Analyzing Avatar Customization Between Groups

Peter Mawhorter*, Sercan Şengün*, Haewoon Kwak†, and D. Fox Harrell*‡
*Computer Science and Artificial Intelligence Laboratory / ‡Comparative Media Studies Program
Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.
†Qatar Computing Research Institute, Hamad Bin Khalifa University, Education City, Doha, Qatar.
E-mail: pmwh@mit.edu, sengun@mit.edu, haewoon@acm.org, fox.harrell@mit.edu

Abstract

As virtual identities such as social media profiles and avatars have become an important avenue of self-expression, it has become important to consider the ways in which existing systems embed the values of their designers. In order to design virtual identity systems which reflect the needs and preferences of diverse users, understanding how virtual identity construction differs between groups is important. This paper presents a new methodology that leverages deep learning for comparative analysis of profile images, with a case study of almost 100,000 avatars from a large online community using a popular avatar creation platform. We use deep learning to assess the "novelty" of each avatar, and then use differential clustering to identify region-specific visual trends among low- and high-novelty avatars. In our case study, we find that avatar customization correlates with increased social activity, and we are able to identify distinct visual trends among U.S.-region and Japan-region profiles. Among these trends, realistic, idealistic, and creative self-representation can be distinguished. We observe that realistic self-expression mirrors national demographics, idealistic self-expression reflects shared mass-media tropes, and creative self-expression propagates within communities.

Index Terms

Avatars, Cultural differences, Unsupervised learning, Artificial neural networks, Data analysis, Image processing, Clustering algorithms, Deep learning

Analyzing Avatar Customization Between Groups

I. Introduction

Virtual identity images—ranging from avatars in videogames and virtual reality to social media profile portraits—have been the subject of sustained research recently because of their central role in identity construction both on social networks and in online videogames. Research such as [1] has underscored that *"Avatars in virtual worlds and social media can impact people's self-perception in the real world and provide proxies for people to engage in communities as players, learners, and doers."* Composed of images as well as text and other data, virtual identities become blended as users project their preferences and values onto their creations [2]. While this projection means that avatars can reveal much about their creators, it is selective and arbitrary, making it a challenge to articulate exactly what we learn about the humans behind these avatars.

Designers recognize that avatars are important, as evidenced by, for example, the ubiquity of purchasable skins as a revenue source in videogames. Various avatar characteristics are related to user behaviors and preferences, such as body type [3], skin color [4], gender predictors [5], body weight [6], color palette [7], and more. In fact, Kao and Harrell have shown that avatar properties can even affect engagement and performance, including learning [8], [9].

Given these impacts, the design of virtual identity systems is an important concern for those building media, videogaming, and virtual reality systems. We also know that designers must also be keenly aware of culturally-specific values when designing systems targeted at international audiences, and this raises an interesting question: how can we best contrast trends in avatar use between different settings, for example across demographic groups or international regions? To address this question, we use unsupervised machine learning to find patterns directly from publicly-available data which includes two different regions at the international level.

The data we use are profiles from Nintendo's Miiverse network (see Fig. 1) [10]. Mii avatars are used across a variety of Nintendo products, and their popularity (both in Japan and globally) is evident from the success of recent products such as the Miitomo smartphone app [11], [12]. The combination of a large international user-base with highly customizable avatars makes Miis perfect for contrasting avatar customization trends between different groups, such as users from the U.S. and Japan regions. Additionally, the Miiverse social network includes detailed profile information such as self-reported expertise, genre preferences, and social network stats, so we can find correlations between avatar appearances and other features related to social networking and user preferences.

We gathered more than 300,000 profiles from Miiverse, selecting active users by finding people who had posted highly-rated levels on the Miiverse Super Mario Maker community [13]. We chose Super Mario maker both for its popularity [14] as well as for the way that its focus on user-generated content encourages interaction among a community of players. By finding profiles who had posted well-received levels, we ensured that all of the profiles we gathered were visible and active within the community. After filtering for only Japanese and U.S. profiles, we removed private accounts (whose friend counts and other social data are hidden) to construct a dataset of 90,124 public profiles. In order to analyze this data and find patterns specific to the U.S. and Japanese contexts, we developed two novel techniques: a deep convolutional neural network to sort images according to their "novelty," and a differential clustering technique to find images exemplary of either the U.S.-region or Japan-region subsets.

Our first technique adds a "novelty" dimension to our data which measures how much each image conforms to overall visual conventions common to the entire dataset

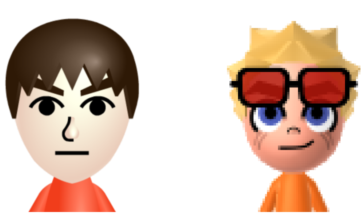

Figure 1. The default male Mii avatar (left), and a Mii customized to look like a specific fictional character (right). Note the cartoonish style and diverse facial features available.

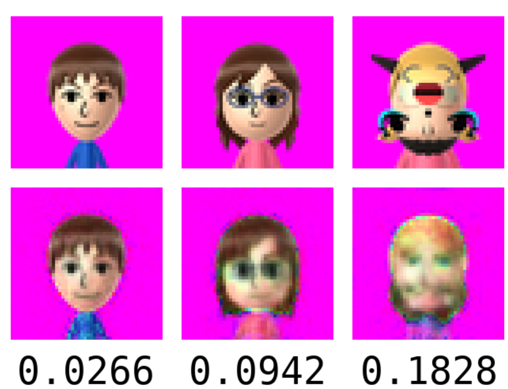

Figure 2. Low-novelty, medium-novelty, and high-novelty Miis (top) along with their reconstructions by our network (bottom) and corresponding novelty values (per-pixel root-mean-squared error of the reconstruction). Note how the network tries (and fails) to reconstruct the high-novelty Mii from more typical facial features. The images are shown at 48×48 pixels, the same resolution the network uses.

(see Fig. 2). To do this, we train an autoencoding convolutional network using modern deep learning methods and then measure its error in reconstructing each image, labelling more-difficult-to-reconstruct images as more-novel. We are able to find several interesting correlations between novelty and profile information such as one's number of friends or interest in certain genres.

The most pronounced correlation with novelty is across the social activity and network size variables, such as number of posts and number of friends. This sustained relationship with both activity and network size, in conjunction with a relationship between novelty and self-reported gaming expertise, could be explained by existing theories that relate virtual identities with player behavior (e.g., results showing that avatar customization is linked to social activity such as [15]). These correlations are consistent with multiple causal links between investment in play, in the social network, and in avatar customization. In particular, we assert that our novelty dimension, which measures how well an avatar can be explained using visual elements common among the entire set, can be used as a rough proxy for the customization effort invested in that avatar, which would corroborate existing work that links customization and engagement (e.g., [16]–[19]).

After sorting images by novelty, we analyze the most- and least-novel portions of our dataset to find images specific to either the U.S. or Japan regions. Our second novel technique is an algorithm to find exemplars inspired by prototype theory (see §II-E) [20] which identifies images that are both maximally distant from opposite-nationality profiles while also having the most same-nationality neighbors. By jointly maximizing these two quantities using a rank-based sort, we can identify images which reflect each country's trends in avatar customization, including instances of realistic, idealistic, and creative self-expression. These trends can be explained by both demographics (for realistic self-expression) and aesthetic preferences (for idealistic and creative self-expression).

In the following sections we highlight important related work that puts ours in context, describe our technical implementation and statistical methodology, present our results, and discuss them. For those interested in the technical details of our system, those are described in §III. Meanwhile, the details of the relationships that we found are discussed in §IV.

II. Related Work

A. Deep Learning

The roots of deep learning can be traced back to early multi-layer networks from the 1980s and '90s, including self-organizing maps [21], [22] which led to the development of convolutional neural networks [23]. Key problems, like regularization for sparse coding [24], had already been addressed by the turn of the millennium.[1] However, the problem of vanishing gradients was a challenge for these early multi-layer architectures, and limited processing power and data made training large networks difficult.

