

Identifying Regional Trends in Avatar Customization

Peter Mawhorter¹, Sercan Şengün², Haewoon Kwak³, and D. Fox Harrell⁴

Abstract—Since virtual identities such as social media profiles and avatars have become a common venue for 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 that reflect the needs and preferences of diverse users, understanding how the virtual identity construction differs between groups is important. This paper presents a new methodology that leverages deep learning and differential clustering 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 novelty discovery to segment the avatars, then cluster avatars by region to identify visual trends among low- and high-novelty avatars. We find that avatar customization correlates with increased social activity, and we are able to identify distinct visual trends among the U.S.-region and Japan-region profiles. Among these trends, realistic, idealistic, and creative self-representation can be distinguished. We observe that the realistic self-expression mirrors regional demographics, idealistic self-expression reflects shared mass-media tropes, and creative self-expression propagates within the communities.

Index Terms—Artificial neural networks, avatars, clustering algorithms, cultural differences, data analysis, deep learning, image processing, unsupervised learning.

I. INTRODUCTION

VIRTUAL identity images—ranging from avatars in video games 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 video games. 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, up to the limits imposed by the sociotechnical affordances of the underlying

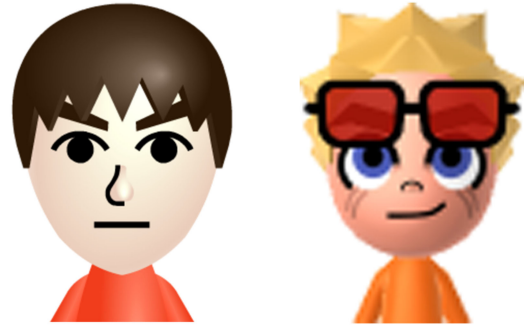


Fig. 1. Default male Mii avatar (left), and a Mii customized to look like a specific fictional character (right). Note the cartoonish style and diverse facial features available.

systems [2]. In any such system, one would expect different user groups to customize their avatars differently. Quantifying those preferences could help the designers to understand how virtual identity systems can better support diverse users, as well as potentially reveal information about the users themselves. Our main research question is, thus, how can we quantify differences in avatar image customization between groups?

To discover how different groups customize their avatars differently, we use neural-network-based novelty detection to segment avatar images according to their relative novelty (see Fig. 2), and then identify differences in customization between the groups of users by finding group-specific avatar clusters. To demonstrate the value of this approach, we have applied it to profiles from Nintendo’s Miiverse network (see Fig. 1) [3] to identify differences in avatar customization between users in the United States and in Japan. Among the resulting differences, customization based on the realistic, idealistic, and creative self-expression can be distinguished. We were also able to find several interesting correlations between novelty and profile information such as one’s number of friends or interest in certain genres. In particular, we assert that our novelty dimension, which measures how well one avatar can be reproduced 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., [4]–[7]).

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 Section III. Meanwhile, the details of the relationships that we found are discussed in Section IV.

Manuscript received December 13, 2017; revised March 12, 2018; accepted May 3, 2018. Date of publication May 15, 2018; date of current version December 23, 2019. This work was supported by the Qatar Computing Research Institute (QCRI)-MIT CSAIL research collaboration. (Corresponding author: Peter Mawhorter.)

P. Mawhorter and S. Şengün are with the Computer Science and Artificial Intelligence Laboratory, Cambridge, MA 02139 USA (e-mail: pmwh@mit.edu; sengun@mit.edu).

H. Kwak is with the Qatar Computing Research Institute, Hamad Bin Khalifa University, Education City, Doha 34110, Qatar (e-mail: haewoon@acm.org).

D. F. Harrell is with the Comparative Media Studies Program Massachusetts Institute of Technology, Cambridge, MA 02139, USA and also with the Computer Science and Artificial Intelligence Laboratory, Cambridge, MA 02139 USA (e-mail: fox.harrell@mit.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TG.2018.2835776

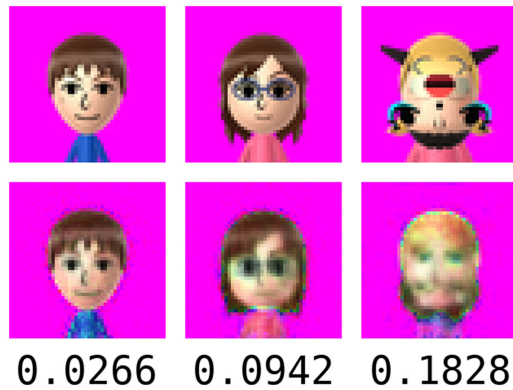


Fig. 2. Low-novelty, medium-novelty, and high-novelty Miis (top) along with their reconstructions by our network (bottom) and corresponding prenormalization novelty values (units are per-pixel-channel root-mean-squared error of each reconstruction in 255-value HSL). The left Mii is very close to the default male Mii, while the center Mii is more novel in part because there are relatively few female-presenting avatars in the dataset. 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.

II. RELATED WORK

A. Deep Learning

Deep learning is a general term that describes machine learning architectures that learn hierarchical representations directly from raw data [8]. In our work, the novelty analysis mechanism passes images through a total of fourteen network layers (hence “deep”) in order to produce both a 128-dimensional compressed image encoding (an automatically learned abstract representation) and a reconstruction attempt that is used to train the network. This is only feasible because of recent advances in network training that have made it much easier to train complex networks and get sensible results from the process.

