6-2017

Culturally-grounded analysis of everyday creativity in social media: A case study in qatari context

D. Fox HARRELL

Sarah VIEWEG

Haewoon KWAK
Singapore Management University, hkwak@smu.edu.sg

Chong-U LIM

Sercan SENGUN

See next page for additional authors

Culturally-Grounded Analysis of Everyday Creativity in Social Media: A Case Study in Qatari Context

Conference Paper · June 2017

DOI: 10.1145/3059454.3059456

CITATIONS
6

READS
257

7 authors, including:

Chong-U Lim
Massachusetts Institute of Technology
11 PUBLICATIONS **270** CITATIONS

SEE PROFILE

Sercan Şengün
Illinois State University
32 PUBLICATIONS **90** CITATIONS

SEE PROFILE

Culturally-Grounded Analysis of Everyday Creativity in Social Media: A Case Study in Qatari Context

D. Fox Harrell
CSAIL, MIT
Cambridge, MA 02139, USA
fox@csail.mit.edu

Sarah Vieweg
QCRI
Hamad Bin Khalifa University
Doha, Qatar
sarahvieweg@gmail.com

Haewoon Kwak
QCRI
Hamad Bin Khalifa University
Doha, Qatar
haewoon@acm.org

Chong-U Lim
Improbable
London, United Kingdom
chong-u@improbable.io

Sercan Sengun
CSAIL, MIT & QCRI
Cambridge, MA 02139, USA
sengun@mit.edu

Ali Jahanian
CSAIL, MIT
Cambridge, MA 02139, USA
ali-design@csail.mit.edu

Pablo Ortiz
CSAIL, MIT
Cambridge, MA 02139, USA
portiz@mit.edu

ABSTRACT

In deploying social media and other information technologies often not designed with MENA (the Middle East and North Africa) cultures in mind, users generate creative approaches to self-representation using virtual identities while preserving their cultural values. To understand and further empower such approaches, we present a mixed-method of computational and qualitative study, focusing on Qatar as a case of such communities in the MENA region. We analyzed a dataset of over 42,000 publicly available social media profiles using computational approaches (archetypal analysis) and qualitatively analyzed a separate set of 255 profiles. We augmented our descriptions with semi-structured interviews. As a result, we delineate a set of five needs/values exhibited by Qatari users supporting their creativity in effectively using virtual identities: Khaleeji[1] features, self-expression, social connections, social monitoring, and physical and virtual identity contrasts. Finally, we propose an initial set of guidelines to support developers of virtual identity systems in better serving these users while preserving their cultural values and creative agency.

ACM Classification Keywords

H.5.3 Information Systems: Information Interfaces and Presentation (e.g., HCI)—*Group and Organization Interfaces*; J.4 Computer Applications: Social and Behavioral Sciences; K.4.0 Computing Milieux: Computers and Society—*General*

Author Keywords

Virtual Identity; Social Media; Avatars; Qualitative Analysis; Archetypal Analysis; Cultural Values; Qatar

INTRODUCTION

The use of virtual identity technologies across platforms such as social media, videogames, and e-commerce sites is pervasive throughout the world. Whether profiles on social media or player characters in videogames, these user representations integrate numerical attributes, graphical images, and related forms of structured data to enable users to communicate, interact, and even express their senses of self [21]. Indeed, virtual identities are sites demonstrating enaction of users' values, from individual preferences to aspects of larger scale cultural norms and mores. Previous research has demonstrated that the way users behave in both the physical and virtual worlds can be influenced by these virtual self representations [58, 22, 48]. Virtual identities, and user behaviors conveyed while deploying them, may reveal aspects of a user's physical-world identity such as their demographic profiles, personality traits, and motivations for play [57, 60]. Furthermore, they may also reveal phenomena that reflect physical-world social constructs, such as conforming to notions of ideal body types or gender-related stereotypes [14, 38, 22, 49].

Here, we contribute an approach to better understanding the relationships between physical-world identities and virtual identities,[2] and the creativity involved in blending between them. The relationship between virtual identity profiles and creativity is underexplored. Forming a virtual identity itself is a creative process in the sense of Goffman's articulation

[1]The term "Khaleeji" means "of the Gulf" in Arabic.

[2]In this paper, "physical-world identity" refers identity experiences that are informed by history, culture, and values in the physical world. We use the term in contrast to "virtual identities," which refer to technical and structural data such as static media assets (e.g., profile pictures) as well as the algorithmic processes for simulating and enacting selves and associated behaviors within a virtual environment.

C+C'17, June 27–30, 2017, Singapore.

ACM ISBN 978-1-4503-4403-6/17/06 ...$15.00.
http://dx.doi.org/10.1145/3059454.3059456

of everyday self-performances that require "*employment of techniques*" to dramaturgically stage one's identity [17]. On one hand, virtual environments may implement restrictive presentation techniques that impair users' self-expression, impression management, and lead to unwarranted stigmatization [20]. On the other hand, virtual identities can impact user's performance and engagement within virtual environments [24] in addition to their creativity [28, 27].

Our aims here are (1) to provide a mixed-methods-based case study accounting for users' needs and values regarding virtual identities in a principled and systematic manner, and (2) to develop guidelines for best practices and design principles for virtual identity system design that support how users creatively maintain cultural values. Toward these ends, we describe how virtual identity systems impact everyday creative self-presentation and the diverse strategies that users form cross-culturally to navigate these impacts. Our case study focuses on the Middle East and North Africa (MENA) region broadly as a set of related underserved communities in terms of culturally-grounded virtual identity systems, and specifically analyzes the Gulf Cooperation Council (GCC) context with a special emphasis on Qatar. To give a few examples of such creative virtual identity usages in the region, when deploying a social media profile image online a mother may use her son's photo, a woman attending a university might use a landscape photo, and anyone may use a cartoony emoji-style image, even in business settings. While such choices could occur in other regions too, here we present techniques to reveal culturally-specific values and deployments of virtual identities. Our long-term goal is that our methods will generalize as an approach for eliciting cultural values around virtual identity use in other contexts.

Paper Overview

The remainder of this paper is structured as follows. We next describe the motivation for our case study and provide an overview of our approach. A section describing Prior Work follows. The Methods section details our procedures and integration of computational and qualitative methods. We then describe our novel computational framework for virtual identity data collection and analysis in the Qatar-AIRvatar Analysis Toolkit section. We describe the user values and needs that we have qualitatively identified along with our computationally discovered categories of user profile style in the Results and Findings section. Based upon these results and findings, we present as a set of principles for virtual identity system development in the Discussion: Toward Baseline Principles and Best Practices section before ending with the Conclusion, Limitations, and Future Work section.