In 2006, Hinton and Salakhutdinov demonstrated that careful weight selection and specific training regimes could be used to effectively conquer the problem of vanishing gradients [25]. This breakthrough and other work in the 2000s (e.g., [26]) led to a renaissance in neural network research and a barrage of impressive results. The combination of increased training efficiency and increasing computational power allowed for larger networks and training sets (see [27] for a seminal example). During this time a flurry of innovations reduced training times, increased accuracy, and helped stave off various undesirable results, many of which we have made use of. In particular, our system uses Rectified Linear Units (ReLUs) [28] and AdaGrad optimization [29] on top of a stacked autoencoding architecture similar to that of Vincent et al. [30] (although we do not incorporate denoising); we also use output normalization as mentioned above [24].

B. Neural Network Applications

Our application of deep convolutional networks to image analysis is inspired by existing research. For example, Häkkinen's use of neural networks for exploring quantitative data prefigures our technique, although we are using a different kind of network and are training on image data instead of coded experiment logs [31]. Another similar application is that of Vega Ezpeleta, who trained a network to identify images of swastikas in profile avatars [32]. The main difference in our work is the use of an unsupervised model, which does not require labeled training data and is thus applicable with much less effort.

Applications of unsupervised deep learning include domains such as medical imaging [33], satellite imagery [34], and, of course, face and object detection [35]. Again, our goal of measuring novelty differentiates our work from these approaches. We also eschew domain specialization in order to keep our approach widely applicable.

C. Virtual Identity

Analyzing virtual identity systems can reveal real-world phenomena, because users project their identities onto their avatars. We use the term virtual identity within a broad framework that includes social media accounts, videogame avatars, and user profiles on e-commerce websites. These virtual identity systems typically mobilize a combination of numerical attributes, images, and other data structures to represent an individual. Individuals, in return, often find ways of creatively using the available options to enact their values, preferences, and self-representations [36]. This process becomes especially crucial for underrepresented communities if virtual identity systems generally fail to provide options to address their values and cultural norms.

Existing research shows that virtual identities can affect one's behaviors and beliefs in the real world [1], [37]. Moreover, avatars have considerable effects on the performance

[1] Olshausen and Field's work deserves special mention because it highlights relationships between convolutional networks and the human vision system and makes the case for developing sparse representations of complex visual patterns.

and engagement of individuals in the virtual world [7], [8], [17], [38], [39]. Analyzing avatar appearance can also reveal norms and stereotypes around issues like body image, gender [5], [40], and other aspects of social identity [15], [41], [42]. Digitally mediated self-expression ultimately has important ramifications for cultural dynamics, including ethnic self-expression and the interactions between ethnic groups in virtual spaces [43]–[45].

The availability of options such as gender [46], [47] and ethnicity [48], [49] that facilitate self-resemblance is also empirically linked to user performance in virtual environments. Studies in non-game social networks have revealed links between avatar appearance and both social network activity and connectivity, mediated by personality traits such as extroversion or narcissism [16]. Unsurprisingly, there appear to be social factors that prompt users to engage in avatar creation or customization (see the discussion section in [18], and page 192 of [50]).

D. Self-Representation

Our discovered categories of self-representation (realistic, idealistic, and creative) are similar to those used in social psychology to talk about self-conception, in particular the distinction between personal and collective identities (see e.g., [51], [52]). However, the distinction that we make is not at the level of personal psychology, but instead it is about how one expresses one's personality. Through ethnographic studies, Turkle makes a distinction between expressing one's real or idealized self and expressing an alternate identity through role-playing [53]. A more fine-grained distinction is made by Neustaedter and Fedorovskaya [54], which includes a four-fold typology including "realistics," "ideals," "fantasies," and "roleplayers." They highlight differences in relation to the real-world self, number of identities, and identity continuity among player groups, whereas we collapse the distinction between realistic and idealistic self-expression to focus solely on whether elements of real-world identity are present. We add an additional category of creative self-expression to highlight the difference between users who reproduce common aesthetics and those who come up with their own. Of course, most virtual identities include aspects of all three forms self-expression, but in our data we find many that emphasize one.

E. Cognitive Categorization

In our analysis of exemplars from different subsets of our data, we draw upon the ideas of prototype theory, as studied empirically by Rosch [20], [55].[2] Along with other more recent work on categorization [56], [57], prototype theory has proved a useful tool for understanding virtual identities (see e.g., [58], [59]). In finding exemplars, we draw on prototype theory to define them as instances which are both surrounded by many other examples from the same group, and distant from members of other groups (thus being central within their local region). Of course, our computational approach uses our trained network's feature space to measure similarity, which may or may not correspond to human judgements, but the relationship to prototype theory is otherwise direct.

III. Methodology

A. Data Acquisition and Pre-Processing

The data for this analysis was scraped from Nintendo's Miiverse social network platform [10] using a custom web crawler to download profiles from users with the highest-rated uploaded levels in the Super Mario Maker community on August 21st, 2016. For each profile scraped, we collected public information from the website, including the profile name, personal information (country and birth date), social information (friend count, following count, follower count, post count, and "yeahs[3]" count), and gaming profile information (skill level, systems owned, and preferred genres). For this analysis, we focus on the country code, social information, and gaming profile information (besides systems owned). Note that the skill level is user-selected from "Beginner," "Intermediate," and "Expert" options, and that users could list up to three preferred genres (many listed none) from the list: "Action," "Adventure," "Fighting," "Puzzle," "Racing," "RPG," "Simulation," "Sports," "Shooter," "Board Game," and "Music."

We selected users from the two largest countries in the dataset (the U.S. and Japan, which are also two of the largest game markets in the world [60]) who had public profiles (private profiles have their social activity information hidden), and for each of these users we downloaded the Mii avatar image from their profile page. It is important to note that there are mechanisms for copying Miis from other users, and Miis can even be saved as a QR code that anyone can scan to create a copy, so the construction of avatars itself can be a social process.

Before we began processing, we first simplified the Mii images using the ImageMagick 'convert' tool for batch processing [61]. Each image was resized from 96×96 to 48×48 pixels, and the alpha channel was removed, resulting in a 3-channel image that was saved to disk in PNG format. During this process a flat fully-saturated magenta background (RGB: 255, 0, 255) was introduced; magenta was chosen because it was a very rare color within the dataset, and it almost never appeared on the borders of the avatars. This dimensionality reduction in terms of resolution and channels greatly speeds up processing times while retaining the human-recognizable features of the original images. Most examples shown in this document are these resized-and-framed images rather than the originals (apologies for the amount of magenta that results).

[2] Although research on categorization with indigenous populations including the work of Rosch and others is problematic in its treatment of their cultures and languages as "simpler" than Western counterparts (see e.g. [20] p. 331), we believe that the conclusions they draw remain informative.

[3] "Yeahs," akin to more well-known "likes," indicate how many posts a user has upvoted.

