The Marginal: A Game for Modeling Players' Perceptions of Gradient Membership in Avatar Categories

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

Abstract

We encounter the results of category formation every day, from demographic categories like race and gender, to role-playing-game classes like "fighter" or "mage". Category membership is often not simply based on the possession of discrete properties but instead constructed from and reflect the highly nuanced relationships (gradience) between members and best-example individuals called "prototypes". In this paper, we present *The Marginal*, an artificial intelligence (AI)-driven game that (1) computationally models the cognitive categories that players develop when customizing videogame avatars and (2) generates challenges for players to use their perception of visual, textual, and numerical data to progress in a game created using these models. We use archetypal analysis, an AI clustering approach for identifying boundary points in data, to generate tasks in *The Marginal* for its gameplay. It shows how AI can be combined with games to model and evaluate cognitive categorization phenomena.

Introduction

We as humans are known to form categories of things we encounter in everyday life – from demographic profiles of race and gender, to other aspects of our lives such as genres of music. In virtual environments, this too may occur in the form of computer role-playing-game (RPG) classes based on roles and abilities. Much of this categorization occurs cognitively and *invisibly*, as social scientists Geoffrey Bowker and Susan Leigh Star point out (Bowker and Star 1999). Challenges occur when what we cognitively perceive as distinctions between concepts do not match the hard-coded, pre-determined categories that we encounter or when these predefined categories of existing systems inadvertently reinforce real-world inequitable phenomena such as stereotyping, discrimination, or marginalization. For example, despite being set in a fantasy setting with fictional characters and races, the commercially successful and critically acclaimed RPG *The Elder Scrolls IV: Oblivion* reflected racial stereotypes through the descriptions and attribute statistics used to represent the ostensibly African "Redguard" race versus the ostensibly French "Bretons" (Harrell 2010). Harrell argues that the underlying computational data structures used for these representations contribute to these undesirable results (Harrell 2009) and suggests that using categorization models that are cognitively grounded can enable us to better understand how these social categorization phenomena occur, with the aims of developing more robust systems that avoid the limitations of existing models. Our motivation stems from seeking to use Artificial Intelligence (AI) algorithmic methods to reveal these invisible, implicit user categories and computationally model them as they emerge from data. Beyond the benefits of being able to quantitatively represent these categories, this serves the social and expressive needs of better supporting the diversity of users and tailoring content their individual preferences and values.

In this paper, we present *The Marginal*[1], which uses AI clustering to computationally model and evaluate the cognitive categories that players form when creating and interacting with avatars in videogames. Players are tasked with identifying an out-of-place avatar termed the "marginal." It differs from the rest based on a set of characteristics covering visual, textual, and numerical data. With multiple factors for consideration, players need to rely on their cognitive perception of similarities (or differences) between the avatars, akin to "centrality gradience" in cognitive categorization theory (Lakoff 1990). The gameplay in *The Marginal* is procedurally generated based on AI models developed from performing an AI-clustering technique called archetypal analysis (AA) on a data set of avatars created by players using an avatar creator we also developed. The goals of *The Marginal* are (1) to enable us as AI researchers to better understand and develop robust approaches for computationally modeling complex phenomena such as category categorization, and (2) demonstrate how such cognitive phenomena can lead to interesting AI-based games, which we believe would be a useful approach to enable us to gain insight into players' perceptions of category formation and classification.

Background & Related Work

We present an overview of relevant theories and concepts from AI and cognitive science to contextualize our aims with the theoretical underpinnings of our approaches.

[1]Prototype Build: http://bit.ly/the-marginal

Cognitive Categorization

We use categorization models from cognitive science in order to characterize the kinds of social identity phenomena we seek to discover. We provide a brief introduction and discuss several of these models. In the traditional "folk" models of categorization, criteria for being a member of a category was based on possessing a fixed set of characteristics. However, this has proven to be ineffective for capturing the nuances that exist when dealing with more complex or dynamic category phenomena like multiple memberships. More recent theories are based on identifying members that are deemed "better examples" of a category than others, which are termed *prototypes* by psychologist Eleanor Rosch (Rosch 1999). Thus, categorization of individuals occurs based on their *perceived distances relative to these prototypes* (i.e., similarity.) Cognitive scientist George Lakoff extends upon these models with what he terms "prototype effects," based on the theory that categorization is an cognitively-grounded and imaginative process involving metaphorical projection (Lakoff 1990). The following list of useful definitions summarized in (Harrell 2010) are used for describing categories from such models, and will be useful for describing the results obtained from our approach.