Case Study Motivation

The GCC context provides several unique features that render it an important and informative region for study. For those who are unfamiliar with the region, this section provides a brief introduction to several cultural characteristics relevant here. The *Khaleeji* identity that encapsulates the people of the Gulf region has transformed profoundly over the last 50 years. The traditional Khaleeji identity bands together around extended family structures (also referred as tribal culture: a membership system that is cited as important as race and class in the region and is interdependent with Khaleeji identity [12]), language (although Arabic is spoken in different dialects throughout the region), and Islamic religious values [4, 5] with some emphasis on gender roles. Khaleeji daily life is prominently impacted by a combination of religio-cultural vision of privacy (typically enforced by Sharia Law at the judicial level) and expectations of upholding the honor of one's family name.

Relevant to the deployment of virtual identities in the region, Abokhodair and Vieweg have described how this mix of cultural, social, and religious values impacts a regional vision of privacy. They describe three aspects related to privacy: awrah, hurma, and haq al-khososyah [2]. Respectively these aspects coincide with the intimate parts of one's body which requires both males and females to cover certain body parts (personal level), the sacredness of certain spaces, home, mosques, and other concepts including awrah (social level), and the local laws that protect hurma (judicial level). These aspects of privacy should not be seen narrowly as religious prescriptions, rather, like most places in the world, privacy is deeply interwoven culturally. This is important to consider since awrah is often misconstrued by people from out-of-culture perspectives since this level of privacy influences the garment choices that the region is sometimes associated with. Previous studies also underline the need for regional religio-cultural individual identities to exist in harmony with the concept of community (umma) where "*a unity of perception, language and civilization [which is] moral and ethical*" is sought [6].

Apart from the rise of the hydrocarbon economy and the effects of modernity, recent changes in Khaleeji identity have been facilitated by four main factors: formation of civil society mechanics, cultural production and exchange in the region, the role of governments in constructing strong national identities, and, most relevant here, the rise of information technologies [4]. Related to the rise of information technologies, the Qatar 2030 National Vision[3] report highlights the need to simultaneously "modernize Qatari society" while preserving cultural traditions. While from our own perspective "modern" and "traditional" are not necessarily oppositional concepts, we mention this here since the National Vision suggests that "modernization" is a key idea in Qatar's self-concept. One aspect of such modernization is the pervasive use of information and communication technologies (ICTs). A 2014 survey shows that 85% of the native population is online and that Qatari "nationals are strongly committed to preserving cultural traditions (94%), and do not see those traditions at odds with a modern, global society" [43]. While Qataris confidently agree that it is possible to preserve their culture and heritage (85%), they also feel their culture should be "more integrated into the modern world" (80%). Although previous research has highlighted some creative practices of Qataris and GCC nationals inside Twitter [1], Facebook [2], and photo-sharing apps in general [3], these interactions and tensions between traditional practices as uptake of emergent uses of technology is not unique to the GCC region. Hence, we posit that our approach is useful to researchers and developers of virtual identity technologies focused on other geographical locations.

[3] www.mdps.gov.qa/en/qnv/Documents/QNV2030_English_v2.pdf

This work aims to formulate a nuanced approach to uncovering everyday user practices and values tensions that arise as negotiate local cultural values in light of values built into technologies that are often designed in other regions such as North America, Asia, and so on. The insight gained from our analysis could be utilized by developers in assessing how their systems' norms and design decisions impact culturally specific groups, and facilitating understanding of how they can support local negotiations of value tensions within cultures instead of relying on false preconceptions from outsider perspectives.

Approach Overview

We implemented a multi-layered mixed-methodology and have assessed a range of current practices exhibiting everyday creativity by Qataris expressing social identity through uses of virtual identity technologies that have not necessarily been designed with GCC values in mind. Although we present this work primarily as a case study, the investigation has resulted in (1) development of novel mixed-methods (qualitative and computational techniques) for understanding user values, needs, and practices in virtual identity systems, and (2) articulation of a set of developer guidelines in the form of principles useful for developing technologies that empower MENA region users to enact traditional values and cultural norms while constructing and deploying virtual identities.

We implemented the *Qatar-AIRvatar* analysis toolkit (an extension of the general *AIRvatar* system [39], which can collect fine-grained data regarding users' avatar customization activities and discover patterns of customization and use in avatars or social media profile images), to perform analysis of GCC social media profiles and avatars. We wrote a script to collect publicly available Instagram profile images and gathered 42,554 Qatari profiles by using the official Instagram API. We then used Qatar-AIRvatar to perform computational analysis of these profiles to discover families of visually similar profile image types using a technique called archetypal analysis.[4]

As a second step, we manually sampled a separate set of 255 Qatari Instagram profiles from users who has checked-in to Qatar within a certain time period and qualitatively analyzed them, examining the content of these accounts one-by-one. We particularly noted the profile pictures, biographical information, and creative techniques deployed by users in navigating the platform. Useful for providing rich, authentic descriptions and quotes illustrating the results that emerged by contrasting our quantitative and qualitative analyses, we additionally conducted 5 in-depth, semi-structured interviews [29] of Qatari virtual identity users (social media and videogames). The interviews were analyzed using Action-Implicative Discourse Analysis, or AIDA [25], and focused on learning about the ideals, beliefs, and practices of the participants as well as how they influence and affect their uses and appropriations of various digital technologies. We focused on social media users and videogame players based upon a cognitive linguistics-based perspective holding that both avatars and profiles form "blended identities," i.e., involving aspects of identity projected from both the physical and virtual worlds [21].

The mixed-methods revealed themes such as Khaleeji (GCC regional) features, social monitoring, self-expression, social connections, and physical and virtual identity contrasts.

[4]While our primary aim is culturally-specific analysis, as opposed to a type of comparative analysis that would suggest a flattening out of intra-country cultural differences, we also collected 101,153 social media profiles from the U.S.A., Brazil, France, Korea, and Sweden for purposes elaborated in the Discussion: Toward Baseline Principles and Best Practices Section. Analysis of the data in these other regions is largely beyond the scope of this paper, however it is worth mentioning that we contrasted our GCC-specific analyses with patterns of use in these other regions to ensure that our focus here is on GCC-specific phenomena and not general trends in Instagram use.