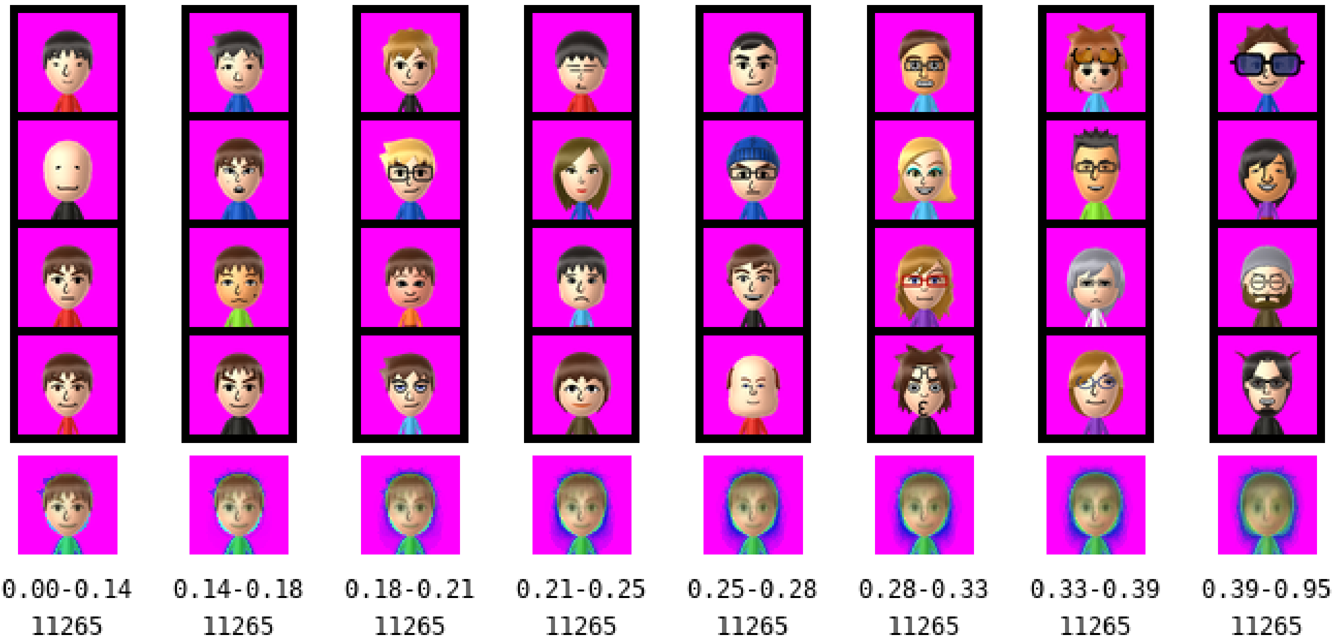

Figure 3. An image lineup for the "novelty" value. Columns include four random samples from a percentile plus a mean across 2,500 random samples from that percentile (the bottom row). Labels are the range of values represented within that percentile (top) and the total number of images in that percentile (bottom). Note how the mean image becomes less distinct as novelty increases, showing how low-novelty Miis are highly similar to each other (and in fact, they are clustered near the default feature settings). The blue halos in the mean images are the results of averaging a minority of overall brownish hair/accessory colors against a majority of the magenta background color in the HSV color space (we did not implement circular averaging for hue).

For processing in `python`, we loaded the base data from a single comma-separated value (CSV) file using the `pandas` scientific data processing library [62]. The CSV file contained image filenames in each row, which were loaded into memory only as-needed using the `scikit-image` library, which was used along with `matplotlib` to produce figures [63], [64]. The final image-processing step consisted of a color-space transformation from the RGB space on disk into the HSL space.

For some of our analyses, we performed a log-transform on the social variables (friends, following, followers, posts, and yeahs) because we expected them to be roughly exponentially distributed. To avoid the singularity at zero but retain its distinctiveness, we used $y = \log{(x + 0.5)}$, which for our integer inputs resulted in negative values for zeros and positive values for other numbers.

Out of an initial set of 305,433 profiles from 79 countries, 202,793 were from the U.S. or Japan, and of these 90,124 were public.[4] It is worth noting that even this sample of 90,000 items is large enough to present some difficulty when using modern consumer hardware (and in this sense we are working with "big data"). For example, a variety of analyses might want to built a matrix of pairwise distances, but a $90,000 \times 90,000$ matrix of floating point numbers would require 30 GB of memory, which can be prohibitive. One of the advantages of our technique is that we are able to load as little as a single image at a time during both training and analysis so our method is extremely scalable. In contrast to novelty assignment, our exemplar-identification step *does* require pairwise distances as input, but for that we rely on a limited neighborhood size to reduce memory requirements.

[4]Filtering in this way biases our results towards more-active, more-engaged users, which is intentional.

B. Network Setup

In order to construct our "novelty" metric, we employ a deep convolutional neural network similar to the work of [30]. However, our network does not introduce extra noise for denoising, because we actually want our network to over-fit on our training data ("novelty" measures the network's *incomplete* generalizability). We also train the network all-at-once instead of training layer-by-layer, which works in large part due clever weight initialization [28]. We set up the network using the `keras` library in `python`, with the `tensorflow` backend [65], [66]. Following general conventions (see e.g., [67], [68]) we set up a convolutional network with two convolutional layers with 3×3 kernels and 32 and 16 pattern units respectively. Each convolutional layer used a rectified linear unit (ReLU) activation function and was followed by a 2×2 max-pooling layer. After these four convolution/pooling layers, our network has an output shape of $12 \times 12 \times 16 = 2,304$ values, which we then flatten and feed into three fully-connected layers with output sizes 512, 256, and 128 to produce the compressed feature representation (the dense layers also used ReLU activation). On this downward path there are a total of 1,349,904 parameters. The choice of 128 output units is arbitrary and easy to adjust, but we felt that this gave the network enough room to learn fairly complex representations of the data while still forcing it to achieve significant compression.

In our autoencoder setup, the network proceeds upwards as a mirror image of the downwards route: three dense layers with 256, 512, and 2,304 output nodes respectively, followed by two 3×3 convolutional layers each preceded by a 2×2 upsampling layer. To get back to a $48 \times 48 \times 3$ output shape, a final convolutional layer was added with 3 units and a 3×3 kernel, unlike the other layers this final layer used a sigmoid activation function. The full network in both directions has a total of 2,704,291 parameters.

For training, we used a mean squared error loss function computed between the input image and the reconstructed image. We added an L1 regularization term based on the activity of the innermost 128-node layer (this is just the sum of the activation values of this layer) to force the network to learn sparse representations; this term had a coefficient of 1×10^{-5} (we added this per [68]; see [24] for the theoretical justification). Adjusting this coefficient affects the representations the network learns, but besides observing poor performance without regularization, we have not explored this parameter in detail.

C. Training and Analysis

For our final results, we trained for 100 epochs on our 90,124 example images, using a batch size of 32.[5] After training, we computed novelty ratings by loading each image, asking the network to reconstruct it, and measuring the root mean squared error (RMSE) between the original and its reconstruction (this is the loss function modulo the L1 regularization term). We then normalized these values across the dataset, so that novelty ratings were between 0 and 1. We also took the 128 dimensions of the internal representation of each image and recorded them as a feature vector for that image. We pruned these representations by ignoring “monotonous” features for which more than 97.5% of the data had identical output values. At this stage, we found that the network always output zero for 82 of our 128 features.[6] Of the remaining 46 features, we pruned 10 more due to monotony.

1) Statistical Analysis: We checked for relationships between our novelty metric and each of the following variables: country, competence (three boolean variables for the three levels), log-transformed versions of the social variables (friends, followers, following, posts, and yeahs), and twelve boolean variables for each of the eleven genres plus `did-not-list-any`. For each variable, we computed either a permutation test for nonzero Pearson’s product-moment correlation [69] between novelty and that variable, or for boolean variables, a Welch’s t-test [70] for unequal mean novelty values between the true and false cases.[7] We performed a total of 21 separate tests against our novelty value, each using the filtered dataset, for $n = 90,124$. After computing p values for each of our tests, we used the Holm-Bonferroni correction for multiple comparisons [71] to ensure a family-wide false positive rate of no more than 5%, and found that 19 of our tests rejected the null hypothesis (of either no relationship or indistinguishable means). The highest passing p value was 0.01233, with a corresponding threshold of 0.01667; the next test, with $p = 0.08275$, failed its threshold of 0.025.

Subsequent analysis of correlation strengths found that the effects were general trends rather than strict relationships (e.g., Pearson’s $r < 0.1$), but this was expected: it would be surprising to find that any of our social or genre variables was determinant of a particular aspect of visual appearance, because people have quite diverse visual tastes. Given that these were general trends, we computed effect sizes in terms of property change per 0.1 change in novelty (10% of the normalized novelty scale) using linear regression, and found that most effects were quite noticeable, especially those related to social variables.

D. Finding Exemplars

After analyzing how novelty relates to other variables, we focused on the low-novelty (bottom 10%) and high-novelty (top 10%) groups. In order to study national-level[8] trends within each group, we searched for the avatars which were exemplars in either the U.S. or Japan subsets. To find these exemplars, we sorted each group according to both distance from the nearest image of the other nationality (Euclidean in feature space) and the number of same-nationality images within that radius.

The former value, termed *separation*, measures how distinct an image is from images of the other nationality, and maxima should be found within nation-specific groups, or else be outliers in the dataset. The latter value, termed *centrality*, distinguishes these cases, where maxima are images which follow a common theme shared by others of the same national origin.