The roots of deep learning can be traced back to early multilayer networks from the 1980s and 1990s, including self-organizing maps [9], [10], which led to the development of convolutional neural networks [11]. The key problems, such as regularization for sparse coding [12], had already been addressed by the turn of the millennium. However, the problem of vanishing gradients was a challenge for these early multilayer 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 [13]. This breakthrough and other work in the 2000s (e.g., [14]) 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 [15] 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) [16] and AdaGrad optimization [17]. We also use a stacked autoencoding architecture similar to that of Vincent *et al.* [18] (although we do

not incorporate denoising), and we use output normalization as mentioned above [12].

B. Novelty Detection

Our use of an autoencoder network to rank the novelty of images is an extension of existing neural-network-based novelty detection schemes [19], [20]. In particular, Sommer *et al.* [20] apply a similar deep learning strategy to discover anomalous phenotypes from microscope images of cells. An overview of the literature on novelty detection [21] highlights one important departure of our work in terms of perspective, however, while novelty detection work is largely focused on separating “normal” from “abnormal” instances, it would be a mistake to view some avatars as “more normal” than others. In fact, we find interesting patterns among both low-novelty and high-novelty segments, and when graphing novelty-versus-image-count, there is no clear point of division between the “normal” and “abnormal” avatars. Following [19], we process novelty as a spectrum, rather than a binary, and our findings show that analysis can productively include all parts of the spectrum.

C. Neural Network Applications

Other applications of deep convolutional networks are also relevant to our work. 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 [22]. Another similar application is that of Vega Ezpeleta, who trained a network to identify images of swastikas in profile avatars [23]. 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.

D. Virtual Identity

Analyzing virtual identity systems can reveal real-world phenomena, because the users project their identities onto their avatars. We use the term virtual identity within a broad framework that includes social media accounts, video game 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 [24]. 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], [25]. Moreover, avatars have considerable effects on the performance and engagement of individuals in the virtual world [5], [26]–[29]. Analyzing avatar appearance can also reveal norms and stereotypes around issues like body image, gender [30], [31], and other aspects of social identity [32]–[34]. Digitally mediated self-expression ultimately has important ramifications for cultural

dynamics, including ethnic self-expression and the interactions between ethnic groups in virtual spaces [35]–[37].

The availability of options such as gender [38], [39] and ethnicity [40], [41] that facilitate self-resemblance is also empirically linked to user performance in virtual environments. Studies in nongame social networks have revealed links between avatar appearance and both social network activity and connectivity, mediated by personality traits such as extroversion or narcissism [4]. Unsurprisingly, there appear to be social factors that prompt users to engage in avatar creation or customization (see the discussion section in [6], and [42, p. 192]).

E. 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., [43] and [44]). 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 in a social setting. Through ethnographic studies, Turkle makes a distinction between expressing one’s real or idealized self and expressing an alternate identity through role-playing [45]. A more fine-grained distinction is made by Neustaedter and Fedorovskaya [46], which includes a fourfold 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 the 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, the most virtual identities include aspects of all the three forms of self-expression, but in our data we find many that emphasize just one.

F. 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 [47], [48].¹ Along with other more recent work on categorization [49], [50], prototype theory has proved a useful tool for understanding the virtual identities (see, e.g., [51] and [52]). 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.

¹Although 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., [47, p. 331]), we believe that the conclusions they draw remain informative.

III. METHODOLOGY

A. Available Materials

For those interested in reproducing our experiment or using our techniques for other purposes, the source code for our analysis is open source and is available at: <https://github.com/solsword/dewpoint> (the `fractionate.py` script runs the main analysis presented here). Although our dataset is not hosted publicly due to potential privacy concerns, interested researchers are encouraged to get in touch with the first author to explore possibilities for collaboration or data sharing.

B. Data Source

Our data comes from Nintendo’s now-defunct Miiverse social network [3]. The Mii avatars are used across a variety of Nintendo products, and their popularity (both in Japan and globally) is evident from e.g., the recent success of the Miitomo smartphone app [53], [54]. 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 United States and from Japan. 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 [55]. We selected users from the two largest countries on the network (the United States and Japan, which are also two of the largest game markets [56]). We chose Super Mario maker both for its popularity [57] as well as for the way that its focus on user-generated content helps build community. 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 (43 424 from the United States and 46 700 from Japan). Filtering in this way biases our results toward more-active, more-engaged users, which is intentional, although it does limit the scope of our findings.

C. Data Acquisition and Preprocessing

The data for this analysis were scraped from Miiverse using a custom web crawler on August 21, 2016. For each profile, we collected public information including the profile name, personal information (country and birth date), social information (friend count, following count, follower count, post count, and “yeahs”² 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

²“Yeahs,” akin to more well-known “likes,” indicate how many posts a user has indicated their approval for.