PRIOR WORK

Very little digital media research has taken place that focuses on Qatari society, culture, and/or heritage, with aspects of [43] as a notable exception. Previous studies seeking to understand the values, motivations, and behaviors of users in creating virtual characters or avatars have used application systems that differ in cultural expectations and context (e.g., *Second Life*, *The Sims*, *World of Warcraft*), which may not generalize well to other cultures (e.g., Qatari culture). In particular, for the purpose of better understanding the cultural nuances that exist between different demographics, contexts, and participants, these approaches have not been able to take into account values built into infrastructures, particularly computational data structures that inevitably incorporate values of the designers' milieus, which may or may not be suitable for the target demographics and cultures of participants.

In terms of computational methods, previous research has shown how users' behaviors in virtual environments are closely related to aspects of their physical worlds identities, such as their ages [53], personality [54], motivations [8], behavior [59], and customization preferences [33]. Additionally, previous work has described *shared technical underpinnings* [21] existing between computational identity representation systems such as character avatars, social networking profiles, a perspective supporting the approach taken here both virtual identity information (e.g., in-game behaviors) and real-world information (e.g., social network profiles) result in identity models that are directly comparable and often interoperable at a technical level.

In terms of qualitative methods, there are many useful survey-based studies of behavioral patterns by users in virtual environments, however, there are also limitations of self-reported surveys for studying users [19]. While useful for particularly subjective notions of identities expressed by users (e.g., preferences), self-reported survey data are often difficult to evaluate and subject to survey bias [7]. While useful for understanding certain characteristics of users (e.g., articulated reasons for making a choice), some aspects of users' experiences cannot be articulated (tacit knowledge), are intrusive to keep track of, or are mentally or physically strenuous for participants or interviewers. These cases are often better suited to unobtrusive data collection and analysis such as performed in the computational aspect of the methods we deploy here.

Another common approach toward understanding links between personal choices/views and virtual behavior is to implement standardized survey instruments or questionnaires.

For example, the BIG-5 International Personality Item Pool (IPIP) [18] survey is a common tool used quantitatively measure personality profiles. While shown to be effective within certain contexts and demographics, such instruments may not be generally applicable to other demographics nor translate well across cultures. These standardized survey instruments are not based on values that emerge from the specific data or research question at hand. In contrast to each of these approaches that are all based on predetermined categories, we seek to understand patterns and categories of virtual identity users and preferences that emerge from the data.

METHODS

Data Collection

Social Media Profiles

We collected publicly available data to study how Qatari users represent themselves in social media through their profile pictures and what kind of creative techniques they employ to include their cultural values in their representations. We collected public profile pictures from Instagram because (1) Instagram is one of the widely used social media services in Qatar [56]: 65% of Qataris use Instagram [23], and (2) Instagram API is well-designed to collect large-scale data. While there can be some differences between public and private Instagram accounts in terms of needs and values, particularly when taking into account the Qatari context, it can pose ethical and logistical challenges to collect private account information.

As Instagram does not support search based on the origin country of users, we designed a three-step process to find users who are highly probable to be Qatari. First, we queried a list of "locations" in Qatar. The locations in Instagram are a set of places where users can tag for their photos. Second, for each of the locations, we iteratively collected recent media objects (photos or videos). We then extract the authors of these recent media objects. Finally, among the set of users we collected, we kept only users whose bios or user names were written in the language that corresponds with the location, Arabic in this case. The language was automatically detected by three widely-used libraries, which are CLD2 (from the Chromium project) [52], Langdetect [42], and LangID [41]. If three libraries were not unanimous in terms of the language detected, we considered the language to be the one indicated by the majority of the libraries.

Interviews

As a step toward elaborating on the situated values and creative approaches of the GCC virtual identity users, we also chose to conduct interviews (5 in total) as they are an effective form of qualitative data elicitation [26, 47]. Since our main motivation in conducting interviews were to obtain indicators describing and naming the phenomena that emerged from our main quantitative and qualitative analysis, we view the resulting authenticity as predilected by the small sample size [51].

The interviews were conducted in Qatar by one of the authors, either in person or via teleconferencing software, with durations ranging between 45 and 90 minutes. The interviewees were all Qatari individuals with experience in virtual identity use in social media and videogames. Three interviewees were male and two were female, with ages ranging from 20 to 28 years old. The interviews took place in English, a language that all participants spoke. The interview flow focused on uncovering indications of features that participants would like to see in region-specific technologies and videogames—at the same time, gaining insight and understandings of how these technologies serve as boundary objects that are conceived of, and designed outside of, GCC cultures. Interviews were all transcribed by the interviewer then carefully analyzed by the interviewer and a separate researcher to identify themes and commonalities with a degree of inter-rater reliability.

Computational Analysis

Qatar-AIRvatar Analysis Toolkit

AIRvatar is an integrated set of tools for collecting, transforming, analyzing, and visualizing data related to virtual identities. In prior work, *AIRvatar* has been used to, for example, identify and characterize videogame players' physical-world values based on telemetry data gathered during the creation of videogame avatars [34, 38, 35, 36, 37, 39]. In this paper, we focused on the *AIRvatar* toolkit's capabilities that allow for the analysis of GCC social-media profiles and profile photos.

Archetypal Analysis

We computationally analyzed 42,554 Qatari Instagram accounts in order to identify different categories of user profile style (in terms of meaningful visual characteristics such as distinguishing photos from synthetic images or text). We used a variant of an unsupervised learning method known as archetypal analysis [13]. We use this clustering technique due to its consilience with prototype-based models of cognitive categorization in cognitive science [46, 30, 34, 32]. In this work, we performed archetypal analysis over our set of Instagram data. Given a specified number of archetypes k and a set of data, an archetypal analysis algorithm produces k archetypes that are extremal representations of the features of the data set. In addition, the algorithm tries to identify archetypes such that each observation in the data set can be be described as a convex combination of the k archetypes produced.

In the past, archetypal analysis has found applications in such diverse topics as astrophysics [10], economics [31, 45], sports analytics [15], and videogame recommender systems [50]. Our interest in the algorithm stems from its proven ability to identify extremal data points within a data set that are both interpretable and able to represent the data set as a whole. Previous works related to modeling social identities [34, 38] has leveraged archetypal analysis in this way.

In order to quantitatively analyze the Instagram data set, here we leveraged the data transformation and data analysis algorithms from *AIRvatar*'s software library. The library itself was implemented in Python using the *scikit-learn* [44] and Python Matrix Factorization (*PyMF*) libraries. In this work, we used the Convex Hull Non-negative Matrix Factorization (CHNMF) algorithm [55] for our analyses in place of the canonical implementation of the archetypal analysis algorithm. The algorithm provides equivalent results to archetypal analysis, but performs better in terms of speed. To aid in our analyses, we leveraged *AIRvatar*'s interactive visualization tool for large-N exploration and analysis of virtual identities.