After sorting images separately in descending order by these separation and centrality metrics, we used the sum of ranks in each ordering (with ties broken arbitrarily) as an exemplar score, with lower scores being better. After picking the best exemplar, we excluded all nearby images (those counted as part of the chosen exemplar’s centrality score) from consideration before picking the next one, so that each exemplar identified represented a distinct cluster. We proceeded to display the top 16 exemplars from each category, as shown in Fig. 4. These exemplars revealed visual aesthetics specific to each country, some

[5] Running on a laptop with an NVidia Georce GTX 950M graphics card, an Intel Core i7-6700HQ processor (2.6 GHz), 16 GB of ram, and a 5,400rpm HDD, training took 3–4 hours. Although this method is time-consuming, it does not require specialized hardware.

[6] This result is due to the L1 regularization term which penalizes the sum of activations. The balance between the sparsity and learning objectives thus regulates the number of active features regardless of how many are available.

[7] We used the `pearsonr` and `ttest_ind` functions from `scipy.stats` in `scipy` version 0.19.1 to compute these statistics

[8] We study trends at the national level because the data is readily available. As pointed out in, e.g., [72], this has its limitations, because nations are multicultural. What we find are prominent trends, not universal properties.

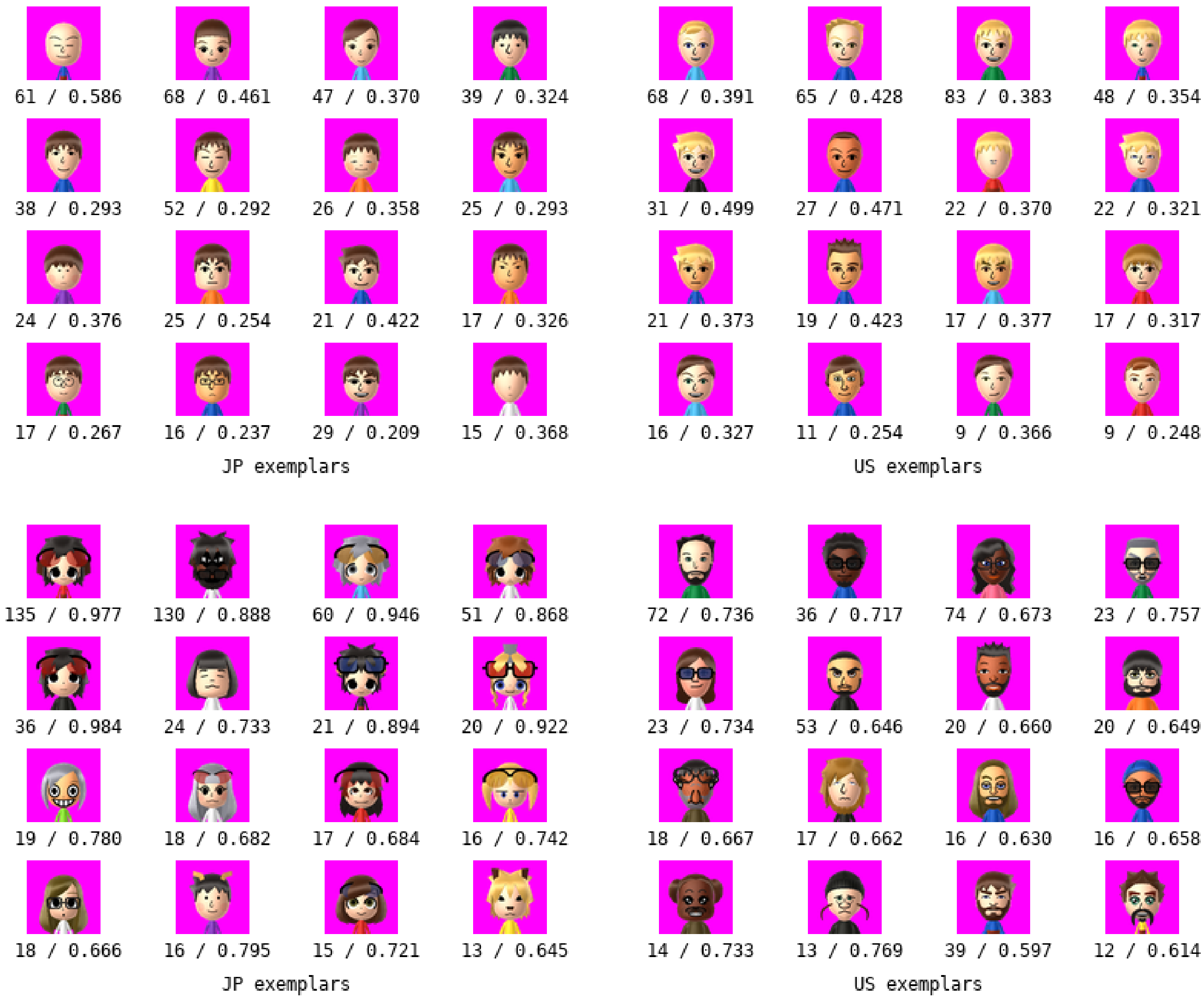

Figure 4. The top 16 low-novelty (top) and high-novelty (bottom) exemplars from the Japan (left) and U.S. (right) subsets. Numbers indicate centrality (number of same-nationality Miis nearby) and separation (feature-space distance to the nearest different-nationality Mii).

of which were distinguishable into the realistic, idealistic, and creative categories of self-expression.

IV. Results

Table I summarizes our statistical results (tests are for correlation with novelty, or difference in novelty by condition for boolean variables), including effect statistic and size where p values are significant. The "Summary" column provides a concise summary of our results, and the details of two relationships are shown in Fig. 5. In the table, failed tests are marked with a red dot before the p value. Effect statistic is either Pearson's r for continuous variables or mean difference in novelty between conditions (δ_μ) for boolean variables. Because Pearson's r does not measure effect size, the effect size column lists linear regression coefficients of each variable against novelty, converted into appropriate units. Effect sizes are reported for a change in novelty of 0.1, which is 10% of the full range 0–1. Note that linear changes in the logistic variables have a multiplicative rather than additive effect. Also note that the relationships in some cases were not strictly linear, but the regression slope gives a general indicator of effect magnitude.

Overall, effect sizes are noticeable, especially when extended to the full range of novelty values (10× the listed effect), despite the fact that our statistic values indicated general trends rather than strict dependencies. For example, a person whose avatar has a novelty value of 0.7 (quite novel) would on average have almost twice as many friends as someone whose avatar has a novelty value of 0.2 (not very novel).[9] For the same pair, the higher-novelty user would be 7% more likely to list at least one preferred genre, and 9% more likely to list role-playing

[9] This ratio is computed as $1.14^5 = 1.925$ taking 1.14 from the "ln(friends)" line in Table I and raising it to the fifth power as each 0.1 difference in novelty multiplies by this amount.