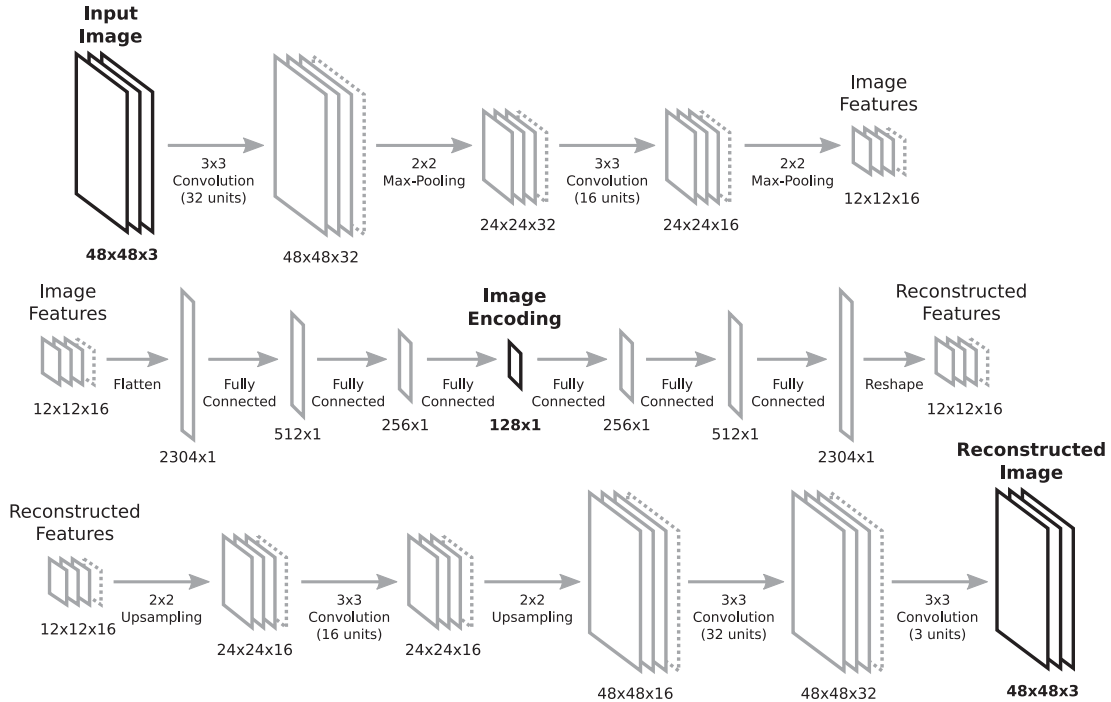


Fig. 3. Network structure used for rating image novelty. The image is encoded to a 128-value vector, which is then decoded to reconstruct the image. L1 regularization is applied to the image encoding during training.

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.” 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 [58]. Each image was resized from 96×96 to 48×48 pixels, and the alpha channel was removed, resulting in a three-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).

For processing in python, we loaded the base data from a single comma-separated value (CSV) file using the pandas scientific data processing library [59]. 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 [60],

[61]. 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.

It is worth noting that even our sample of 90 000 items is large enough to present some difficulty when using modern consumer hardware. For example, a variety of analyses might want to build 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.

D. Network Setup

In order to construct our novelty metric, we employ a deep convolutional neural network similar to the work of [18]. 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 ini-

tialization [16]. We set up the network using the keras library in python, with the tensorflow backend [62], [63].

Fig. 3 shows the network structure we used, including two convolutional layers with ReLU activation followed by max-pooling, and three fully connected layers (also using ReLU activation) to achieve a 128-dimensional encoding, which is a fairly conventional setup (see, e.g., [64] and [65]). On this encoding path, there are a total of 1 349 904 parameters. The choice of 128 output units is arbitrary, but we felt that this gave the network enough room to learn fairly complex representations while still forcing it to achieve significant compression.

In our autoencoder setup, the decoding portion of the network is a mirror image of the encoding part. To get back to a $48 \times 48 \times 3$ output shape, we use a final convolutional layer with 3 units and a 3×3 kernel, using a sigmoid activation function, unlike the other layers. 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 [65]; see [12] 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.

E. Training and Analysis

For our final results, we trained for 100 epochs on our 90 124 example images, using a batch size of 32.³ We used keras' AdaGrad optimizer [17], with the default base learning rate, epsilon and decay parameters (0.01, 1×10^{-7} , and 0, respectively). On a typical run, the network's initial average loss value is ~ 0.015 , after ten epochs it reaches ~ 0.007 , and at the end of training, its loss is ~ 0.0045 . After training, we computed novelty ratings by loading each image, asking the network to reconstruct it, and measuring the root-mean-squared error between the original and its reconstruction (this is the loss function without the L1 regularization term). We then normalized these values across the dataset, so that novelty ratings were between 0 and 1 (the values in Fig. 2 are not normalized). 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 (this is due to the L1 regularization term which penalizes the sum of activations; the balance between sparsity and learning regulates the number of active features regardless of how many are available). Of the remaining 46 features, we pruned 10 more due to monotony, this pruning affected

feature-space coordinates used for differential clustering but not novelty values, which used the full feature set as built into the network. The average monotony value of pruned features that were not single-valued was 99.28%, meaning that on average just 653 avatars affected by the pruning of each of the 10 extra features pruned.

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 [66] between novelty and that variable, or for boolean variables, a Welch's *t*-test [67] for unequal mean novelty values between the true and false cases (we used the `pearsonr` and `ttest_ind` functions from `scipy.stats` version 0.19.1 to compute these statistics). We performed a total of 21 separate tests against our novelty value, each using the filtered dataset, for $n = 90124$. After computing *p* values for each of our tests, we used the Holm-Bonferroni correction for multiple comparisons [68] 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.

F. Differential Clustering

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 regional⁴ trends within each group, we searched for the avatars which were exemplars in either the U.S. or Japan subsets. We used a novel differential clustering technique to find these exemplars: We sorted each group according to both distance from the nearest image from the other region (Euclidean in feature space) and the number of same-region images within that radius.