Qualitative Analysis

To validate and explain the quantitative analysis, we manually browsed a separate set of Instagram profiles who had posted photos from Qatar within a recent time frame. Since we could not assume that these users were Qatari, or even Khaleeji, we took the step of performing manual analysis of likely profiles to identify GCC nationals, especially noting Qataris, given our focus on issues such as cultural preservation in light of information and communication technology use. This analysis involved looking at profile photos and translating any biographical information from Arabic to English. We also used such information for ascertaining gender when possible. After eliminating tourists and expatriates, we further studied a representative set of 255 Instagram accounts that we could confidently believe to be either Qatari or from another GCC nationality and expressing relevant cultural markers.

A secondary form of qualitative analysis involved analyzing the 5 interviews to uncover ideals, beliefs, and creative practices of Qatari individuals in creating, customizing, and deploying virtual identities. We used an "Action-Implicative Discourse Analysis" (AIDA) approach [25] augmented with grounded theory techniques [16] to analyze the interviews. These approaches allowed us to assess the interview data, situate it within Khaleeji culture, and make legitimate claims about foundations for design principles borne from the data.

RESULTS AND FINDINGS

Archetypal Analysis Results and Findings

Before performing archetypal analysis on the data, we first had to choose the number of archetypes k. Choosing an optimal value for k is quite difficult in practice, though there are a number of well-established techniques to aid in the decision-making process. To find the optimal number of archetypes k with which to analyze the Qatari Instagram data, we chose to first leverage the Cattell scree test, which is widely used to find the optimal number of factors for describing a data set.

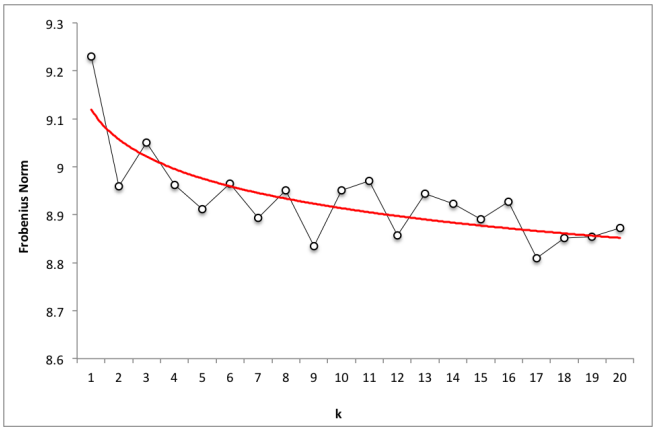

Figure 1. The black line is a log-scale scree plot of Frobenius norms of models produced using Convex Hull Non-negative Matrix Factorization for varying k. The red line is a logarithmic regression of the scree plot.

Figure 1 shows a scree plot [9] of the Frobenius norms[5] of models produced using CHNMF for varying k. We observe that the "elbow" of the graph, and thus, the optimal number of clusters according to the Cattell scree test lies somewhere between $k = 5$ and $k = 9$.

To narrow down this range to a single value, we first observe that there are two noteworthy local minima in our scree plot. One can be seen at $k = 9$ and the other can be seen at $k = 17$. As there is only a marginal decrease in the resulting model's error measure at $k = 17$ that appears to be spurious when considering the overall trend of the graph, and since we found that the archetypes at $k = 9$ archetypes were easier for humans to interpret (four researchers found the greatest consensus of interpretation of the clusters at $k = 9$), we settled on $k = 9$ being the best value of k for our archetypal analysis.

Figure 2 shows the archetypal images produced using *AIRvatar*. The column labeled 'Prototype' shows the archetypal image for each of the $k = 9$ clusters. The column labeled 'Top-Weighted' shows the 9 images that are most similar to a given archetype within the data set. After setting a threshold parameter in the *AIRvatar* toolkit, this column's cells were populated with images that were at minimum 30% similar to the archetypal image for a particular row.

The set of archetypes resulted in 9 clusters. The clusters were identified by recurring dominant themes within them, and although a few cluster members were not in line with the dominant themes, the participating researchers independently decided on similar themes with an acceptable range of inter-researcher reliability. Also utilizing the results of our qualitative analysis (which we go into further detail in the next section), we have named these clusters as follows. Male avatars exhibited two distinct clusters with traditional attires (#2, *Traditional Male*) and non-traditional attires (#6, *Non-traditional Male*). Within the cultural context the decision to have an avatar with traditional or non-traditional clothing may present consequences. On one hand, using an avatar with traditional attire instantly marks the user as a person from the GCC region and someone of presumably Arabic origins (important in Qatar due to the small percentage of nationals). This might also initiate prejudice and bigoted backlash online in broader global contexts (a theme that was illustrated by one of our interviewees who described encountering "racism" online). At the same time, it may also present the user with social capital [11] if the user of the avatar is a business owner that deals with local people or is a member of a traditional family structure. Two other clusters that contained both male and female avatars had either playful selfie-like characteristics (#5, *Selfie*) or illustrative representations (#1, *Illustrative*). Using illustrative representations is observed to be a common creative strategy in terms of protecting modesty and privacy, especially for female avatars, but also for males. One cluster that we believed to be mainly female (#7, *Thematic Photos*) utilized thematic

[5]The objective function of CHNMF seeks to minimize $\|V - \hat{V}\|_F^2$ where V is the original data set and $\hat{V}$ is the product of the two matrices produced by CHNMF. Note that $\|X\|_F^2$ is the Frobenius norm of the matrix X, and is defined as the Euclidean distance between the elements of X.