Table I
STATISTICAL RESULTS

Variable	p Value	Statistic	Effect Size (per +0.1 novelty)	Summary
country:US	8.3×10^{-27}	$\delta_\mu = 0.00835$	+1.5 US%	More moderate-novelty in U.S.
skill:Beginner	1.2×10^{-09}	$\delta_\mu = -0.00555$	-0.71 Beginner%	Novelty is correlated with experience (more experts, fewer beginners + intermediates).
skill:Intermediate	4.1×10^{-07}	$\delta_\mu = -0.00408$	-0.7 Intermediate%	
skill:Expert	5.9×10^{-19}	$\delta_\mu = 0.00721$	+1.2 Expert%	
ln(friends)	2.7×10^{-183}	$r = 0.096$	×1.14 friends	
ln(following)	8.1×10^{-206}	$r = 0.102$	×1.18 following	Novelty (∼customization) is correlated with social network size and activity.
ln(followers)	2.3×10^{-241}	$r = 0.11$	×1.19 followers	
ln(posts)	8.6×10^{-75}	$r = 0.0609$	×1.06 posts	
ln(yeahs)	8.2×10^{-117}	$r = 0.0764$	×1.17 yeahs	
‹no genres›	7.0×10^{-28}	$\delta_\mu = -0.00951$	-1.4 ‹no genres›%	
genres(Action)	0.083	–	–	
genres(Adventure)	0.1	–	–	
genres(Fighting)	0.0023	$\delta_\mu = 0.00277$	+0.38 genres(Fighting)%	Interest in RPG and music games correlates strongly with novelty. Overall, novelty correlates with more stated preferences (more engagement) except in the case of sports games.
genres(Puzzle)	0.0055	$\delta_\mu = 0.0031$	+0.27 genres(Puzzle)%	
genres(Racing)	0.0058	$\delta_\mu = 0.00315$	+0.28 genres(Racing)%	
genres(RPG)	4.4×10^{-52}	$\delta_\mu = 0.0137$	+1.8 genres(RPG)%	
genres(Simulation)	1.8×10^{-17}	$\delta_\mu = 0.0128$	+0.63 genres(Simulation)%	
genres(Sports)	1.1×10^{-08}	$\delta_\mu = -0.00912$	-0.42 genres(Sports)%	
genres(Shooter)	0.012	$\delta_\mu = 0.00279$	+0.27 genres(Shooter)%	
genres(Board Game)	7.2×10^{-08}	$\delta_\mu = 0.0128$	+0.23 genres(Board Game)%	
genres(Music)	5.0×10^{-72}	$\delta_\mu = 0.027$	+1.5 genres(Music)%	

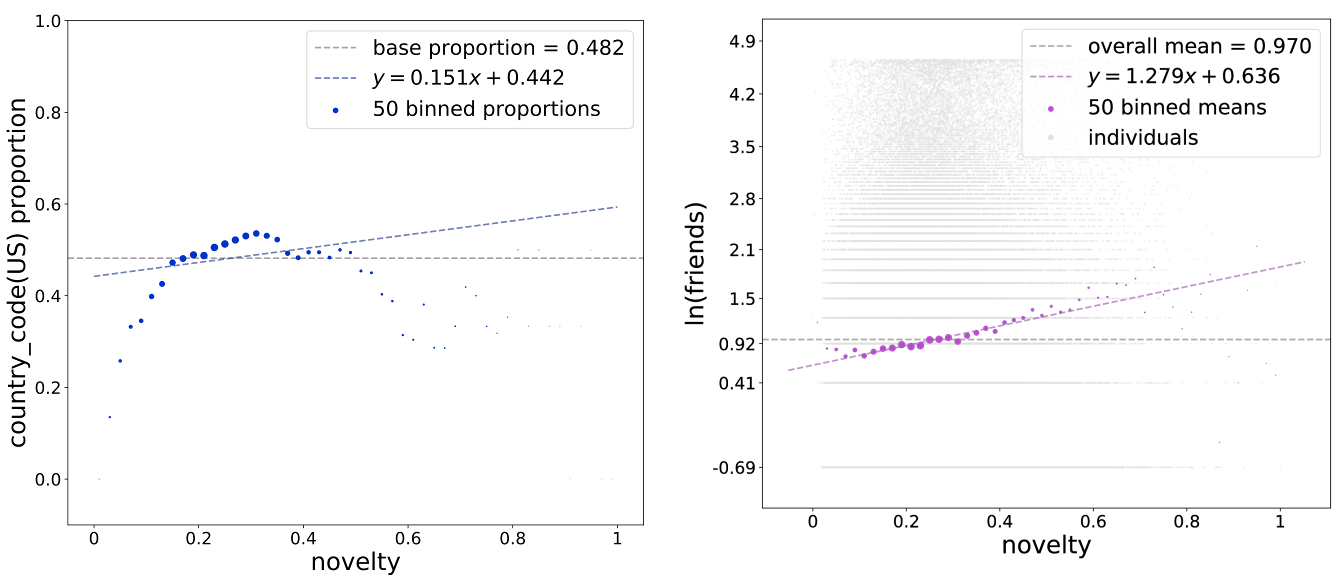

Figure 5. Two graphs showing relationships with novelty. The left graph shows the proportion of US (vs. JP) profiles, while the right graph shows the logarithm of a profile's friends count. The plots group the novelty axis into 50 equally-sized bins, and the height of each point represents the proportion of U.S. profiles (or average log-friends) within that bin, while the size of each point represents the number of items in that bin (dot areas are proportional to bin counts). Most items fall approximately within the 0.1–0.4 range of novelty values. The colored lines are regression lines for the raw data (not the binned proportions plotted), while the dotted grey lines show the overall proportion (or mean) of the full dataset. The right-hand graph includes the raw data in grey (with significant overplotting).

games as a preferred genre than the lower-novelty user.

Figure 5 plots two relationships, including regression lines used to calculate effect sizes. Although country code in Fig. 5 demonstrates a clearly non-linear relationship, most of our variables produced graphs (not shown) more similar to the log-friends graph indicating linear (albeit still noisy) trends across novelty values. Summarizing Table I, we found that country code had a nonlinear relationship with novelty, while the various skill categories taken together indicate that increased novelty is correlated with increased expertise. Likewise, the social variables indicated a significant relationship between novelty and sociability (with substantial effect sizes) and novelty was also correlated with an increase in genre preferences, with especially strong relationships with the RPG and Music genres, and an inverted relationship with the Sports genre.

V. Discussion

A. Customization and Sociability

The main result from our novelty analysis is a clear link between novelty and sociability. Friend count, following count, followers count, number of posts, and number of yeahs are all positively correlated with novelty, and although these are noisy effects, their effect sizes are substantial. Self-reported expertise is also correlated with novelty, which points to a link with engagement. These effects make sense if there is a link between novelty and customization effort (and the grouping of default Miis into the low end of the novelty spectrum means that there is at least some relationship).

Existing research shows that avatar appearance and customization are related to engagement in games [8], [17], [38], and in fact both social interaction and appearance customization are motivating factors for online play [73]. Additionally, avatar appearance has been directly linked to social activity and network size in non-game social networks [16], and other studies suggest multiple possible links between social network size and avatar customization, including simple peer pressure [18]. There is even evidence specific to the Nintendo Mii avatar system that users may prefer some associations over others based on avatar appearance [74], and that expectations about future interaction influence avatar creation [75]. In fact, [75] and [50] report anecdotal evidence that Mii creation can be seen as either a distraction from or a central attraction of the Wii gaming experience (see pages 7 and 192 of the respective studies).

All of this evidence points to links between avatar customization and sociability, and our findings corroborate this. The further implication is that novelty as a metric does provide at least an approximate measure of avatar customization.

B. The Novelty Metric

One of our main research questions was whether our novelty metric would provide a useful way to understand variation within our dataset. Looking at the visualization in Figure 3, we are reassured that the novelty measure is able to distinguish between a common core of avatars close to the default features and a variety of other avatars ranging from relatively minor modifications to distinctive and exaggerated faces. Furthermore, visualizations of individual output dimensions (not shown) indicate that the learned features index meaningful concepts such as shirt- or hair-color, such that images difficult for the network to reconstruct would likely also be judged as more novel by humans.[10] The relationships between novelty, social activity, and self-reported expertise also show that the metric provides useful insight into our data.

Given the completely unsupervised nature of our novelty metric, it can be used to rank any kind of image dataset, and does not require manual labelling. We also expect that it can be extended to other kinds of data by applying alternate network setups. The main utility of our technique lies in its ability to foreground the margins of a dataset. As demonstrated by our exemplar analysis, the ability to analyze common and uncommon segments of the dataset separately can reveal patterns that would be impossible to detect when just examining central tendencies.

C. National Trends

Figure 4 shows the top 16 exemplars identified from the U.S. and Japan subsets of the data within both the low-novelty and high-novelty groups. What is immediately apparent from the low-novelty images is a significant appearance trend which follows demographic divides. Of the 16 top U.S. exemplars, nine have blond(e) hair, a trait that is much more prevalent in the U.S. than in Japan.[11] Given that creating avatars which reflect one's real-life appearance is a common mode of self-expression [54], it is entirely unsurprising that demographic differences should be apparent from our exemplars. The presence of other exemplars with features that differ between the U.S. and Japan regions, such as brown skin, reinforces this point.[12]

Another interesting aspect of the low-novelty exemplars is the presence of blank-faced Miis in both sets (the 16th JP and 7th U.S. exemplars in the top half of Fig. 4). These are the aforementioned counter-examples to the link

[10]Comparing our results with human judgements is a critical piece of future work; there are also some rare but direct counter-examples.

