also observed the following characteristics:

- Archetype 1 had the highest "Strength" and "Endurance", but lowest "Intelligence". Archetype 1 had the biggest "Size".
- Archetype 2 was relatively balanced across the attributes, with highest "Willpower" and "Personality".
- Archetype 3 had highest "Intelligence", "Agility" and "Speed", but lowest "Willpower". Archetype 3 had a relatively small "Size".

4.2 Ultima IV

In *Ultima IV*, we found $k = 3$ to be optimal. All three were pure archetypes. The ternary plot in Figure 2(b) of the Appendix visualizes the α coefficients of these archetypes. We also observed that :

- Archetype 1 had the lowest values across all attributes.
- Archetype 2 had the highest values across all attributes, except for "Intelligence" and "Magic Points".
- Archetype 3 had the highest "Intelligence" and "Magic Points".

4.3 AIRvatar

For characters created using *AIRvatar*, we found $k = 3$ to be optimal. The bar plot in Figure 1 shows the three archetypes obtained, represented with the same six RPG attributes. We observed the following:

- Archetype 1 had highest "Intelligence" and "Wisdom" attributes, but lowest "Strength" and "Endurance".
- Archetype 2 had the highest "Strength", "Endurance", and "Dexterity" attributes, but the lowest "Wisdom".
- Archetype 3 had the highest "Charm" but lowest "Dexterity".

5 FINDINGS

5.1 Classes, Roles, and Category Gradience

In *Oblivion*, we found that each **archetype corresponded with the primary roles of the game**, namely "Fighter" (Archetype 1), "Mage" (Archetype 2), and "Thief" (Archetype 3). We used descriptions in the *Unofficial Elder Scrolls Pages* [15], to help identify these roles from obtained archetypes. "Fighters" '*rely heavily upon melee combat to attack enemies, expect to receive a lot of damage rely upon high health...*', "Mages" '*avoid combat, use decoys, and rely upon magical attacks.*' *Magicka*, used for spells and magic, is affected by both "Intelligence" (Capacity) and "Willpower" (Regeneration). A "Thief" '*relies upon sneak attacks and avoids face-to-face combat, uses a poisoned bow as a primary means of attack,*' corresponding to the high "Speed" and "Dexterity" (Bow Accuracy) attributes.

Likewise, in *Ultima IV*, we observed from our results that each **archetype corresponded with characters of primary roles in the game**. Katrina the Shephard is Archetype 1 as her description in the *Unofficial Ultima IV Strategy Wiki* [21] states "...she has the lowest attributes, no magic power and a limited selection of equipment; start the game with her if you're looking for a challenge". Archetype 2 corresponds to "Iolo the Bard", who has the highest "Dexterity" described as "probably the most important attribute because it rules the probability of hitting enemies, avoiding traps and dodging enemies."Archetype 3 corresponds to "Mariah the Mage", with highest "Intelligence" (determines maximum "Magic Points".)

For characters created by players in *AIRvatar*, we observed from our results that the archetypes corresponded with **traditional RPG roles used in games**, which we term "Intelligent/Wise-Cleric" (Archetype 1), "Physical-Fighter," (Archetype 2) and "Charming-Thief" (Archetype 3). We the descriptions of traditional *Dungeons and Dragons* classes to match against the highest-scoring attributes of each archetype to identify these roles. Magic using "Mages/Clerics" focus on magic, and generally have lower strength. "Fighters" are usually strong in attack and defense, but usually have little to no magic capabilities, while "Thieves" often are in-between, but have high capabilities in social skills, cunning and stealth.