The former value, termed *separation*, measures how distinct an image is from images from the other region, and maxima should be found within region-specific groups, or else be outliers in the dataset. The latter value, termed *centrality*, distinguishes

³Running on a laptop with an NVidia GeForce GTX 950M graphics card, an Intel Core i7-6700HQ processor (2.6 GHz), 16 GB of ram, and a 5400 r/min HDD, training took 3–4 h. Although this method is time-consuming, it does not require specialized hardware.

⁴We study trends at the regional level because the data are readily available. As pointed out in, e.g., [69], this has its limitations, because these regions are multicultural. What we find are prominent trends, not universal properties.

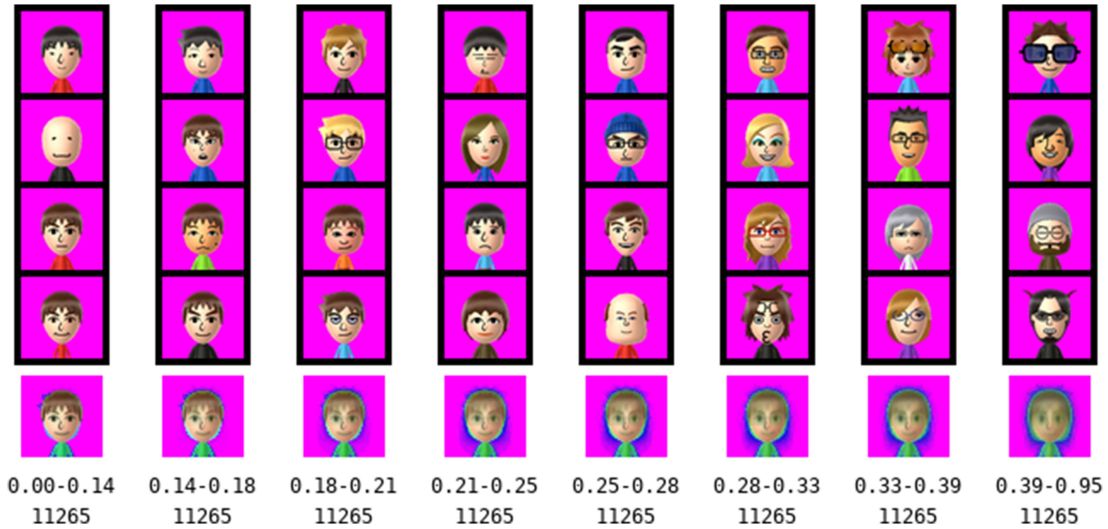


Fig. 4. Image lineup for the novelty value. Columns include four random samples from a percentile plus a mean across 2500 random samples from that percentile (the bottom row). Labels are the range of values represented within that percentile (top) and the total number of images in that percentile (bottom). 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).

these cases, where maxima are images which follow a regional theme.

After sorting images separately in descending order by separation and centrality, 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. This iterative procedure identified clusters that were unique to each group (in our case, each region), hence the term "differential clustering." We proceeded to display the top eight exemplars from each category, as shown in Fig. 5. These exemplars revealed visual aesthetics specific to each region, some 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. 6. In the table, nonsignificant 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, in 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, the linear changes in the logistic variables have a multiplicative rather than an additive effect. Also note that the relationships

in some cases were not strictly linear, but the regression slope gives a general indicator of magnitude.

Overall, the effect sizes are noticeable, especially when extended to the full range of novelty values ($10\times$ 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). 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 games as a preferred genre than the lower-novelty user.

Fig. 6 plots two relationships, including regression lines used to calculate effect sizes. Although country code in Fig. 6 demonstrates a clearly nonlinear 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 correlation between novelty and sociability. The friend count, follower count, following count, number of posts, and number of yeahs are all positively correlated with novelty, and although these



Fig. 5. Top eight low-novelty (top) and high-novelty (bottom) exemplars from the Japan (left) and U.S. (right) subsets. Numbers indicate centrality (number of same-region Miis nearby) and separation (feature-space distance to the nearest different-region Mii).

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 correlation 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). Of course, this result is confined to users with public profiles, since we do not have data on the social networks or activity of private profiles.

Existing research shows that the avatar appearance and customization are related to engagement in games [5], [26], and [28], and in fact both the social interaction and appearance customization are motivating factors for online play [70]. Additionally, the avatar appearance has been directly linked to social activity and network size in nongame social networks [4], and other studies suggest multiple possible links between social network size and avatar customization, including simple peer pressure [6]. There is even evidence specific to the Nintendo Mii avatar system that users may prefer some associations over others based on avatar appearance [71], and that expectations about future interaction influence avatar creation [72]. In fact, McLaughlin *et al.* [72] and Higgins *et al.* [42] 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 multiple links between sociability and avatar customization, and our findings corroborate this (of course, the question of causation is unclear, and there may in fact be multiple influences in both directions). The further implication is that novelty as a metric does provide at least an approximate measure of avatar customization.

B. Novelty Metric

For our technique to find trends in avatar customization, the novelty metric must successfully capture important variation within the dataset. Looking at the visualization in Fig. 4, 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, the visualizations of individual output dimensions (not shown) indicate that the learned features index meaningful concepts such as shirt- or hair-color, such that the images difficult for the network to reconstruct would likely also be judged as more novel by humans (comparing our results with human judgements is a critical piece of future work; there are also some rare, but direct, counter-examples). 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 labeling. 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

Fig. 5 shows the top eight 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 eight top U.S. exemplars, seven have blond(e) hair, a trait that is much more prevalent in the United States than in Japan.⁵ Given that creating avatars which reflect one’s real-life appearance is a common mode of

⁵Although representative data on hair color are difficult to come by, National Longitudinal Study of Youth data from 1985 [73], [74] suggest that perhaps 10% of the population has blonde hair in the United States, 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 [75], suggesting a much lower prevalence.