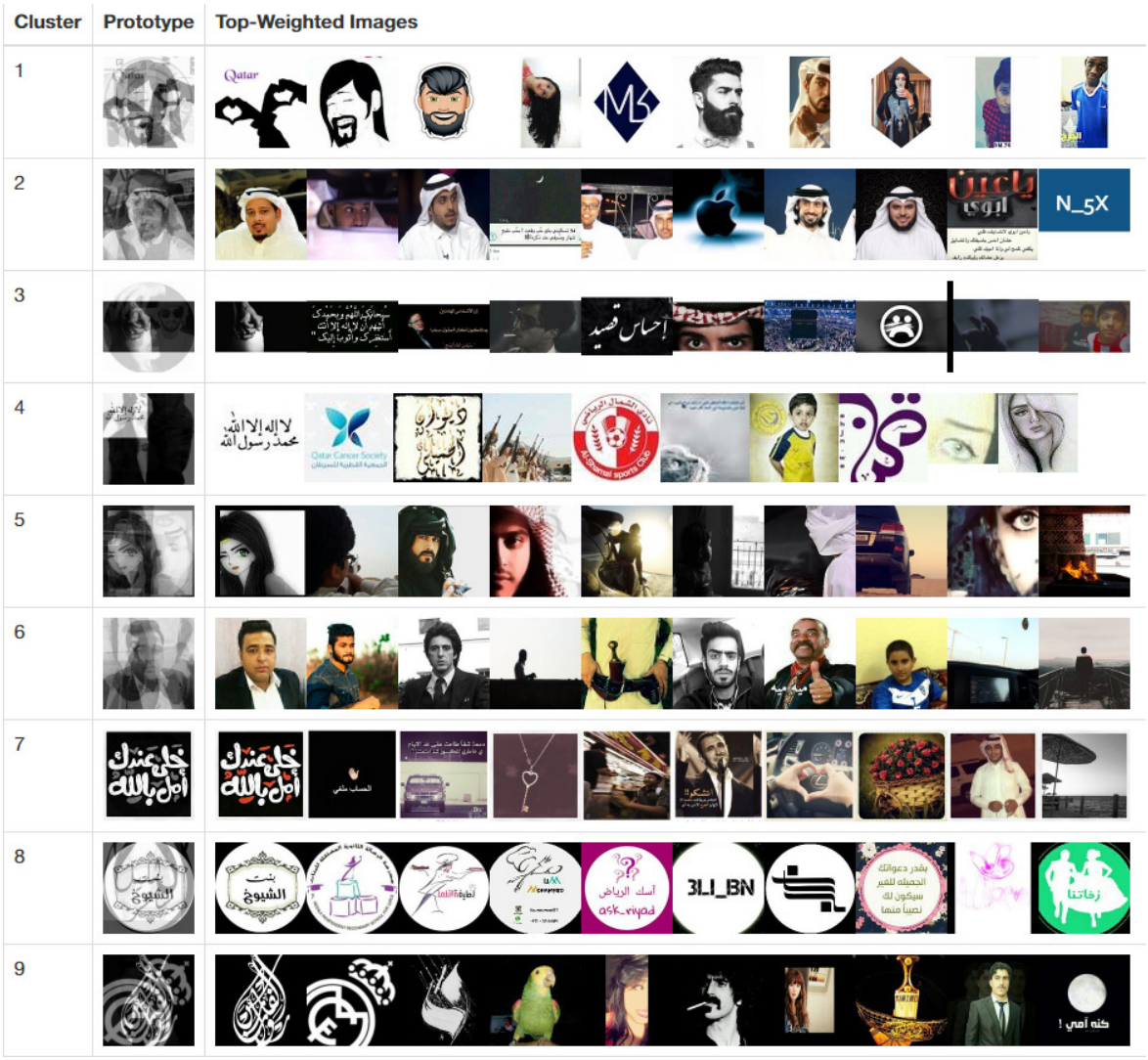

Figure 2. Archetypal profile images from Qatari Instagram accounts.

(e.g., romantically-themed) photos. The themes of the photos focused around flowers, heart symbols, scenery, horses, and cars (especially SUVs for males). This is also another creative strategy for users of the region who value privacy (another theme that has been mentioned in our interviews). Of the three clusters that had more graphical characteristics, one turned out to contain more religious text and symbols (#3, *Faith*), and the others had corporate and commercial logos (#4, *Commercial*) and typographic messages (#8, *Typographic*). Finally, the final cluster (#9, *Other*) compiled together photographic or iconic symbols that had lower descriptive values combined.

Qualitative User Analysis Results and Findings

The hand-coding of a separate set of 255 Instagram profiles resulted in 69 female (27%) and 111 male (44%) users, with an additional 75 profiles (29%) whose genders were unspecified. We coded accounts that seemed to be business-related as unspecified gender types (such as accounts providing Qatari or GCC phone numbers which showcased fashion and jewelry items), unless they were associated with the services of a specific person (such as a hairdresser, make-up artist, or photographer), in which case we assigned the account to the self-presented gender of service provider. Many female users deploy social media for running small home-based businesses, an example of a site where users enact self-expressive power and social engagement through creative profile construction.

Almost a third of the accounts (32%) were private, while the rest were public. The ratio of public accounts for males (73%) was slightly higher than females (65%). Although some might not expect this high ratio for females due to privacy concerns, we can speculate that our hand-coding activity caught more active Instagram users who tend to keep their accounts public in addition to the presence of female local merchants and service professionals such as mentioned above. However as we examined the profiles, we confirmed our inference that many Khaleeji females tend to utilize creative surrogate images instead of personal photos (of themselves, their faces, etc.), transferring the issue of privacy from account access to account content. 40 out of 69 female accounts (58%) did not

show their faces in their profiles photos, depending instead on other thematic images to represent themselves. These images ranged from photos of objects, scenery, flowers, children, close-up details of their faces (mostly both eyes, one eye, or lips), or illustrations. Khaleeji men often used photos of (presumably their own) faces—only 28 male accounts (25%) did not use their faces in the profile photos. The men appeared in either traditional (37%) or non-traditional clothing. In addition, they might sometimes utilize images of horses, falcons, cars (especially SUVs), scenery, and professional athletes.

Merging the results of qualitative analysis with interviews provided us insight on recurring categories which were further consolidated into 5 overarching themes (Table 1). The emerging themes suggest a diversity of values and creative strategies related to virtual identities created among users in GCC countries. There were also gender-related differences between the categories we uncovered.

Mixed-Methods Results and Findings

Comparing our computational clusters with the qualitative themes, we saw that the virtual identity uses in the clusters were each in-line with one qualitative theme at least (Table 2). This lead us to the conclusion that the five emergent themes encompassed unique challenges the members of GCC cultures faced while using these virtual identity systems and some of the creative solutions they employed to overcome them.

The theme of *Khaleeji Features* indicates that participants were enthusiastic about having the option of using traditional, culturally-specific features in their avatars, or seeing characters in videogames that have these features. For example, consider the prevalence of the ghutra[6] in cluster #2 and Arabic writing in cluster #3 in Figure 2. From the qualitative analysis we have observed that a majority of the profile images had visual clues that hinted Khaleeji origins of the users. From the interviews, one female participant referred to the character *Shaheen* in the videogame *Tekken 7*: *"If you know the game Tekken [...] they have a new character and they called it Shaheen. He is based on a Saudi person, and he wears the ghutra, and he has the Arabian beard, the Khaleeji beard, and he has a falcon. It was really fun to play with."* Even when Khaleeji features were not present, users would find creative ways to invent them. The same female participant mentioned a strategy used by Arabic speaking users in utilizing the pronunciation of numbers and letters in creating profile names to indicate their ethnic origins. A male participant described his efforts in hacking game codes to *"change the player names, how they look, and make it like [Qatari]."* The culturally diversifying features mainly focused on clothing, skin color, and hair/beard style. Participants also mentioned that there are significant differences of beard and ghutra styles between the men in GCC, and very few virtual environments implement this form of cultural diversity. Female participants mentioned the necessity of features which reflect modesty in their avatars.