- Prototypes: The "best example" members of categories.
- Naturalization: The deepening familiarity of interactions within a given social group or category.
- Membership: Encountering and interacting with objects within certain social groups and increasingly engaging in naturalized relationships with them.
- Marginalization: A result of enforced naturalization where members of exist outside of social groups, are less prototypical members of communities, or characterized by possessing multiple memberships.

Thus, being close to an exemplary prototype, and thus categorically associated with it, is termed "centrality." The notion of "gradience" occurs because "members within the category boundaries may be more or less central" (Lakoff 1990). Using such models and concepts from the sociology of classification, we can formally describe categorization-related phenomena such as stereotypes (commonly held, but often misleading, category expectations), ideals (culturally valued categories, which may not be typically encountered), and paragons (categories defined in terms of individual members who represent an ideal or its opposite.) Consider the difference between an ideal husband, which is that of a good provider and is faithful, versus a stereotypical husband, which might represent one that is bumbling and beer-bellied. These definitions enable us to describe and reason about the cognitively grounded categorization models and social identity phenomena in what are termed Idealized Cognitive Models (ICM) (Lakoff 1990). We seek to use AI to computationally model ICMs, enabling us to algorithmically represent cognitive categorization-based phenomena such as centrality gradience. Our aims are similar to other works that use games to model and evaluate cognitive constructs like real-world demographic and personality profiles (Tekofsky et al. 2015; van Lankveld et al. 2011). We next describe the AI approach that we used to create *The Marginal*.

Archetypal Analysis

Archetypal analysis (AA) (Cutler and Breiman 1994) is a clustering method for reducing the dimensionality data and representing it as a *convex combination* of a set of key data points called **archetypes**. For example, applying AA on a dataset of soccer players and their statistics (Eugster 2011) computationally revealed and represented the following four archetypes – "offensive player," "center forward," "defender," and "weak player." Every individual player in the entire data set could then be represented as a hybrid mixture of these archetypes (Seth and Eugster 2014). 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:

$$z_j = \sum_{i=1}^{n} \beta_{ij} x_i \qquad (1) \qquad\qquad \hat{x}_i = \sum_{j=1}^{k} \alpha_{ji} z_j \qquad (2)$$

Equation 1 means each archetype z_j resembles (i.e., represented using the same feature variables) as the data and Equation 2 specifies that each data point x_i can then be represented as a weighted combination of the archetypes. The objective function minimizes the residual sum of squares:

$$RSS = ||x_i - \hat{x}_i||^2 = ||x_i - \sum_{j=1}^{k} \alpha_{ji} z_j||^2 \qquad (3)$$

under the constraints that the coefficients $\sum \alpha_{ij} = 1$ $\alpha_{ij} \geq 0$ and weights $\sum \beta_{ji} = 1$ $\beta_{ji} \geq 0$. These ensure the archetypes *meaningfully resemble* and are *convex mixtures* of the data. These archetypes are located on the data convex hull (Cutler and Breiman 1994) and are represented as combinations of individual points, making them more easily interpretable (Bauckhage and Thurau 2009), unlike other dimensionality reduction techniques like principal component analysis (Jolliffe 2005) and non-negative matrix factorization (Lee and Seung 1999). We used the convex-hull non-negative matrix factorization (CHNMF) algorithm (Thurau, Kersting, and Bauckhage 2009) to construct our AA models. As detailed later, the α-coefficients from model centrality gradience as they quantitatively represent how close an individual in the dataset is to an archetype. Previous work in (Risi et al. 2014) used clustering via self-organizing maps (SOM) for generating gameplay. To our knowledge, beyond our work on behavioral archetypes modeling gender-related differences based on timing data (Lim and Harrell 2015d), AA clustering has not been previously used for evaluating social categorization phenomena with games.

System Architecture & Design

In Figure 1, we provide an overview of the various existing systems, or systems that were made, that supported the development of *The Marginal*. We first provide a brief overview of how the systems work together before going into greater detail into each of them in subsequent sections.