TABLE I
STATISTICAL RESULTS

Variable	<i>p</i> 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%	
$\ln(\text{friends})$	2.7×10^{-183}	$r = 0.096$	$\times 1.14$ friends	Novelty (\sim customization) is correlated with social network size and activity.
$\ln(\text{following})$	8.1×10^{-206}	$r = 0.102$	$\times 1.18$ following	
$\ln(\text{followers})$	2.3×10^{-241}	$r = 0.11$	$\times 1.19$ followers	
$\ln(\text{posts})$	8.6×10^{-75}	$r = 0.0609$	$\times 1.06$ posts	
$\ln(\text{years})$	8.2×10^{-117}	$r = 0.0764$	$\times 1.17$ years	
⟨no genres⟩	7.0×10^{-28}	$\delta_\mu = -0.00951$	-1.4 ⟨no genres⟩%	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(Action)	0.083	—	—	
genres(Adventure)	0.1	—	—	
genres(Fighting)	0.0023	$\delta_\mu = 0.00277$	+0.38 genres(Fighting)%	
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)%	

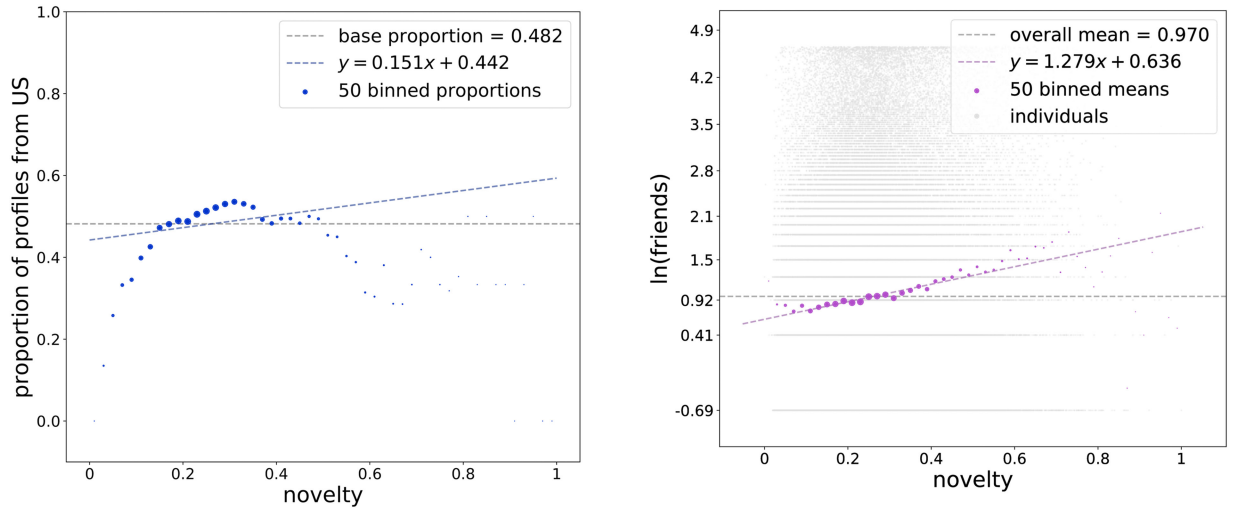


Fig. 6. Two graphs showing relationships with novelty. The left graph shows the proportion of U.S. (versus Japanese) profiles, while the right graph shows the logarithm of a profile's friends count. The plots group the novelty axis into 50 equally sized bins, and the height of each point represents the 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 gray lines show the overall proportion (or mean) of the full dataset. The right-hand graph includes the raw data in gray (with significant overplotting).

self-expression [46], it is entirely unsurprising that demographic differences should be apparent from our exemplars. The presence of other exemplars with features that differ between the United States and Japan regions, such as brown skin, reinforces this point.⁶

⁶Note that, the presence of brown-skinned Miis in the high-novelty U.S. group exemplifies the machine learning tendency to amplify biases [76], [77]. Because the brown-skinned Miis are rare and quite different from the default Mii in raw

Another interesting aspect of the low-novelty exemplars is the presence of a blank-faced Mii (the 7th U.S. exemplar in the top half of Fig. 5). This is one of the aforementioned counter-

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 toward 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 [78], and results should be interpreted with this in mind.

examples to the link between novelty and customization: This near-facelss Mii requires clever customization, but nonetheless has a low-novelty score. The network is able to reconstruct this face with little error not because it is common, but as a side effect of optimizing for the full dataset. Although this example cuts against the notion that novelty is linked to customization effort, the number of such examples (probably less than a hundred) 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 common themes, but they instead represent a kind of play within the avatar-creation system. By inspecting more exemplars than shown in Fig. 5, we have found that this particular idea appears in both the regions, but its specifics differ by region. A similarly creative example is the second high-novelty Japan-region exemplar, which has an all-black 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 mostly Japanese title and which uses the Japanese 3DS interface [79] has more than 380 000 views, while a similar video on the first page of results with an English title and interface [80] 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 drive regional differences in aesthetics.