[11]Although representative data on hair color are difficult to come by, National Longitudinal Study of Youth data from 1985 [76], [77] suggest that perhaps 10% of the population has blonde hair in the U.S., while a recent article from Japan notes that many high schools in Tokyo require parents to certify their children's hair color if it is not black [78], with the reported number of forms turned in suggesting a much lower prevalence.

[12]Note that the presence of brown-skinned Miis in the high-novelty U.S. group exemplifies the machine learning tendency to amplify biases [79], [80]. Because brown-skinned Miis are rare and quite different from the default Mii in raw pixels, the algorithm classifies them as more-novel than similar tan-skinned Miis, even though both might be equivalently realistic self-representations. The novelty metric is biased towards the predominant features of the population's self-representations, which are in turn biased by visual norms from other media and selection bias in determining the population. Our technique is thus an example of inequitable technology, similar to Kodak film from the middle of the 20th century [81], and results should be interpreted with this in mind.

between novelty and customization: these near-featureless Miis require clever customization, but nonetheless have low novelty scores. The network is able to reconstruct these faces with little error not because it optimized for them, but as a side effect of optimizing for the full dataset. Although these examples cut against the notion that novelty is linked to customization effort, the number of such examples (perhaps a few dozen) is so small that our statistics are not significantly affected.

These anomalous exemplars are examples of creative self-expression: they are not realistic, nor do they reference shared media or stereotypes, but instead represent a kind of play within the avatar-creation system. Apparently, this particular idea is popular in both Japan and the U.S., but its implementation differs by region, hence the different exemplars. A similarly creative example is the second high-novelty Japan-region exemplar, which has a blacked-out face achieved using creatively-positioned sunglasses and eyebrows. Its presence as an exemplar with 130 neighbors (Japan-region Miis more-similar than the most-similar U.S. Mii) indicates that this Mii is somehow region-specific, hinting that the people creating similar Miis are not simply coming up with the same idea independently. Indeed, a little searching finds YouTube videos showing how to create Miis like this; among videos found when searching for “monster Mii” on YouTube, the video with a mostly Japanese title and which uses the Japanese 3DS interface [82] has more than 380,000 views, while a similar video on the first page of results with an English title and interface [83] has only about 32,000 views. This suggests one possible explanation for the region-specific popularity of this Mii design: language-specific external media may help create regional differences in aesthetics.

The final category of self-expression that we expect to see is idealistic self-expression, where avatars reference ideals and/or paragons, which may be nation-specific (see [56] for a discussion of ideals and paragons). Looking at the Japan-region side of the high-novelty exemplars, we can see one such trend centering around cute-looking Miis, especially those with large eyes and very small mouths and noses. Unlike the specific method for the creation of a pure-black face by carefully positioning features in unconventional ways, a “cute” aesthetic is more general, as evidenced by the variety of Japan-region exemplars with a cute appearance. Instead of being a meme with a single inventor, the cute appearance of these Miis can be considered a kind of national aesthetic, driven by Japanese media such as manga and anime. This aesthetic has been identified by other researchers (e.g., [84], [85]) as an enduring theme within Japanese popular culture, so finding it among the Japan-region exemplars is unsurprising.

D. Exemplar Analysis

The larger takeaway from our exemplar analysis is that avatar images alone contain rich information about the specific contexts of their creation. The nation-specific trends in realistic, idealistic, and creative self-expression that appear among exemplars show the promise of this differential clustering technique for analyzing group-specific aesthetics. Separating the Miis by novelty values was also an important first step, as the exemplars reveal different kinds of self-expression between the low- and high-novelty groups. Our technique is able to identify the most salient divergent images even in datasets that have a lot of overlap, and those reveal important trends among the groups being analyzed. Because of this, it should be generally useful in analyzing data with known class where the classes are not separable into clusters, and as it works at the feature level, it can potentially be applied to other kinds of data where a feature space can be established.

VI. Conclusion

When approaching a dataset containing hundreds of thousands of images, aggregate analysis of the raw data is infeasible for humans. We have developed a new analysis approach which uses a deep neural network trained as an autoencoder to assign novelty values to each image, while at the same time producing a sparse encoding for the data. Using this method, we found links between expertise, sociability, and novelty; these can be explained by existing literature if novelty is seen as a proxy for the player’s customization effort. Our technique groups Miis similar to the default together on the low-novelty end of the spectrum, while Miis with exaggerated and creatively placed features are mostly assigned high novelty values. This novelty dimension can be used to separate low- and high-novelty groups for further analysis.

To find trends in self-expression by nationality, we proceeded to apply a new differential clustering technique to find exemplars from the low- and high-novelty groups, and these exemplars included examples of three key modes of self-expression. Realistic self-expression echoed demographic differences between countries, especially in the low-novelty group. At the same time, idealistic self-expression was identifiable in the high-novelty exemplars, with cute Miis from the Japan region being a specific example. Finally, some Miis were the result of individual creativity, but showed up as exemplars because of region-specific propagation (e.g., via language-specific YouTube videos).

As an unsupervised machine learning technique, our novelty metric is generalizable, and our method for identifying exemplars can likewise be applied to any feature space, such as those that result from latent semantic analysis. These tools, like any machine learning technique, must be used carefully because they tend to amplify and disguise the biases that manifest in their input data (see footnote 12 above). However, they can help to organize and understand otherwise unwieldy image datasets, especially when central tendencies trend towards an uninformative default, or when multiple classes of interest overlap to the point where standard clustering techniques break down.

Ultimately, our analysis makes concrete the potential of culturally-grounded virtual identity system design. Realistic, idealistic, and creative self-expression manifest differently in different cultures and regions, giving rise to diverse needs and requirements. Accordingly, system designers need to be aware of design principles that support and empower diverse communities in digital self-expression.

Acknowledgements

This research was funded by the Qatar Computing Research Institute (QCRI)-MIT CSAIL research collaboration.

References

[1] D. F. Harrell and S. V. Harrell, “Imagination, computation, and self-expression: Situated character and avatar mediated identity”, *Leonardo Electronic Almanac*, vol. 17, no. 2, 2012.

[2] D. F. Harrell and C.-U. Lim, “Reimagining the avatar dream: Modeling social identity in digital media”, *Comm. ACM*, vol. 60, no. 7, pp. 50–61, Jun. 2017, ISSN: 0001-0782. DOI: 10.1145/3098342. [Online]. Available: http://doi.acm.org/10.1145/3098342.

[3] K. L. Nowak and C. Rauh, “The influence of the avatar on online perceptions of anthropomorphism, androgyny, credibility, homophily, and attraction”, *Journal of Computer-Mediated Communication*, vol. 11, no. 1, pp. 153–178, 2005, ISSN: 1083-6101. DOI: 10.1111/j.1083-6101.2006.tb00308.x. [Online]. Available: http://dx.doi.org/10.1111/j.1083-6101.2006.tb00308.x.

[4] T. C. Peck, S. Seinfeld, S. M. Aglioti, and M. Slater, “Putting yourself in the skin of a black avatar reduces implicit racial bias”, *Consciousness and Cognition*, vol. 22, no. 3, pp. 779–787, 2013.

[5] R. A. Dunn and R. E. Guadagno, “My avatar and me–Gender and personality predictors of avatar-self discrepancy”, *Computers in Human Behavior*, vol. 28, no. 1, pp. 97–106, 2012.

[6] E. Dean, S. Cook, M. Keating, and J. Murphy, “Does this avatar make me look fat? Obesity and interviewing in Second Life”, *Journal For Virtual Worlds Research*, vol. 2, no. 2, 2009.

[7] D. Kao and D. F. Harrell, “Exploring the effects of dynamic avatars on performance and engagement in educational games”, in *Games+Learning+Society 2016*, 2016.

[8] ——, “Toward avatar models to enhance performance and engagement in educational games”, in *2015 IEEE Conference on Computational Intelligence and Games*, IEEE, 2015, pp. 246–253.