The theme of *Social Monitoring* revealed issues and concerns regarding privacy. For example, in Figure 2 consider the prevalence of illustrative avatars in cluster #1 and thematic images such as hands in the shape of a heart or flowers in cluster #7 instead of user photos. Such examples paralleled our qualitative analysis, which revealed fewer examples of full-face photo usage than other regions. Regardless of gender, Khaleeji users creatively came up with aestheticized strategies like using shadows and light, silhouettes, back of the head profiles, framing half the face, or focusing on a part of the face (typically eyes). The theme was also particularly meaningful when deciding to add new friends. For example, from the interviews, the second female participant said: *"I feel like my account is my privacy, I only add my friends—I wouldn't want a random stranger knowing what I'm watching right now, what I'm playing right now, or what my status is...."* An additional aspect regarding monitoring and privacy, and the wish to be viewed as upright and respectable revolves around the usage of microphones in videogame sessions for females. It is not uncommon for people who do not know each

Themes	Recurring Codes
1. Khaleeji Features	Common
	• Traditional Clothing • Avatar Skin Color • Khaleeji Diversity
	Female
	• Modesty
	Male
	• Avatar Beard and Hair Style
2. Social Monitoring	Common
	• Self-monitoring • Monitoring by Others
	Female
	• Privacy • Microphone Use
3. Self-expression	Common
	• Self-likeness of Avatars • Cosmetic Avatar Features
	Female
	• Gender Related Avatar Features
4. Social Connections	Common
	• Connections with Family
	Female
	• Connections with Strangers
	Male
	• Fear of Bigoted Backlash • Relationship Seeking
5. Physical and Virtual Identity Contrasts	Common
	• Non-permeability • Escapism

Table 1. Overarching themes that emerged during the interviews.

[6]A ghutra is a type of headdress traditionally worn by Arab men.

	Khaleeji Features	Social Monitoring	Self-expression	Social Connections	Physical and Virtual Identity Contrasts
Traditional Male	✓		✓	✓	
Non-traditional Male			✓	✓	
Selfie			✓		✓
Illustrative		✓			✓
Thematic Photos		✓			✓
Religious	✓				
Commercial				✓	
Typographic				✓	✓

Table 2. Comparing computational clusters with qualitative themes.

other in a physical context to connect via an online videogame and communicate via microphone. However, in the GCC (among other regions) a female voice is considered personal and private, and not something to be exposed to outsiders. Therefore, when and how female gamers use microphones is a point of contention—some use them, some do not, and some will only listen, and not speak, but use text to offer responses and thoughts. Another practice that we see as an extension of privacy concerns and deem as a creative process was the practice of creating a persona or a paragon instead of self-likeness. In social media profiles this emerged as using the images and quotes of celebrities and fictional characters. In videogames this emerged as creatively selecting regional paragons and modeling virtual identity around it such as this female participant's response: *"I wanted it to look like the [ancient deity Ra], so I made it extremely skinny and black and tall and weird and bug-eyed [...] Even the way the character would act, I would make it act snobby, like some sort of deity."*

Participants also expressed their commitment to implementing creative details in their avatars that empowered their *Self-expression*—such as skin color and fashion. For example, in Figure 2 consider again the prevalance of traditional clothing in cluster #1 in addition to the use of highly stylized thematic images featuring rich shadows in cluster #5 and the use of celebrity and "cool" posed images in cluster #6 and #9. In our qualitative results, one female participant gave this creative example of preserving head covering in videogames: *"the healer [character's] clothes are usually all covered, even with a hoodie and everything so I found it interesting, I did just realize it last month that my avatar is really well-covered like I am"*), and character origin (the second female participant gave this example from the videogame *Skyrim: "In Skyrim, I took the Khajiit tribe, and they're not even human, they're anthropomorphic type of animal, they're based on cheetahs, and the reason I picked them is because they come from the desert—so it reminded me of myself..."*[7] Also 2 (one female, one male) out of 5 participants mentioned an avatar gender-based choice for videogame character profession, as illustrated by one male participant's comment: *"If the character is some type of swordsman, I don't know why—or like an axe wielder, I would prefer like a chunky guy with a beard [...] but if it's like a magician or something, I would say an elfin woman."*

The theme of *Social Connections* underlined the purposes of online connections established by participants. For example, in Figure 2 consider how the clothing in clusters #2 and #5 and the use of Arabic writing, particularly in clusters #3, #4, #7, and #9, can facilitate recognition and connections between Khaleeji individuals. A dominant sub-theme was in the differences between making formal connections (with strangers) versus informal connections (with friends and family members). The qualitative analysis revealed how the framing of formal photos (with traditional attire) and non-formal photos varied, with formal photos including more of the body, similar sitting positions, and direct angles. Another concern about social connections emerged from the interview where the first female participant expressed worry about representing her Khaleeji identity in digital environments, as she thought this might lead to complications or problems around bigotry, described as racism: *"[...] I feel there would be more racism involved if there were more [Khaleeji] options to display certain things."* Participants were acutely aware that their backgrounds are in the minority in greater gaming communities—culturally, racially/ethnically, and with regards to nationality. Due to the nature of many videogames that allows for or even necessitates communication with people from all over the world, concern, and maybe even fear, are ever-present potentials vis-à-vis the communication patterns and routines participants regularly undertake. This, in addition to the great significance placed on privacy in Khaleeji culture, are cause for great care when it comes to connecting with others in digital contexts.