We validate this based on the free-text responses that players provided for their avatars, in addition to customizing their characters. We provide selected responses from the highest scoring players for each archetype to highlight this behavior:

1. **Archetype 1 (Intelligent/Wise-Cleric):** "*Stephanie is a wandering wolf* ***mage****. She was born to a poor family, but her parents did their best to support her academic ventures. She studied hard and was eventually admitted to the nation's most prestigious* ***arcane*** *academy.*"
2. (**Archetype 2 (Physical-Fighter)**: "*Gerald ... is a veteran of many wars in Elibca, serving as a* ***knight*** *and later as a* ***general*** *for the kingdom of Calmenia ... living the remainder of his life in modesty as he nurses* ***old scars***." & "*Saya is an independent* ***Mercenary*** *selling her contract not to the highest bidder, but to those she deems in the most need of her services. Secretly, she dreams of becoming a Paladin some day but believes that she has* ***far too candor in her speech*** *and methodology to fit in ...* "
3. **Archetype 3 (Charming-Thief):** "*She is* ***friendly*** *and ready to reach out to the other villages. She* ***prefers talking to fighting****, but is tough enough to fight if she needs to.* "

These results shows that AA can effectively model implicit categories, such as intended player roles and relationships between attributes from analyzing raw statistical attribute data. For example, in both *Oblivion* and *AIRvatar*, "Strength" and "Intelligence" attributes are always maximized on different archetypes, while "Strength" and "Endurance" were be maximized on archetypes together. Additionally, with archetypes corresponding to prototypical player roles, we observed that **each individuals could meaningful represented as mixtures of these archetypes**, corresponding to hybrid roles intended by most designers.

5.2 Revealing Stereotypes, Marginalization, and Inequity

5.2.1 Race-related Stereotyping

From the archetypal analysis results on characters in *Oblivion*, we were able to observe that **some of the in-game races were deemed more "prototypical" with respect to player roles and that we could observe how these in-game races reflected real world stereotypes**. To visualize this, we make use of the ternary plot of results shown in Figure 2(a) of the Appendix. This is best visualized using a ternary plot, as shown in Figure 2(a) of the Appendix. We observe that the ostensibly Norwegian "Nords" are viewed as archetypal Fighters, the ostensibly French "Bretons" as archetypal Mages, and ostensibly South American "Bosmers" as archetypal Thieves. Additionally, the obstensibly African-American "Redguards" stereotypically close to the physical-fighter archetype with no characteristics of the intelligence-mage archetype, though exhibiting some stealth-thief archetype characteristics. This corresponds with findings by Harrell in his assessment of racial stereotypes in *Oblivion* [9].

5.2.2 Gender-related Inequity & Marginalized Characters

In *Oblivion* we also note that **for each race, male characters are consistently deemed more prototypical than their female counterparts than their female counterparts.** This is illustrated in Figure 2(a), where for each archetype, the male characters are always at least as close, or closer to the archetypes, than their counterpart female characters. Insight into the significance of characters being closer to the centers (i.e., further away from archetypes) is highlighted in the design choices made in *Ultima IV*, wherein the NES version of *Ultima IV*, "Julia" was replaced by a male character "Julius", with no modification to the stats. From the ternary plot in Figure 2(b) of the Appendix, it can be seen that "Julia" is the character with negligible "Intelligence" and "MP" attributes and located between the overall lowest and highest-performing archetypes, possessing multiple memberships. This computational modeling of a **less prototypical individual would, by Lakoff's definitions [11], represent the marginalization of that individual**. We hypothesize that the implications of this made it seem "low-stakes" to swap her gender within the game and that it might have been more difficult to swap the genders of an archetype instead (i.e., making a Katrina a male to have the lowest stats or Iolo a female while having the highest stats.) To validate the effects of marginalization (being further away from archetypes), we sampled characters created with *AIRvatar* that had coefficient values $.3 \leq \alpha_k \leq 0.6$ for *all three archetypes*. These reflected characters that players created to be less prototypical.

- **Character #41:** "*Pinkie is a girl with a* ***unique gift*** *for magic, . . . works* ***best in a team but can hold her own*** *when needed.*"
- **Character #102:** "*A spellcaster . . . a love for* ***forbidden magics****.* ***Chaotic good****, generally* ***tries to do the right thing but isn't afraid*** *to crack a few eggs to make an omlette.*"