Heroes of Elibca is an avatar creator we developed that where players created avatars in the style of traditional 16-bit role-playing-game (RPG) characters. All created avatars and analytical data is collected by our analytics system *AIRvatar*,

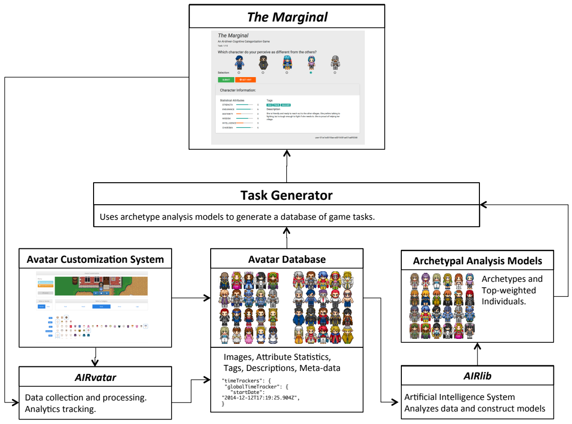

Figure 1: A system architecture diagram showing the various systems used to create *The Marginal*. Players first customize characters using *Heroes of Elibca*, with data collected by *AIRvatar* used to construct a database of player avatars, including images, text, etc. Our AI system *AIRlib* analyzes this data to create archetypal analysis models. These models are used by a task generator for gameplay in *The Marginal*.

that processes and stores it on our server. This provided us with a database of created avatars, each composed of various assets like images and text descriptions. Currently, our database contains 191 created avatars that were obtained as a result from previous user study. More details on both *AIRvatar* and the user-study involving *Heroes of Elibca* are available in (Lim and Harrell 2015a; 2015c). This database is used an AI-system *AIRlib* that constructs various different computational models using the data. *The Marginal* makes use of archetypal analysis (AA) models, which are first processed by a task generator to develop different challenges for the player. Finally, *The Marginal* presents tasks to the player as an interactive game via a graphical user interface (GUI).

Customizing Avatars in *Heroes of Elibca*

In order to provide players with a variety of ways to represent their avatar, *Heroes of Elibca* provides four different characteristics of the avatar that can be customized. These characteristics are based on the taxonomy of technical components of identity representation systems in (Harrell 2009).

Attribute Stats Players customized both their character's visual appearance and statistical attributes values of six commonly used videogame attributes (strength, endurance, dexterity, wisdom, intelligence, and charisma) on a 7-point scale. Each attribute was defaulted to 4 points. A total of 27 attribute points were to be allocated to each player's avatar.

Tags and Descriptions Players were provided with two ways provide text-based representations for their created avatars. First, players were asked to provide simple one-word *tags* for their avatars (e.g., "strong, clever, brooding"). Secondly, players were provided input for *free-text entry*, enabling them to provide more verbose descriptions of their avatars. Both inputs were optional and each had an example help text as a guide for players.

Images Players had two choices of avatar genders[2] to select from, in turn providing access to a gender-specific base image and assets across the five customization categories to customized the visual appearance of their characters – hair, head, body, arms, and legs. In each category, several sub categories of assets gave players more fine-grained control over their avatar's appearance, e.g., gloves for hands or pads for shoulders under the arm category. Players were provided an animated preview of their avatars and could rotate the view in any of the four directions. Each created character is 32×48 in pixel dimensions, with an animation preview.

AI Model Construction using *AIRlib*

We constructed separate AA models for each characteristic. Given a player i and characteristic j, each avatar is represented as a feature vector v_{ij}. A data set of all $N = 191$ player avatars for a particular characteristic j would be represented as a $N \times M$ matrix V_j for performing non-negative matrix factorization. The feature vector v_{image} was created by flattening the 32×48 avatar image. Since each pixel is represented by RGBA values, the number of features $M = 32 \times 48 \times 4 = 6144$. The feature vector $v_{attributes}$ has $M = 6$ to correspond to the six types of customizable attributes. Both v_{tags} and $v_{description}$ were constructed using bag-of-words (BOW) representation. For tags, the number of features M corresponds to the number of unique tags observed across the data set. For descriptions, the number of features M was the total number of unique English word terms observed after tokenizing each of the text descriptions in the data set, capped at $M = 500$ as a default parameter.