Finally, some codes that related to the comparisons between physical world experiences of participants versus their virtual identity-based experiences were consolidated under the *Physical and Virtual Identity Contrasts* theme. For example, again in Figure 2, consider the anonymizing effects of illustrative avatars in cluster #1, Arabic writing in clusters #3, #4, and #8, artfully stylized images in clusters #4, #5, and #9, and photos of people other than users (and even a parrot) in clusters #6 and #9. Users from the region emphasized virtual identities as personal creative spaces such as this female participant who described game worlds as *"dream worlds [where] everyone just does whatever he or she wants, without any judgment"* and another as *"a gateway to a different world."* Another interesting component of this theme was the non-permeability of the virtual, as summarized by the first female participant as; *"Everything that happens online stays online [...] I feel that the majority of people, if they know they're anonymous then they probably do things they wouldn't do if they weren't."* However

[7] Cheetahs historically they occupied other habitats in the MENA region, if not actually deserts.

this is not a phenomena directed solely for female users as one male participant notes: *"it sucks for the [men too], we get the same treatment"*, then later adds *"we all have a different identity online"*. This "separation" between the physical and virtual selves is prevalent globally in social media settings, however here we highlight culturally-specific phenomena in the context of the GCC and attendant social consequences that may ensue if social media users and videogame players make missteps that are cause for regionally-shared perspectives on public shame or embarrassment. The idea of anonymity is important all over the world, but when it is offered and utilized by those in the GCC the option to engage and participate in various digital communications may offer even more outlets than are typically permitted in face-to-face environments. One important outcome of these physical and virtual identity contrasts was a familiar practice of maintaining multiple identities and/or accounts as is discussed in the next section.

Overall the merging of the results of computational and qualitative analysis revealed the strategies of appropriations by Khaleeji users in these virtual environments, indicating the push toward modernity in the region, while the preservation of tradition remains a predominant goal.

DISCUSSION: TOWARD BASELINE PRINCIPLES AND BEST PRACTICES

It is useful to consider the broader implications and limitations of the results above. Our results characterize a set of ways that Qatari users construct, utilize, deploy virtual identities and employ creative strategies while doing so. Our study is culturally-grounded in the sense that it aims to characterize data collected in a specific region. Our study is not intended to reveal a characterization of a people or culture at large, and we do not claim that it is an exhaustive account. Indeed, Qatari culture, like any other culture, is not monolithic and individuals may fall outside of the categories articulated in our analyses. For the same reason, our approach is not to construct generalizing cross-cultural comparisons of the form: "Qataris prefer X, where as Swedes prefer Y." However, it has been useful to contrast discovered clusters of social media profile image use in diverse regions for a more parsimonious reason. We aimed to verify that some of our findings are not characterizations of social media use irrespective of region or culture. Hence, we did perform archetypal analysis of data in other regions and, indeed, discovered clusters that did not exist in Qatar and vice versa. For example, in France there were fewer formal business portrait photos and more images of types including candid selfies, photos of users' pets, photos of couples, and more revealing attire regardless of gender. In Sweden there were more "activity portraits," which indicated the apparent hobbies or professions of users (e.g., a kayaker in a river or a baker in an apron in front of a shop window). We did not note faith-oriented profiles as an archetypal deployment in France or Sweden. Such distinctions should not be seen as characterizing the cultures at large in the respective countries, but rather as evidence that archetypal analysis can be useful for discovering diverse culturally-grounded and specific virtual identity image deployments and not only general trends that cut across all regions alike. We leave more systematic studies to compare regional differences of virtual identity usage for future work.

Despite the robust alignment of our computational and qualitative results, our approach faced some constraints and limitations. Of particular note, some of the values and practices described above, such as modesty and social monitoring, also influence a number of Qatari virtual identity users to maintain multiple accounts, both public and private. While the use of public vs. private accounts is not unique to Qatar or the GCC, informal discussions with informants from the region indicate that the deployments of social media profile content in private accounts for networks of immediate family members, close friends, and broader social connections would likely all results in quite distinct findings. Nonetheless, when understood as an account of public identity performance, our results above reveal a coherent portrait of an array of Qatari virtual identity deployments commensurate with the needs and values elicited by our interviews and described in prior work.

Culturally-Grounded Design Principles and Guidelines for Virtual Identity System Developers

When considering our results as restricted to public sphere virtual identity use, they suggest a preliminary set of design principles that could help guide virtual identity systems developers creating applications for use in the region. These principles emerge from considering the values and needs our study elicited in light of two distinct types of virtual identity creation, use, and deployment features that developers must consider. First, *embedded* features are those that are directly enabled by technical affordances of systems. For example, filters that can be applied to photographs, pre-built customization options when constructing avatars, and default images or avatars are all embedded features in that developers explicitly implement the features to enable users to self-represent. In contrast, the second type of virtual identity features are those that are *enacted* by users without direct influence of developers. For example, the choices of profile images that users upload, the content of text chosen to be entered into open text fields such as "Name," and text that users post to each others' feeds are all enacted, or user performed, features of virtual identity systems. The distinction between embedded and enacted features is important to note because while direct design principles can be offered to developers regarding embedded features such as adding specific types of customization options, it is a more subtle challenge to offer principles for developers to support actions of users involving externally generated content. Hence, the following two subsections articulate two very different sets of principles that reflect the distinctions between embedded and enacted virtual identity system features.

Principles for Embedded Virtual Identity Features

The following are example recommendations that virtual identity system features developers could implement in Qatar and the GCC in order to support local needs and values for creative self-represenation. Taken together, such recommendations form the basis for a set of culturally-grounded design principles for virtual identity systems. These principles are particularly appropriate for social media-based virtual iden-

tities, as they emerge from our Instagram studies, but could likely generalize to other forms of virtual identity.

- *Khaleeji Features & Social Monitoring*: Avatar creators for generating non-photographic (or cartoony) virtual identity representations incorporating a wide array of local self-representational styles (e.g., clothing and grooming preferences) enable traditional articulation of Khaleeji features. Such avatar creators can also support needs regarding modesty, particularly regarding values precluding publicly deployed female photographic representations.

- *Social Monitoring*: Using accounts verified for gender, users can define settings regarding the preferred communications methods, e.g., voice chat, speech-to-text functions, text chat, and/or symbol-based chat (using prebuilt elements to construct commonly used phrases). Similarly, unobtrusive vocal anonymizing features can be deployed in settings in which communication is required, but values, needs, or tensions regarding modest self-presentation emerge.

- *Social Connections*: Rather than only enabling groupings of social connections according to standardized distinctions such as between public vs. private accounts, regionally and culturally-specific distinctions such as *mahram* (spouses and unmarriageable kin) vs. *non-mahram* connections could be implemented as pre-built categories with requisite affordances for enabling and restricting access and providing appropriate self-representations to other users.

- *Physical and Virtual Identity Contrasts*: When engaging other users on social media involving international connections, as noted in our interviews, GCC members have been recipients of harassment based upon stereotypes and prejudices from the physical world. At the same time, female users have encountered gender-specific forms of harassment online. Specific community guidelines and moderation explicitly targeting regionally-specific phenomena, in addition to targeted outreach to inform users about these issues, could be used to ameliorate some of these forms of abuse.