5.2.3 Gender-related Stereotyping

In the results of characters created using *AIRvatar*, we observed that **players constructed characters with more homogeneous gender distributions between archetypes and also when close to the archetypes**. We define close as individuals with coefficient values $\alpha_k >= 0.80$. In Table 1 of the Appendix, we observe that all three archetypes had a mixture of male and female avatars close to each of them. Both Archetype 1 ("Intelligent/Wise-Cleric") and Archetype 3 ("Charming-Thief") had more female avatars closer to the the archetypes than male avatars, while Archetype 2 ("Physical-Fighter") had more male avatars closer to it. These results share similarities with those of *Oblivion*, *Ultima IV*, as well as our precious analyses in [14] where males avatars were associated with more physical roles, and female avatars with magic-related roles. For the "Charming-Thief" role, neither females nor males were closely associated with it – showing that "Thief"-like roles have less gender stereotyping associated with them. These results appears to suggest that taken collectively, players seek to reduce the degree of marginalization or privilege of either gender relative to what designers commonly portray. We hypothesize that perhaps, in the absence of a well-known game series, people relied more on real-world gender stereotypes. Thus, these results may reveal what people do without being restricted to canonical classes and roles – an observation perhaps useful for developers incorporating race and gender into their designs.

6 LIMITATIONS & FUTURE WORK

Here we discuss several limitations of our approach and describe potential avenues for overcoming them with future work and directions.
1. Determining the number of archetypes The approach we outlined in Section 3 adopts Occam's Razor [2] in that we pick the lowest number of archetypes k from the minimization of the residual sum-of-squares (RSS). However, this may not always be effective, with the result possibly being that the archetypes discovered are not sufficient to *adequately represent* the rest of the data points. For example, with $k = 3$ archetypes applied to results from *AIRvatar*, we discovered that no close individuals ($\alpha_j >= .9$) for one of the archetypes. It is possible that other metrics for determining k could be employed (e.g., choosing higher values of a scree plot's elbow.)
2. Normalizing Statistical Attributes While there are similarities between the statistical attributes used for defining characters in various videogames, there are issues with standardizing the number, the descriptions, and the effects that each attribute has. Also, there is a tension between the gaming use of these terms like "Intelligence" or "Wisdom." and their real meanings. Additionally, different games use different numerical scales (e.g., upon-100 in *Oblivion* but upon-7 in *AIRvatar*) for these attributes. It is difficult to translate the significance of each point due to different granularities. A standardized list and scale would be useful for such cross-platform comparisons.
3. Representation Beyond Statistical Attributes Representation in computing systems spans across several other technical components of the system, including graphical assets and textual descriptions [8]. Our next step is to analyze additional data collected using *AIRvatar*, which include the images of the constructed avatars, textual descriptions made by players, and other behavioral data obtained using the analytical capabilities of *AIRvatar*. We believe that these additional sources of information will enable further insight into the types of social phenomena that players experience and encounter through virtual representations in videogames and other computing systems.

7 CONCLUSION

We have demonstrated a novel approach to computationally model cognitively grounded social identity phenomena in videogames using archetypal analysis (AA). Previous work in this area has relied on qualitative methods (e.g., self-reported surveys) to identity and assess the presence of social identity-related issues such as marginalization, stereotyping, and discrimination. We demonstrated AA's effectiveness for modeling gender-related marginalization and biases like males being represented as closer archetypes than females and race-related stereotypes like in-game races possessing attributes that reflect characteristics of real-world stereotypes. AA was also able to reveal implicit categories like prototypical RPG roles used in videogames, which had implications to such race and gender-related phenomena. Being able to reveal such emergent phenomena through analyzing the data structures and designs of systems mean that computing systems can be analyzed in a systematic way, enabling quantifiable insight to be gained while minimizing the common effects of subjective evaluations such as survey bias. We believe that these findings contribute towards substantiating the use of AI to better understand the effects of virtual characters on players behaviors.