The Marginal

Having described the underlying systems, we now present a detailed overview of *The Marginal*, which uses the aforementioned AI models of category gradience for gameplay.

Gameplay

In each run of *The Marginal*, players are presented with a series of tasks. Each task has the same basic goal: players have to rely on their cognitive perception of centrality gradience to identify the single avatar, out of a set of candidates, that differs the most greatly from the rest, henceforth termed the "marginal". However, the reasons for the marginal differing can be based on any of the following four characteristics: (1) its visual appearance, (2) the distribution of its attributes, (3) its associated tags, and (4) the text description of the avatar. While such components do not manifest in the real-world (e.g., people do not view one another in terms of numeric stats), they are often employed in computational identity representation systems like videogames and can reinforce socially undesirable constructs (Harrell 2009). It is worth noting here that these similarities and differences are not determined through a simple calculation of overlapping characteristics (e.g., equivalent values for attributes or common words) but are instead calculated based on the AA

[2] We follow role-playing conventions here, but recognize the distinction between gender and sex. In future work, we seek representations that decouple biological sex and gender.

models of centrality gradience. This should be particularly apparent for characteristics like the images and text descriptions and we cover a more detailed explanation of this in a later section. Players can request for a hint, which informs them on which particular characteristic to focus on. Finally, at the end of each run, the game presents a summary of the player's performance in accurately identifying the marginal across the series of tasks. Players could optionally specify which characteristics contributed to each result.

User Interface

A screenshot of the user interface (UI) of *The Marginal* is shown in Figure 2. The main elements of the UI are the (1) progression indicator, (2) images of the candidate avatars, (3) radio buttons for player selection, (4) a breakdown of attribute values, (5) associated tags, and (6) a text description. Items (4)–(6) refresh whenever a selection is made, since they reflect the properties of the currently selected avatar. Additional elements are (7) the hint button to guide players, and (8) a generated user identification (ID) string for anonymous data collection purposes. Not shown in the figure is a modal dialog the prompts the player to confirm their choice.

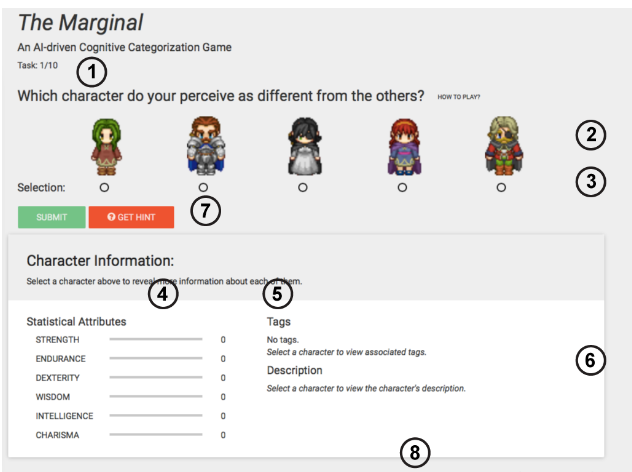

Figure 2: Screenshot of the UI of *The Marginal*. The goal is to identify which of the five candidates is different from the rest across various characteristics (attributes, tags, descriptions, images) as computed by an underlying AA model.

Sample Narrated Playthrough

As mentioned briefly earlier, the challenge in *The Marginal* lies in being able to discover what characteristics are shared by the other $N-1$ of the candidates that make the N^th one the marginal. Because there are four characteristics to look at, players need to carefully study each avatar in order to make an informed guess. Furthermore, because the similarities and differences are modeled using "centrality gradience" rather than discrete factors, there is substantial nuance that makes all characters different from one another in various ways. We illustrate this challenge with a narrated playthrough of some of *The Marginal*'s generated tasks.