The final category of self-expression that we expect is the idealistic self-expression, where avatars reference ideals and/or paragons, which may be region-specific (see [49] 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 regional aesthetic, possibly driven by Japanese media, such as manga and anime. This aesthetic, known as “可愛い” (“kawaii,” which translates literally as “cute”) has been identified by other researchers (e.g., [81] and [82]) as an enduring theme within the 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 region-specific trends in the 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, analysis of the raw data is infeasible for humans. We have developed an analysis approach which uses a deep neural network for novelty discovery and to produce a sparse encoding for the data. Using this method, we found links between expertise, sociability, and novelty, these can be explained by the existing literature if novelty is seen as a proxy for the player’s customization effort. Miis similar to the default are grouped 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 region, we proceeded to apply a 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. The realistic self-expression echoed demographic differences between countries, especially in the low-novelty group. At the same time, the 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, novelty analysis is generalizable to other datasets, 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 6). However, they can help to organize and understand otherwise unwieldy image datasets, especially when central tendencies trend toward an uninformative default, or when multiple classes of interest overlap to the point where standard clustering techniques break down.

Ultimately, the combination of novelty segmentation and differential clustering was able to successfully reveal distinct usage patterns associated with the different regions in our dataset. Our analysis of the results revealed ways in which the realistic, idealistic, and creative self-expression manifest differently in different regions. Accordingly, system designers need to be aware of design principles that support and empower diverse communities in digital self-expression.