[9] ——, “Toward avatar models to enhance performance and engagement in educational games”, in *2015 IEEE Conference on Computational Intelligence and Games*, IEEE, 2015, pp. 246–253.

[10] Nintendo. (2012). Miiverse social network. The Miiverse network was taken offline by Nintendo on 11/8/2017., [Online]. Available: https://miiverse.nintendo.net/ (visited on 10/16/2017).

[11] S. Strom. (2016). Miitomo reaches 1 million users in Japan, [Online]. Available: http://www.ign.com/articles/2016/03/20/miitomo-reaches-1-million-users-in-japan (visited on 10/16/2017).

[12] T. Whitehead. (2016). Miitomo tops iOS chart in the US and hits three million global downloads, [Online]. Available: http://www.nintendolife.com/news/2016/04/miitomo_tops_ios_chart_in_the_us_and_hits_three_million_global_downloads (visited on 10/16/2017).

[13] Nintendo. (2015). Miiverse Super Mario Maker community, [Online]. Available: https://miiverse.nintendo.net/titles/6437256808751874777/6437256808751874782 (visited on 10/16/2017).

[14] B. Hillier. (2016). Super Mario Maker is pretty damn popular, [Online]. Available: https://www.vg247.com/2016/05/18/super-mario-maker-is-pretty-damn-popular/ (visited on 10/16/2017).

[15] C.-U. Lim and D. F. Harrell, “Modeling player preferences in avatar customization using social network data: A case-study using virtual items in team fortress 2”, in *2013 IEEE Conference on Computational Intelligence in Games*, IEEE, 2013, pp. 1–8.

[16] E. Y. Ong, R. P. Ang, J. C. Ho, J. C. Lim, D. H. Goh, C. Lee, and A. Y. Chua, “Narcissism, extraversion and adolescents'self-presentation on Facebook”, *Personality and Individual Differences*, vol. 50, no. 2, pp. 180–185, 2011.

[17] R. Ng and R. Lindgren, “Examining the effects of avatar customization and narrative on engagement and learning in video games”, in *2013 18^{th} International Conference on Computer Games: AI, Animation, Mobile, Interactive Multimedia, Educational Serious Games*, Jul. 2013, pp. 87–90. DOI: 10.1109/CGames.2013.6632611.

[18] M. M. Strano, “User descriptions and interpretations of self-presentation through Facebook profile images”, *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, vol. 2, no. 2, 2008.

[19] Y.-H. Lee and D. Y. Wohn, “Are there cultural differences in how we play? Examining cultural effects on playing social network games”, *Computers in Human Behavior*, vol. 28, no. 4, pp. 1307–1314, 2012, ISSN: 0747-5632. DOI: https://doi.org/10.1016/j.chb.2012.02.014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0747563212000568.

[20] E. H. Rosch, “Natural categories”, *Cognitive Psychology*, vol. 4, no. 3, pp. 328–350, 1973.

[21] T. Kohonen, “Automatic formation of topological maps of patterns in a self-organizing system”, in 2^{nd} *Scandanavian Conference on Image Analysis*, Espoo, Finland, 1981, pp. 214–220.

[22] K. Fukushima and S. Miyake, “Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position”, *Pattern Recognition*, vol. 15, no. 6, pp. 455–469, 1982.

[23] Y. LeCun, B. E. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. E. Hubbard, and L. D. Jackel, "Handwritten digit recognition with a back-propagation network", in *Advances in Neural Information Processing Systems*, 1990, pp. 396–404.

[24] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by V1?", *Vision Research*, vol. 37, no. 23, pp. 3311–3325, 1997, ISSN: 0042-6989. DOI: http://dx.doi.org/10.1016/S0042-6989(97)00169-7. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0042698997001697.

[25] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks", *science*, vol. 313, no. 5786, pp. 504–507, 2006.

[26] M. Ranzato, C. Poultney, S. Chopra, and Y. LeCun, "Efficient learning of sparse representations with an energy-based model", in *Advances in Neural Information Processing Systems*, 2007, pp. 1137–1144.

[27] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks", in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[28] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks", in 14^{th} *International Conference on Artificial Intelligence and Statistics*, 2011, pp. 315–323.

[29] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization", *Journal of Machine Learning Research*, vol. 12, no. Jul, pp. 2121–2159, 2011.

[30] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders", in 25^{th} *International Conference on Machine Learning*, ser. ICML '08, Helsinki, Finland: ACM, 2008, pp. 1096–1103, ISBN: 978-1-60558-205-4. DOI: 10.1145/1390156.1390294. [Online]. Available: http://doi.acm.org/10.1145/1390156.1390294.

[31] P. M. H. Häkkinen, "Neural network used to analyze multiple perspectives concerning computer-based learning environments", *Quality and Quantity*, vol. 34, no. 3, pp. 237–258, Aug. 2000, ISSN: 1573-7845. DOI: 10.1023/A:1004763519135. [Online]. Available: https://doi.org/10.1023/A:1004763519135.

[32] E. Vega Ezpeleta, "Identifying illicit graphic in the online community using the neural network framework", PhD thesis, Uppsala University, 2017. [Online]. Available: http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-325810.

[33] H.-C. Shin, M. R. Orton, D. J. Collins, S. J. Doran, and M. O. Leach, "Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4d patient data", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1930–1943, 2013.

[34] F. Zhang, B. Du, and L. Zhang, "Saliency-guided unsupervised feature learning for scene classification", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 4, pp. 2175–2184, 2015.

[35] Q. V. Le, "Building high-level features using large scale unsupervised learning", in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 8595–8598. DOI: 10.1109/ICASSP.2013.6639343.

[36] D. F. Harrell, S. Vieweg, H. Kwak, C.-U. Lim, S. Şengün, A. Jahanian, and P. Ortiz, "Culturally-grounded analysis of everyday creativity in social media: A case study in qatari context", in *2017 SIGCHI Conference on Creativity and Cognition*, ser. C&C '17, Singapore, Singapore: ACM, 2017, pp. 209–221, ISBN: 978-1-4503-4403-6. DOI: 10.1145/3059454.3059456. [Online]. Available: http://doi.acm.org/10.1145/3059454.3059456.

[37] N. Yee, J. N. Bailenson, and N. Ducheneaut, "The proteus effect", *Communication Research*, vol. 36, no. 2, pp. 285–312, 2009.

[38] D. Kao and D. F. Harrell, "Exploring the impact of avatar color on game experience in educational games", in *2016 CHI Conference on Human Factors in Computing Systems*, ACM, 2016, pp. 1896–1905.

[39] R. A. Ratan and M. Dawson, "When Mii is me: A psychophysiological examination of avatar self-relevance", *Communication Research*, vol. 43, no. 8, pp. 1065–1093, 2016.

[40] N. Ducheneaut, M.-H. Wen, N. Yee, and G. Wadley, "Body and mind: A study of avatar personalization in three virtual worlds", in *SIGCHI Conference on Human Factors in Computing Systems*, ACM, 2009, pp. 1151–1160.

[41] C.-U. Lim and D. F. Harrell, "Revealing social identity phenomena in videogames with archetypal analysis", in 6^{th} *International AISB Symposium on AI and Games*, 2015.

[42] S. Şengün, "Why do I fall for the elf, when I am no orc myself? the implications of virtual avatars in digital communication", *Comunicação e Sociedade*, vol. 27, pp. 181–193, 2015.

[43] L. Nakamura, "Race in/for cyberspace: Identity tourism and racial passing on the internet", *Works and Days*, vol. 13, no. 2, 1995.

[44] ——, *Cybertypes: Race, Ethnicity, and Identity on the Internet*. New York, NY: Routledge, 2002.

[45] ——, *Digitizing Race: Visual Cultures of the Internet*. Mnneapolis, MN: U of Minnesota Press, 2008.

[46] A. L. Baylor and Y. Kim, "Pedagogical agent design: The impact of agent realism, gender, ethnicity, and instructional role", in *Intelligent Tutoring Systems*, Springer, vol. 3220, 2004, pp. 592–603.