ACKNOWLEDGEMENTS

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

REFERENCES

[1] Christian Bauckhage and Christian Thurau, 'Making archetypal analysis practical', in *Pattern Recognition*, 272–281, Springer, (2009).
[2] Anselm Blumer, Andrzej Ehrenfeucht, David Haussler, and Manfred K Warmuth, 'Occam's razor', *Information processing letters*, **24**(6), 377–380, (1987).
[3] Geoffrey C. Bowker and Susan Leigh Star, *Sorting Things Out: Classification and its Consequences*, MIT Press, 1999.
[4] Raymond B Cattell, 'The scree test for the number of factors', *Multivariate behavioral research*, **1**(2), 245–276, (1966).
[5] Adele Cutler and Leo Breiman, 'Archetypal analysis', *Technometrics*, **36**(4), 338–347, (1994).
[6] Manuel JA Eugster, 'Archetypal athletes', *arXiv preprint arXiv:1110.1972*, (2011).
[7] Erving Goffman, *Stigma: Notes on the Management of Spoiled Identity*, 1963.
[8] D Fox Harrell, 'Computational and cognitive infrastructures of stigma: Empowering identity in social computing and gaming', *Proceedings of the 7th ACM Conference on Cognition and Creativity*, 49–58, (2009).
[9] D. Fox Harrell. Toward a theory of critical computing. CTheory, ctheory.net/articles.aspx?id=641, 2010.
[10] Ian Jolliffe, *Principal component analysis*, Wiley Online Library, 2005.
[11] George Lakoff, *Women, Fire, and Dangerous Things: What categories reveal about the mind*, 1990.
[12] George Lakoff, *Moral politics: How liberals and conservatives think*, University of Chicago Press, 2010.
[13] Daniel D Lee and H Sebastian Seung, 'Learning the parts of objects by non-negative matrix factorization', *Nature*, **401**(6755), 788–791, (1999).
[14] Chong-U Lim and D Fox Harrell, 'Toward telemetry-driven analytics for understanding players and their avatars in videogames', in *In CHI'15 Extended Abstracts on Human Factors in Computing Systems*, (2015).
[15] Oblivion: Character Creation. The Unofficial Elder Scrolls Pages, 1995.
[16] Eleanor Rosch, 'Principles of categorization', *Concepts: Core readings*, 189–206, (1999).
[17] Otto Santa Ana, *Brown tide rising: Metaphors of Latinos in contemporary American public discourse*, University of Texas Press, 2002.
[18] Sohan Seth and Manuel J. A. Eugster, 'Probabilistic archetypal analysis', Technical report, arXiv.org, (2014).
[19] Rafet Sifa, Christian Bauckhage, and Anders Drachen, 'Archetypal game recommender systems', *Proc. KDML-LWA*, (2014).
[20] Claude M Steele and Joshua Aronson, 'Stereotype threat and the intellectual test performance of african americans.', *Journal of personality and social psychology*, **69**(5), 797, (1995).
[21] Ultima IV: Quest of the Avatar - Companions. StrategyWiki, 2013.
[22] Nick Yee and Jeremy Bailenson, 'The proteus effect: The effect of transformed self-representation on behavior', *Human communication research*, **33**(3), 271–290, (2007).

A COEFFICIENT TABLES

Description	α_1	α_2	α_3	Player Gender	Avatar Gender
Archetype 1 ("Intelligent/Wise-Cleric")	***1.00**	0.00	0.00	Female	Female
	0.90	0.00	0.10	Female	Female
	0.82	0.17	0.01	Male	Male
Archetype 2 ("Physical-Fighter")	0.00	***1.00**	0.00	Male	Male
	0.00	***1.00**	0.00	Female	Female
	0.00	***1.00**	0.00	Male	Male
	0.00	***1.00**	0.00	Male	Male
	0.00	***1.00**	0.00	Male	Male
	0.00	0.88	0.12	Male	Female
	0.00	0.87	0.13	Male	Male
	0.14	0.86	0.00	Male	Male
	0.15	0.85	0.00	Male	Male
	0.14	0.85	0.02	Female	Female
	0.00	0.83	0.17	Female	Male
	0.00	0.80	0.20	Male	Male
Archetype 3 ("Charming-Thief")	0.10	0.00	***0.90**	Male	Male
	0.00	0.11	0.89	Female	Female
	0.15	0.00	0.85	Female	Female

Table 1. Table of characters created with *AIRvatar* with high α coefficients to each archetype. Values ≥ 0.90 are bolded. * marks the closest individual(s) of each archetype.