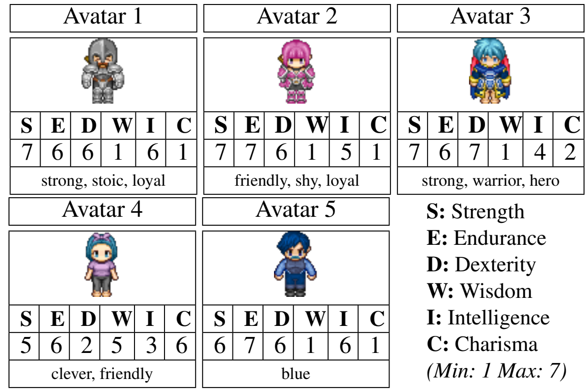

Figure 3: Candidates avatars from a task in *The Marginal*. One is the marginal (most dissimilar from rest), which players have to analyze the attributes, tags, and images to identify. Some tags and text descriptions omitted for simplicity.

The Challenge Consider the candidate avatars shown in Figure 3 and their associated characteristics. For simplicity, we omit text descriptions and selectively reduced the some tags. We welcome readers to attempt to identify the marginal before we provide the narrated playthrough.

The first characteristic we shall look at are the images. We might first see that the grayscale colors of Avatar 1 are different from the others, which feature mainly shades of pink and blue. But another categorization could exist, such as the fact that Avatars 1–3 are all wearing some form of armor, unlike Avatars 4–5. Thus, considering images alone appears insufficient to isolate the marginal, since the two different ways of categorizing the characters so far do not yield a consistent resultant marginalized character. The next characteristic we look at are the tags. Now, we see that the "loyal" tag categorizes Avatars 1 and 2, but Avatar 1 also shares the "strong" tag with Avatar 3, while Avatar 2 shares the "friendly" tag with Avatar 4. Avatar 5 looks suspiciously out of place with just the single "blue" tag, but for the sake of completeness, let us consider the distribution of attribute values next. We first observe that all avatars have high "endurance." We could categorize Avatars 1–3 together as a group due to have maximum "strength" values and Avatars 4–5 as another group as having moderately high "strength" values. However, we also discover that all avatars, except for Avatar 4, have low "charisma" and "wisdom" values. To add to that last train of thought, we also note that all avatars, except Avatar 4 again, have high "dexterity" values. Thus, we this might be a more compelling case for Avatar 4 being the marginal. This is the AA model's answer, as explained next.

Underlying Computational Model

We previously described a sample narrated playthrough that illustrated the challenge and an possible approach to determining the solution. We now use the underlying computational model of *The Marginal* to explain how the each candidate avatar was quantitatively represented to construct the scenario. First, we chose to use an AA model of the dataset with $k = 3$ archetypes, previously shown in (Lim and Har-

	S	E	D	W	I	C
A1	6.6	**6.7**	**6.7**	1.0	4.9	1.2
A2	1.2	3.3	4.3	**7.0**	5.7	5.4
A3	**7.0**	**7.0**	4.0	1.0	1.0	**7.0**

(Computed Archetypes)

α	S	E	D	W	I	C
.99	7	7	6	1	5	1
.98	7	6	6	1	6	1
.97	7	7	7	1	4	1
.96	6	7	6	1	6	1
.87	7	6	7	1	4	2

(A1's Top-5 Individuals)

α	S	E	D	W	I	C
.00	3	4	3	7	3	7
.00	2	2	4	7	5	7
.00	3	5	3	7	3	7
.00	3	3	3	5	6	7
.00	5	6	2	5	3	6

(A1's Bottom-5 Individuals)

Figure 4: The tables show the underlying models used to generate the task described. A selected archetype A1 and its top and lowest-five weighted individuals are shown. Green rows indicate those sampled to produce five candidates.

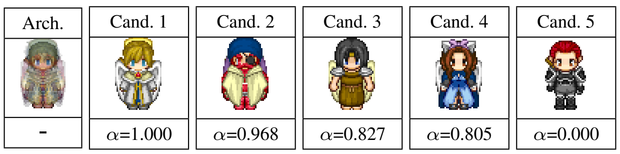

Figure 5: Images generated for a task using an archetype from the AA model (left-most image). They here shown in order of decreasing weights, but are shuffled in-game.

rell 2015b) to result in archetypes with sufficient distinctions from one another. The top table of Figure 4 shows the three archetypes and their attributes, illustrating these distinctions where A1 has high "endurance" and "dexterity," A2 has high "wisdom," and A3 has high "strength," "endurance," and "charisma." Again, we highlight that, being computationally derived from our AA models, the distinctions between archetypes are based on holistic nuances rather than discrete overlapping of features (e.g., attributes.)