[47] R. E. Guadagno, J. Blascovich, J. N. Bailenson, and C. Mccall, "Virtual humans and persuasion: The effects of agency and behavioral realism", *Media Psychology*, vol. 10, no. 1, pp. 1–22, 2007.

[48] J. A. Pratt, K. Hauser, Z. Ugray, and O. Patterson, "Looking at human–computer interface design: Effects of ethnicity in computer agents", *Interacting with Computers*, vol. 19, no. 4, pp. 512–523, 2007.

[49] R. B. Rosenberg-Kima, E. A. Plant, C. E. Doerr, and A. L. Baylor, "The influence of computer-based model's race and gender on female students' attitudes and beliefs towards engineering", *Journal of Engineering Education*, vol. 99, no. 1, pp. 35–44, 2010.

[50] H. C. Higgins, J. K. Horton, B. C. Hodgkinson, and S. B. Muggleton, "Lessons learned: Staff perceptions of the Nintendo Wii as a health promotion tool within an aged-care and disability service", English, *Health Promotion Journal of Australia*, vol. 21, no. 3, pp. 189–95, Dec. 2010. DOI: 10.1071/HE10189. [Online]. Available: http://www.publish.csiro.au/he/HE10189 (visited on 12/04/2017).

[51] M. B. Brewer and W. Gardner, "Who is this "We"? Levels of collective identity and self representations", *Journal of Personality and Social Psychology*, vol. 71, no. 1, p. 83, 1996.

[52] M. A. Hogg, "Social identity theory", in *Understanding Peace and Conflict Through Social Identity Theory*, Springer, 2016, pp. 3–17.

[53] S. Turkle, "Constructions and reconstructions of self in virtual reality: Playing in the MUDs", *Mind, Culture, and Activity*, vol. 1, no. 3, pp. 158–167, 1994.

[54] C. Neustaedter and E. Fedorovskaya, "Presenting identity in a virtual world through avatar appearances", in *Graphics Interface 2009*, Canadian Information Processing Society, 2009, pp. 183–190.

[55] E. Rosch and B. B. Lloyd, *Cognition and Categorization*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1978, vol. 1.

[56] G. Lakoff, *Women, Fire, and Dangerous Things*. Chicago, IL: University of Chicago Press, 1987.

[57] G. Bowker and S. L. Star, *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: The MIT Press, 1999.

[58] D. F. Harrell, "Computational and cognitive infrastructures of stigma: Empowering identity in social computing and gaming", in 7^{th} *ACM Conference on Creativity and Cognition*, ACM, 2009, pp. 49–58.

[59] D. F. Harrell, "Toward a theory of critical computing: The case of social identity representation in digital media applications", 2009.

[60] Newzoo. (2017). Top 100 countries by game revenues, [Online]. Available: https://newzoo.com/insights/rankings/top-100-countries-by-game-revenues/ (visited on 10/18/2017).

[61] ImageMagick Studio LLC, *ImageMagick open-source image processing toolkit*, 1999. [Online]. Available: http://www.imagemagick.org (visited on 08/30/2017).

[62] W. McKinney, "Data structures for statistical computing in Python", in 9^{th} *Python in Science Conference*, S. van der Walt and J. Millman, Eds., 2010, pp. 51–56.

[63] S. van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, and T. a. Yu, "scikit-image: Image processing in Python", *PeerJ*, vol. 2, e453, Jun. 2014, ISSN: 2167-8359. DOI: 10.7717/peerj.453. [Online]. Available: https://doi.org/10.7717/peerj.453.

[64] J. D. Hunter, "Matplotlib: A 2D graphics environment", *Computing in Science Engineering*, vol. 9, no. 3, pp. 90–95, May 2007, ISSN: 1521-9615. DOI: 10.1109/MCSE.2007.55.

[65] F. Chollet *et al.* (2015). Keras, [Online]. Available: https://github.com/fchollet/keras (visited on 08/28/2017).

[66] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Y. Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems, [Online]. Available: http://tensorflow.org/ (visited on 08/28/2017).

[67] F. Chollet. (Jun. 5, 2016). Building powerful image classification models using very little data, [Online]. Available: https://blog.keras.io/building-powerful-image-classification-models-using-very-little-data.html (visited on 08/28/2017).

[68] ——, (May 14, 2016). Building autoencoders in Keras, [Online]. Available: https://blog.keras.io/building-autoencoders-in-keras.html (visited on 08/28/2017).

[69] K. Pearson, "Note on regression and inheritance in the case of two parents", *Royal Society of London*, vol. 58, pp. 240–242, 1895, ISSN: 03701662. [Online]. Available: http://www.jstor.org/stable/115794.

[70] B. L. Welch, "The significance of the difference between two means when the population variances are unequal", *Biometrika*, vol. 29, no. 3/4, pp. 350–362, 1938, ISSN: 00063444. [Online]. Available: http://www.jstor.org/stable/2332010.

[71] S. Holm, "A simple sequentially rejective multiple test procedure", *Scandinavian Journal of Statistics*, vol. 6, no. 2, pp. 65–70, 1979, ISSN: 03036898, 14679469. [Online]. Available: http://www.jstor.org/stable/4615733 (visited on 12/04/2017).

[72] D. Oyserman, H. M. Coon, and M. Kemmelmeier, *Rethinking individualism and collectivism: Evaluation of theoretical assumptions and meta-analyses*. 2002.

[73] N. Yee, “Motivations for play in online games”, *CyberPsychology & Behavior*, vol. 9, no. 6, pp. 772–775, 2006.
[74] C. P. Principe and J. H. Langlois, “Children and adults use attractiveness as a social cue in real people and avatars”, *Journal of Experimental Child Psychology*, vol. 115, no. 3, pp. 590–597, 2013.
[75] C. McLaughlin, J. Vitak, and J. Crouse, “Online identity construction and expectation of future interaction”, in *2011 44th Hawai'i International Conference on System Sciences*, IEEE, 2011, pp. 1–10.
[76] Bureau of Labor Statistics, U.S. Department of Labor, “National Longitudinal Survey of Youth 1979 cohort”, 1985, (round 7) Produced and distributed by the Center for Human Resource Research, The Ohio State University. Columbus, OH: 2014.
[77] ——, (2017). NLS investigator, [Online]. Available: https://www.nlsinfo.org/investigator/pages/login.jsp (visited on 10/19/2017).
[78] S. Doi and I. Minetoshi. (2017). Survey: 57% of Tokyo high schools demand hair-color proof, [Online]. Available: http://www.asahi.com/ajw/articles/AJ201705010035.html (visited on 10/19/2017).
[79] K. Crawford. (2016). Artificial intelligence's white guy problem, [Online]. Available: https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html (visited on 10/25/2017).
[80] T. Bolukbasi, K.-W. Chang, J. Zou, V. Saligrama, and A. Kalai, “Quantifying and reducing stereotypes in word embeddings”, in *2016 ICML Workshop on #Data4Good: Machine Learning in Social Good Applications*, New York, NY, 2016.
[81] L. Roth, “Looking at shirley, the ultimate norm: Colour balance, image technologies, and cognitive equity”, English, *Canadian Journal of Communication*, vol. 34, no. 1, pp. 111–136, 2009, ISSN: 07053657. [Online]. Available: https://search.proquest.com/docview/219534659?accountid=12492 (visited on 12/04/2017).
[82] 麦茶。. (2014). 「3DS」未確認生物の作り方!! (mii スタジオ), [Online]. Available: https://www.youtube.com/watch?v=Tc4vE6h_B8I (visited on 10/19/2017).
[83] purplemonkey582. (2011). How to make scary halloween Mii (black monster), [Online]. Available: https://www.youtube.com/watch?v=EVzD3-AzApk (visited on 12/04/2017).
[84] S. Kinsella, “Cuties in Japan”, *Women, Media and Consumption in Japan*, pp. 220–254, 1995.
[85] C.-Y. C. Cheng and H.-Y. Chen, “The possibility of cultural proximity between Taiwan, Singapore and Japan in kawaii culture”, in *2014 3rd International Conference on Innovation, Communication and Engineering*, CRC Press, Guiyang, Guizhou, PR China, 2015, p. 417.