REFERENCES

- [1] D. F. Harrell and S. V. Harrell, "Imagination, computation, and self-expression: Situated character and avatar mediated identity," *Leonardo Electron. Almanac*, vol. 17, no. 2, pp. 74–91, 2012.
- [2] D. F. Harrell and C.-U. Lim, "Reimagining the avatar dream: Modeling social identity in digital media," *Commun. ACM*, vol. 60, no. 7, pp. 50–61, Jun. 2017. [Online]. Available: <http://doi.acm.org/10.1145/3098342>
- [3] Nintendo. (2012) Miiverse social network. The Miiverse network was taken offline by Nintendo on 11/8/2017.
- [4] E. Y. Ong *et al.*, "Narcissism, extraversion, and adolescents self-presentation on Facebook," *Personal. Individ. Differ.*, vol. 50, no. 2, pp. 180–185, 2011.
- [5] R. Ng and R. Lindgren, "Examining the effects of avatar customization and narrative on engagement and learning in video games," in *Proc. 18th Int. Conf. Comput. Games: AI, Animat., Mobile, Interactive Multimedia, Edu. Serious Games*, Jul. 2013, pp. 87–90.
- [6] M. M. Strano, "User descriptions and interpretations of self-presentation through Facebook profile images," *Cyberpsychol.: J. Psychosocial Res. Cyberspace*, vol. 2, no. 2, 2008. [Online]. Available: <https://cyberpsychology.eu/article/view/4212/3253>
- [7] Y.-H. Lee and D. Y. Wahn, "Are there cultural differences in how we play? Examining cultural effects on playing social network games," *Comput. Human Behav.*, vol. 28, no. 4, pp. 1307–1314, 2012. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0747563212000568>
- [8] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [9] T. Kohonen, "Automatic formation of topological maps of patterns in a self-organizing system," in *Proc. 2nd Scandinavian Conf. Image Anal.*, Espoo, Finland, 1981, pp. 214–220.
- [10] K. Fukushima and S. Miyake, "Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position," *Pattern Recognit.*, vol. 15, no. 6, pp. 455–469, 1982.
- [11] Y. LeCun *et al.*, "Handwritten digit recognition with a back-propagation network," in *Proc. Adv. Neural Inf. Process. Syst.*, 1990, pp. 396–404.
- [12] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by V1?" *Vis. Res.*, vol. 37, no. 23, pp. 3311–3325, 1997. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0042698997001697>
- [13] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [14] M. Ranzato, C. Poultney, S. Chopra, and Y. LeCun, "Efficient learning of sparse representations with an energy-based model," in *Proc. Adv. Neural Inf. Process. Syst.*, 2007, pp. 1137–1144.
- [15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
- [16] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. Stat.*, 2011, pp. 315–323.
- [17] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *J. Mach. Learn. Res.*, vol. 12, no. Jul, pp. 2121–2159, 2011.
- [18] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proc. 25th Int. Conf. Mach. Learn.*, 2008, pp. 1096–1103. [Online]. Available: <http://doi.acm.org/10.1145/1390156.1390294>
- [19] S. Hawkins, H. He, G. Williams, and R. Baxter, "Outlier detection using replicator neural networks," in *Proc. Int. Conf. Data Warehousing Knowledge Disc.*, 2002, pp. 170–180.
- [20] C. Sommer, R. Hoefler, M. Samwer, and D. W. Gerlich, "A deep learning and novelty detection framework for rapid phenotyping in high-content screening," *Mol. Biol. Cell*, vol. 28, pp. 3428–3436, 2017. [Online]. Available: <http://www.molbiolcell.org/content/early/2017/09/25/mbc.E17-05-0333.abstract>
- [21] M. A. Pimentel, D. A. Clifton, L. Clifton, and L. Tarassenko, "A review of novelty detection," *Signal Process.*, vol. 99, pp. 215–249, 2014. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S016516841300515X>
- [22] P. M. H. Häkkinen, "Neural network used to analyze multiple perspectives concerning computer-based learning environments," *Quality Quantity*, vol. 34, no. 3, pp. 237–258, Aug. 2000. [Online]. Available: <https://doi.org/10.1023/A:1004763519135>
- [23] E. Vega Ezpeleta, "Identifying illicit graphic in the online community using the neural network framework," Ph.D. dissertation, Uppsala Univ., 2017. [Online]. Available: <http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-325810>
- [24] D. F. Harrell *et al.*, "Culturally-grounded analysis of everyday creativity in social media: A case study in Qatari context," in *Proc. SIGCHI Conf. Creativity Cognition*, 2017, pp. 209–221. [Online]. Available: <http://doi.acm.org/10.1145/3059454.3059456>
- [25] N. Yee, J. N. Bailenson, and N. Ducheneaut, "The proteus effect," *Commun. Res.*, vol. 36, no. 2, pp. 285–312, 2009.
- [26] D. Kao and D. F. Harrell, "Toward avatar models to enhance performance and engagement in educational games," in *Proc. IEEE Conf. Comput. Intell. Games*, 2015, pp. 246–253.
- [27] D. Kao and D. F. Harrell, "Exploring the effects of dynamic avatars on performance and engagement in educational games," in *Proc. Games+Learning+Society*, 2017, pp. 305–316.
- [28] D. Kao and D. F. Harrell, "Exploring the impact of avatar color on game experience in educational games," in *Proc. CHI Conf. Human Factors Comput. Syst.*, 2016, pp. 1896–1905.
- [29] R. A. Ratan and M. Dawson, "When Mii is me: A psychophysiological examination of avatar self-relevance," *Commun. Res.*, vol. 43, no. 8, pp. 1065–1093, 2016.
- [30] R. A. Dunn and R. E. Guadagno, "My avatar and me—Gender and personality predictors of avatar-self discrepancy," *Comput. Human Behav.*, vol. 28, no. 1, pp. 97–106, 2012.
- [31] N. Ducheneaut, M.-H. Wen, N. Yee, and G. Wadley, "Body and mind: A study of avatar personalization in three virtual worlds," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2009, pp. 1151–1160.
- [32] 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 *Proc. IEEE Conf. Comput. Intell. Games*, 2013, pp. 1–8.
- [33] C.-U. Lim and D. F. Harrell, "Revealing social identity phenomena in video games with archetypal analysis," in *Proc. 6th Int. AISB Symp. AI Games*, 2015, pp. 12–17.
- [34] 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.
- [35] L. Nakamura, "Race in/for cyberspace: Identity tourism and racial passing on the internet," *Works Days*, vol. 13, no. 2, pp. 181–193, 1995.
- [36] L. Nakamura, *Cybertypes: Race, Ethnicity, and Identity on the Internet*, New York, NY, USA: Routledge, 2002.
- [37] L. Nakamura, *Digitizing Race: Visual Cultures of the Internet*, Minneapolis, MN, USA: Univ. Minnesota Press, 2008.
- [38] A. L. Baylor and Y. Kim, "Pedagogical agent design: The impact of agent realism, gender, ethnicity, and instructional role," in *Proc. Intell. Tutoring Syst.*, vol. 3220, 2004, pp. 592–603.
- [39] R. E. Guadagno, J. Blascovich, J. N. Bailenson, and C. McCall, "Virtual humans and persuasion: The effects of agency and behavioral realism," *Media Psychol.*, vol. 10, no. 1, pp. 1–22, 2007.
- [40] J. A. Pratt, K. Hauser, Z. Ugray, and O. Patterson, "Looking at human-computer interface design: Effects of ethnicity in computer agents," *Interacting Comput.*, vol. 19, no. 4, pp. 512–523, 2007.
- [41] 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," *J. Eng. Edu.*, vol. 99, no. 1, pp. 35–44, 2010.
- [42] 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," *Health Promotion J. Australia*, vol. 21, no. 3, pp. 189–95, Dec. 2010. [Online]. Available: <http://www.publish.csiro.au/HE/HE10189>
- [43] M. B. Brewer and W. Gardner, "Who is this 'We'? Levels of collective identity and self representations," *J. Personal. Social Psychol.*, vol. 71, no. 1, pp. 83–93, 1996.
- [44] M. A. Hogg, "Social identity theory," in *Proc. Understanding Peace Conflict Through Social Identity Theory*, 2016, pp. 3–17.
- [45] S. Turkle, "Constructions and reconstructions of self in virtual reality: Playing in the MUDs," *Mind, Culture, Activity*, vol. 1, no. 3, pp. 158–167, 1994.
- [46] C. Neustaedter and E. Fedorovskaya, "Presenting identity in a virtual world through avatar appearances," in *Proc. Graph. Interface*, 2009, pp. 183–190.
- [47] E. H. Rosch, "Natural categories," *Cognitive Psychol.*, vol. 4, no. 3, pp. 328–350, 1973.
- [48] E. Rosch and B. B. Lloyd, *Cognition and Categorization*, vol. 1. Hillsdale, NJ, USA: Lawrence Erlbaum Associates, 1978.