Second, the system randomly selects Archetype A1 for a given task. The tables at the bottom of Figure 4 show the top and bottom-five weighted individuals of A1. We note that the top weighted individuals have unique attribute distributions, but are all highly weighted and deemed close to A1 (high centrality). Bottom individuals are also each unique, but have low weights and are far from A1 (low centrality.) Randomly selecting four top and one bottom individual produces the five candidates from our playthrough in Figure 3.

Another Task Example: Images In order to provide a better understanding on the effectiveness of these AA models for its use in generating tasks in *The Marginal*, we provide an additional example that focuses on images. The same process from the previous section is undertaken (i.e., choose a $k = 3$ AA model and randomly select 4 top individuals and 1 bottom individual), and the selected archetype and candidates are shown in Figure 5

We see that candidates avatars 1–4 possess cloaks, wings, and/or capes-like accessories – a characteristic represented by the archetype. Candidate 5 has neither and is the marginal

Attr			Tags			Desc			Imgs			Overall		
A	E	%	A	E	%	A	E	%	A	E	%	A	E	%
3	3	1.0	1	2	.50	1	2	.50	1	3	.33	6	10	.60
1	2	.50	1	3	.33	0	2	0.0	0	3	0.0	2	10	.20
2	2	1.0	3	3	1.0	2	2	1.0	1	3	.33	8	10	.80
1	2	.50	1	3	.33	2	2	1.0	2	3	.67	6	10	.60
3	3	1.0	3	3	1.0	0	2	0.0	0	2	0.0	6	10	.60
2	2	1.0	2	3	.67	0	2	0.0	0	3	0.0	4	10	.40
2	2	1.0	1	3	.33	1	2	.50	2	3	.67	6	10	.60
1	2	.50	0	2	0.0	2	3	.67	3	3	1.0	6	10	.60
2	2	1.0	3	3	1.0	0	2	0.0	2	3	.67	7	10	.70
2	2	1.0	1	2	.50	1	3	.33	3	3	1.0	7	10	.70
19	22	.86	16	27	.59	9	22	.40	14	29	.48	58	100	.58
Key: — **A**: No. of Answered Tasks in Agreement — **E**: No. of Tasks Encountered														

Figure 6: The table summarizes the results of a pilot study conducted with N=10 participants. Each row represents the results of each player's playthrough that included 10 tasks.

in this case. This demonstrates the robustness of the AA models, since the same underlying algorithmic modeling is used for generating this task involving visual images.

Intended Outcomes

Our goals of *The Marginal* are to: (1) enable players to gain insight into their own cognitive perception of category membership based on different factors (e.g., visual, comprehension, interpretation) and (2) enable us as researchers to evaluate AI models are at modeling centrality gradience. Toward achieving goal (1), at the end of the playthrough, players are presented with a performance report with a breakdown of scores of the player's of their performance. For goal (2), we plan to conduct a large scale user study that would provide us with the required data for analysis. We outline a pilot study the provides a preliminary evaluation of our approach.

Pilot Study

As a preliminary evaluation of *The Marginal*, we conducted a pilot study with 10 participants, 8 of whom were graduate students in university and 2 were working professionals, and between the ages of $25 - 32$ years old. Only 1 was a female, which is a skewed demographic. However, the aim of was not meant to be conclusive, but to test system functionality and gain early insight into results for further study.

Experimental Design

Each player was linked to an online version of *The Marginal* to play on their own computers. Data and results were remotely collected with *AIRvatar*. Each player had given 10 tasks (2 per characteristic and 2 randomly chosen), selected from a generated pool of 40 tasks. Figure 6 shows a summary of these results with a player per row.