B BAR PLOTS

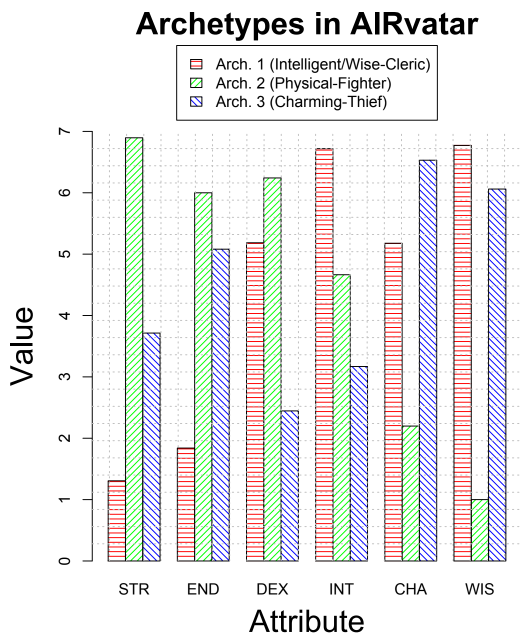
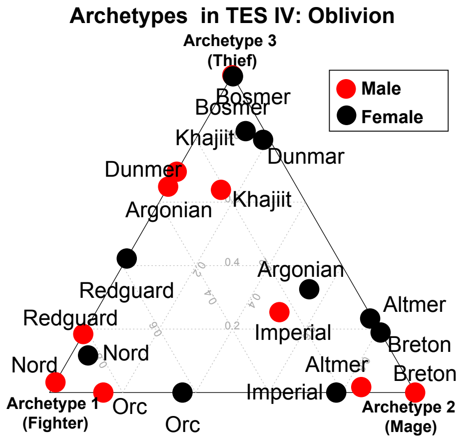
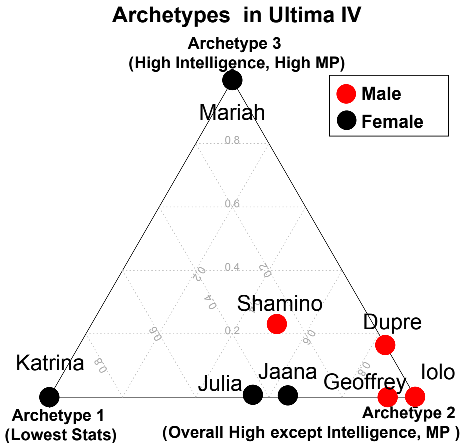


Figure 1. The plot above shows the $k = 3$ archetypes obtained from archetypal analysis on the data set of players and their statistical attribute allocations to each of their avatars. Due to convexity constraints, archetypes can be meaningfully represented with the same features of the original data.

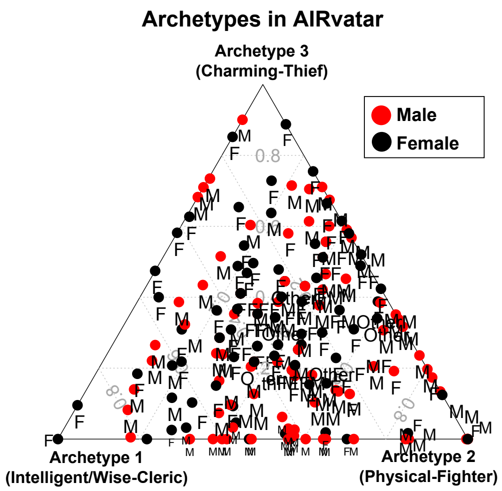
C TERNARY PLOTS



(a) Oblivion



(b) Ultima IV



(c) AIRvatar

Figure 2. Ternary plots representing characters as mixtures of archetypal archetypes in *The Elder Scrolls IV: Oblivion*, *Ultima IV*, and from our *AIRvatar* system. Labels for (a) denote races in *Oblivion*, (b) denote names in *Ultima IV*, and (c) player gender in *AIRvatar*.