- [49] G. Lakoff, *Women, Fire, and Dangerous Things*. Chicago, IL, USA: Univ. Chicago Press, 1987.
- [50] G. Bowker and S. L. Star, *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA, USA: MIT Press, 1999.
- [51] D. F. Harrell, "Computational and cognitive infrastructures of stigma: Empowering identity in social computing and gaming," in *Proc. 7th ACM Conf. Creativity Cognition*, 2009, pp. 49–58.
- [52] D. F. Harrell, "Toward a theory of critical computing: The case of social identity representation in digital media applications," *CTheory*, 2010.
- [53] S. Strom, "Miitomo reaches 1 million users in Japan," 2016. [Online]. Available: <http://www.ign.com/articles/2016/03/20/miitomo-reaches-1-million-users-in-japan>
- [54] T. Whitehead, "Miitomo tops iOS chart in the US and hits three million global downloads," 2016. [Online]. Available: http://www.nintendolife.com/news/2016/04/miitomo_tops_ios_chart_in_the_us_and_hits_three_million_global_downloads
- [55] Nintendo, "Miiverse Super Mario Maker community," 2015.
- [56] Newzoo, "Top 100 countries by game revenues," 2017. [Online]. Available: <https://newzoo.com/insights/rankings/top-100-countries-by-game-revenues/>
- [57] B. Hillier, "Super Mario Maker is pretty damn popular," 2016. [Online]. Available: <https://www.vg247.com/2016/05/18/super-mario-maker-is-pretty-damn-popular/>
- [58] ImageMagick Studio LLC, "ImageMagick open-source image processing toolkit," 1999. [Online]. Available: <http://www.imagemagick.org>
- [59] W. McKinney, "Data structures for statistical computing in Python," in *Proc. 9th Python Sci. Conf.*, 2010, pp. 51–56.
- [60] S. van der Walt *et al.*, "scikit-image: Image processing in Python," *PeerJ*, vol. 2, Jun. 2014, Art. no. e453. [Online]. Available: <https://doi.org/10.7717/peerj.453>
- [61] J. D. Hunter, "Matplotlib: A 2-D graphics environment," *Comput. Sci. Eng.*, vol. 9, no. 3, pp. 90–95, May 2007.
- [62] F. Chollet *et al.*, "Keras," 2015. [Online]. Available: <https://github.com/fchollet/keras>
- [63] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015. [Online]. Available: <http://tensorflow.org/>
- [64] F. Chollet, "Building powerful image classification models using very little data," 2016. [Online]. Available: <https://blog.keras.io/building-powerful-image-classification-models-using-very-little-data.html>
- [65] F. Chollet, "Building autoencoders in Keras," 2016. [Online]. Available: <https://blog.keras.io/building-autoencoders-in-keras.html>
- [66] K. Pearson, "Note on regression and inheritance in the case of two parents," *Roy. Soc. London*, vol. 58, pp. 240–242, 1895. [Online]. Available: <http://www.jstor.org/stable/115794>
- [67] 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. [Online]. Available: <http://www.jstor.org/stable/2332010>
- [68] S. Holm, "A simple sequentially rejective multiple test procedure," *Scandinavian J. Stat.*, vol. 6, no. 2, pp. 65–70, 1979. [Online]. Available: <http://www.jstor.org/stable/4615733>
- [69] D. Oyserman, H. M. Coon, and M. Kemmelmeier, "Rethinking individualism and collectivism: Evaluation of theoretical assumptions and meta-analyses," vol. 128, no. 1, pp. 3–72, 2002.
- [70] N. Yee, "Motivations for play in online games," *CyberPsychol. Behav.*, vol. 9, no. 6, pp. 772–775, 2006.
- [71] C. P. Principe and J. H. Langlois, "Children and adults use attractiveness as a social cue in real people and avatars," *J. Exp. Child Psychol.*, vol. 115, no. 3, pp. 590–597, 2013.
- [72] C. McLaughlin, J. Vitak, and J. Crouse, "Online identity construction and expectation of future interaction," in *Proc. 44th Hawaii Int. Conf. Syst. Sci.*, 2011, pp. 1–10.
- [73] 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, USA: 2014.]
- [74] Bureau of Labor Statistics, "NLS investigator," 2017. [Online]. Available: <https://www.nlsinfo.org/investigator/pages/login.jsp>
- [75] S. Doi and I. Minetoshi, "Survey: 57% of Tokyo high schools demand hair-color proof," 2017. [Online]. Available: <http://www.asahi.com/ajw/articles/AJ201705010035.html>
- [76] K. Crawford, "Artificial intelligence's white guy problem," 2016. [Online]. Available: <https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html>
- [77] T. Bolukbasi, K.-W. Chang, J. Zou, V. Saligrama, and A. Kalai, "Quantifying and reducing stereotypes in word embeddings," in *Proc. ICML Workshop #Data4Good: Mach. Learn. Social Good Appl.*, arXiv: 1606.06121v1, 2016, pp. 41–45.
- [78] L. Roth, "Looking at shirley, the ultimate norm: Color balance, image technologies, and cognitive equity," *Canadian J. Commun.*, vol. 34, no. 1, pp. 111–136, 2009. [Online]. Available: <https://www.cjc-online.ca/index.php/journal/article/view/2196/3069>
- [79] 麦茶 [3DS] 未確認生物の作り方 !! mii スタジオ. 2014. [Online]. Available: https://www.youtube.com/watch?v=Tc4vE6h_B8I
- [80] purplemonkey582, "How to make scary halloween Mii (black monster)," 2011. [Online]. Available: <https://www.youtube.com/watch?v=EVzD3-AzApk>
- [81] S. Kinsella, "Cuties in Japan," in *Women, Media Consumption Japan*, L. Skov and B. Moeran, Eds. Honolulu, HI, Hawaii: Univ. Hawai'i Press, 1995, pp. 220–254.
- [82] C.-Y. C. Cheng and H.-Y. Chen, "The possibility of cultural proximity between Taiwan, Singapore, and Japan in kawaii culture," in *Proc. 3rd Int. Conf. Innovation, Commun. Eng. Guiyang, Guizhou, China: CRC Press*, 2015, pp. 417–419.