Preliminary Results

We discovered that overall (last column), players identified the marginal in agreement with the AA models an average of 58% of the time, which is promising given that a random choice between 5 candidates for each task would give a baseline performance of 20%. Out of the four characteristics, our models predicted players' choices best with attribute values (86%), tags (59%), images (48%) and worst with descriptions (40%). We suspect that (1) a small set of features (M=6) for attributes enabled k=3 prototypes to be a sufficient AA model for distinguishing between candidates, and (2) numerical values (attributes) and discrete words (tags) were easier to comprehend than free-prose text and images. Detecting pixel-based differences may not be as apparent and future work to model the discrete categorical item choices is planned. However, performance across all of the characteristics were better than random selection. These preliminary findings, though early, point toward our approach being effective at modeling players' perception of cognitive categorization and centrality gradience in avatars.

Discussion

We discuss the implications of our findings from *The Marginal* along with limitations and plans for future work.

Data-driven Approaches for Implicit Phenomena

Recall that for each task, both the marginal and the rest of the candidates are not manually categorized by hand. Rather, categories are modeled using our underlying AI models, in turn computed based on a data set of actual avatars created by players. This emergent, "bottom-up" approach better reflects the implicit values and preferences of players compared to system creators determining them on their own. Studying players' avatar customization behaviors in this way has previously shown interesting findings, such as players conforming to common RPG class or gender-related stereotypes (Lim and Harrell 2015b; 2015c). The broader implications could involve alternative data sets based on other aspects of society in the real world (e.g., demographic, education, employment), using it to model and reveal society's values of other socially-charged topics. This forms an approachable way to help others better understand such nuanced phenomena, akin to (Talton et al. 2009)'s use of "landmarks" to introduce the notion of generative spaces of trees.

AI Clustering and Categorization Dynamics

In *The Marginal*, we made use of AI models constructed using archetypal analysis (AA). The reason for doing so was due to its effectiveness at identifying extremal individuals in a data set that are both interpretable and could be used to represent the entire data set (Drachen et al. 2014). However, there are several other clustering approaches such as non-negative matrix factorization (NMF), k-means clustering, and principal component analysis, that each result in cluster models that reveal other interesting aspects of the data. Additionally, there are several other "prototype effects" from cognitive categorization theory besides marginalization that can be investigated, such as stereotypes, ideals, and paragons (Lakoff 1990). Combining both sets of models can be effective at modeling the nuances of social category dynamics (Harrell et al. 2014) , including complex concepts of stigma and impression management (Goffman 1963). We believe there exists a close relationship between AI clustering and category dynamics and that such cross-disciplinary projects can help us develop robust models for capturing the complexities of categorization-based social phenomena.

Limitations & Future Work

There are several aspects of how gameplay could be improved. Firstly, while limiting models to $k = 3$ archetypes appears adequate for comparatively low dimensionality features (e.g., attribute statistics), features with high dimensionality like tags, descriptions, and image data likely require more archetypes. Secondly, each task currently only uses AA models over a single characteristic (e.g., attributes, tags, descriptions, images) and never a combination. A task could produce an intended marginal based on one characteristic, but with a chance that a non-marginal candidate is a marginal based on another. In our narrated playthrough, we saw signs of such problems where, even though Avatar 4 was the marginal based on attributes, Avatar 5 could also be a marginal is based on tags. Thus, further work could go into (1) filtering out tasks that pose such problems or (2) developing ways to combine AA models before generating a task. The latter approach is considerably more difficult, but would have deeper implications for developing more nuanced AI models. Thirdly, using the top and bottom-5 weighted examples maximized the differences between the marginal and the others. We could increase the difficulty by using moderately weighted examples to make distinctions less evident. Lastly, we plan to conduct a more thorough user study in order to evaluate the effectiveness of games like *The Marginal* in allowing players to learn more about cognitive categorization-related concepts, as well as to give insight into how to extend work for improving our models.

Conclusion

In this paper, we have presented *The Marginal*, a game that uses artificial intelligence (AI) to develop computational models of cognitive categories of avatars based on numerical, textual, and image data. Players in *The Marginal* are presented with a group of avatars and tasked with identifying a single marginalized individual called the "marginal," that is categorically different from the rest. Tasks differ based on the degree of similarity between the candidates and players are not informed of which characteristics to focus on in order to make a decision. The tasks in *The Marginal* are not hand-crafted, but are instead procedurally generated based on archetypal analysis models that were constructed from a dataset of player-created avatars. This meant that any categorization that separated avatars were created in a "bottom-up" manner, emerging from the data, and reflected actual categories that players' had implicitly developed. Our intention was to show that games using AI, such as *The Marginal* can be used as a means to assess the computational models of players' perception of categorization phenomena. We seek to further develop such technologies as tools to critically assess such phenomena that may provide better insight into how implicit categories are formed by players.

Acknowledgment

We thank the reviewers for their valuable feedback. This material is based upon work supported by the National Science Foundation under Grant No. 1064495. We also acknowledge the support of a QCRI-CSAIL Collaboration Grant.

References

Bauckhage, C., and Thurau, C. 2009. Making archetypal analysis practical. In *Pattern Recognition*. Springer. 272–281.

Bowker, G. C., and Star, S. L. 1999. *Sorting Things Out: Classification and its Consequences*. MIT Press.

Cutler, A., and Breiman, L. 1994. Archetypal analysis. *Technometrics* 36(4):338–347.

Drachen, A.; Thurau, C.; Sifa, R.; and Bauckhage, C. 2014. A comparison of methods for player clustering via behavioral telemetry. *arXiv preprint arXiv:1407.3950*.

Eugster, M. J. 2011. Archetypal athletes. *arXiv preprint arXiv:1110.1972*.

Goffman, E. 1963. *Stigma: Notes on the Management of Spoiled Identity*. Touchstone.

Harrell, D. F.; Kao, D.; Lim, C.-U.; Lipshin, J.; Sutherland, A.; Makivic, J.; and Olson, D. 2014. Authoring conversational narratives in games with the chimeria platform. In *Proceedings of 9th Annual Conference on Foundations of Digital Games*.

Harrell, D. F. 2009. 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.

Harrell, D. F. 2010. Toward a theory of critical computing. CTheory, ctheory.net/articles.aspx?id=641.

Jolliffe, I. 2005. *Principal component analysis*. Wiley Online Library.

Lakoff, G. 1990. *Women, Fire, and Dangerous Things*. Cambridge University Press.

Lee, D. D., and Seung, H. S. 1999. Learning the parts of objects by non-negative matrix factorization. *Nature* 401(6755):788–791.

Lim, C.-U., and Harrell, D. F. 2015a. Developing computational models of players' identities and values from videogame avatars. In *Proceedings of the 10th International Conference on the Foundations of Digital Games*.

Lim, C.-U., and Harrell, D. F. 2015b. Revealing social identity phenomena in videogames with archetypal analysis. In *Proceedings of the 6th International AISB Symposium on AI and Games*.

Lim, C.-U., and Harrell, D. F. 2015c. Toward telemetry-driven analytics for understanding players and their avatars in videogames. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '15, 1175–1180.

Lim, C.-U., and Harrell, D. F. 2015d. Understanding players' identities and behavioral archetypes from avatar customization data. In *Proceedings of the IEEE Conference on Computational Intelligence and Games*.

Risi, S.; Lehman, J.; D'Ambrosio, D. B.; and Stanley, K. O. 2014. Automatically categorizing procedurally generated content for collecting games. In *Proceedings of the 5th FDG Workshop on Procedural Content Generation in Games*.

Rosch, E. 1999. Principles of categorization. *Concepts: Core readings* 189–206.

Seth, S., and Eugster, M. J. A. 2014. Probabilistic archetypal analysis. Technical report, arXiv.org.

Talton, J. O.; Gibson, D.; Yang, L.; Hanrahan, P.; and Koltun, V. 2009. Exploratory modeling with collaborative design spaces. *ACM Transactions on Graphics* 28(5):167.

Tekofsky, S.; Spronck, P.; Goudbeek, M.; Plaat, A.; and Van den Herik, J. 2015. Past our prime: A study of age amp; play style development in battlefield 3. *Computational Intelligence and AI in Games, IEEE Transactions on* PP(99):1–1.

Thurau, C.; Kersting, K.; and Bauckhage, C. 2009. Convex non-negative matrix factorization in the wild. In *Data Mining, 2009. ICDM'09. Ninth IEEE International Conference on*, 523–532. IEEE.

van Lankveld, G.; Spronck, P.; Van den Herik, J.; and Arntz, A. 2011. Games as personality profiling tools. In *Proceedings of the IEEE Conference on Computational Intelligence and Games*, 197–202. IEEE.