The list above is not exhaustive, but rather is representative of principles for guiding culturally-specific embedded virtual identity system features driven by values and needs elicited in the GCC rather than introspectively determined by outsiders.

Principles for Enacted Virtual Identity Features

While it may appear challenging to provide developers with a set of principles supporting practices user enacted virtual identity features because users' actions and content cannot always be anticipated, developers can, in fact, provide categories to structure users' activities based upon assessing patterns in user behavior. For example, our computational results suggest the design principles supporting user enacted features such as:

- *Predefined Categories for User Image Categorization*: Enabling users to tag profile images could support ensuring virtual identity deployments appropriate to different settings and viewers. For example, tags for Formal vs. Informal, Traditional vs Non-traditional, and Photographic vs. Illustrative could be used to help automate display of appropriate and relevant user images.

- *Automated Processing*: While these features may be less on the immediate horizon, algorithmically robust techniques such as face recognition applied to photographs could be used for enhancing, altering, or redacting photographic images depending on the image's setting or potential viewer, according to local needs and values related to social monitoring and social connections. Similarly, biometric measures such as fingerprint scans could be used to restricting access to content per the same types of needs and values.

The features described above represent our first consolidation of "lessons learned" from our case study articulated as actionable principles that we hope can be put to use by system developers located in and/or targeting the GCC region.

CONCLUSION, LIMITATIONS, AND FUTURE WORK

The central aim of this paper has been to present a novel approach to analyzing and better understanding culturally-grounded virtual identity uses, the cross-cultural creative ways underserved users navigate these identities, and demonstrating those results via a mixed-methods case study in Qatar. We have also articulated a novel mixed methods approach including a computational platform of our own inventions for discovering and understanding user values and practices in virtual identity systems. A final aim has been to present important steps toward articulating a set of baseline principles and best practices for developing information and communication technologies that empower Qataris in everyday creative practices of enacting their needs and values using virtual identity technologies. As described above, the dataset here is limited to public accounts. Although studying private accounts may reveal important different cultural nuances, deploying our mixed-methods approach on public accounts is a step towards discovering cultural needs for both the public and private domains. By discovering implicit patterns in data, our computational methods reveal tacit values in social media profile use rather than explicitly stated user values. We further suggest that one future work direction would be to use the *AIRvatar* social media profile constructor to collect data on behavioral patterns in profile customization as done in [40] for revealing additional tacit values expressed by users.

In future work we aim to further develop our recommended principles and practices for developers via (1) applying our approach to new case studies focused on other systems for creating avatars for videogames and virtual environments, and (2) running experiments in a controlled experimental setting where we can use *AIRvatar* to collect fine-grained customization data as users construct avatars such that we can analyze not only avatars as an end-products, but also the processes of creating them. Additionally, we aim to adapt a virtual system of our own design to culturally-specific needs and values of Qatar by enacting our own design principles and evaluating the resultant system. Ultimately, we hope that users will be able to adapt our approach to support the needs and values of diverse communities and subgroups globally, especially for the purposes of empowering people to participate in digital media via virtual identities in ways that support their own cultural needs and values.

ACKNOWLEDGMENTS

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

REFERENCES

1. Norah Abokhodair, Sofiane Abbar, Sarah Vieweg, and Yelena Mejova. 2016. Privacy and twitter in Qatar: Traditional values in the digital world. In *WebSci '16 Proceedings of the 8th ACM Conference on Web Science*. ACM, 66–77.
2. Norah Abokhodair and Sarah Vieweg. 2016. Privacy & social media in the context of the Arab Gulf. In *DIS '16 Proceedings of the 2016 ACM Conference on Designing Interactive Systems*. ACM, 672–683.
3. Norah Abokhodair, Sarah Vieweg, and Adam Hodges. 2017. Accepted. Photo sharing in the Arab Gulf: Expressing the collective and autonomous selves. In *CSCW 2017 Proceedings of the 20th ACM Conference on Computer-Supported Cooperative Work and Social Computing*. ACM.
4. Abdulaziz Saud Al-Babtain. 2016. Foreword. Oxford Gulf and Arabian Peninsula Studies Forum Autumn 2016 Issue, oxgaps.org. (2016).
5. Lulwa Abdulla Al-Misned. 2016. Understanding the Evolution of the Khaleeji Identity. Oxford Gulf and Arabian Peninsula Studies Forum Autumn 2016 Issue, oxgaps.org. (2016).
6. Nazih N. Ayubi. 1995. *Over-Stating the Arab State: Politics and Society in the Middle East*. I.B.Tauris & Co Ltd., New York, NY.
7. Christian Bauckhage, Kristian Kersting, Rafet Sifa, Christian Thurau, Anders Drachen, and Alessandro Canossa. 2012. How players lose interest in playing a game: An empirical study based on distributions of total playing times. In *Proceedings of the IEEE Computational Intelligence and Games Conference*. IEEE, 139–146.
8. Alessandro Canossa, Josep B Martinez, and Julian Togelius. 2013. Give Me a Reason to Dig Minecraft and Psychology of Motivation. *In Proceedings of IEEE Conference on Computational Intelligence and Games* (2013).
9. Raymond B Cattell. 1966. The scree test for the number of factors. *Multivariate behavioral research* 1, 2 (1966), 245–276.
10. Ben HP Chan, Daniel A Mitchell, and Lawrence E Cram. 2003. Archetypal analysis of galaxy spectra. *Monthly Notices of the Royal Astronomical Society* 338, 3 (2003), 790–795.
11. Chao-Ming Chiu, Meng-Hsiang Hsu, and Eric T.G. Wang. 2006. Understanding knowledge sharing in virtual communities: An integration of social capital and social cognitive theories. *Decision Support Systems* 42, 3 (2006), 1872–1888.
12. Miriam Cooke. 2014. *Tribal Modern: Branding New Nations in the Arab Gulf*. University of California Press, Los Angeles, CA.
13. Adele Cutler and Leo Breiman. 1994. Archetypal analysis. *Technometrics* 36, 4 (1994), 338–347.
14. Robert Andrew Dunn and Rosanna E. Guadagno. 2012. My avatar and me: Gender and personality predictors of avatar-self discrepancy. *Computers in Human Behavior* 28, 1 (2012), 97 – 106. DOI: http://dx.doi.org/10.1016/j.chb.2011.08.015
15. Manuel JA Eugster. 2011. Archetypal athletes. *arXiv preprint arXiv:1110.1972* (2011).
16. Barney G Glaser. 1967. *The discovery of grounded theory: Strategies for qualitative research*. Aldine Publishing.
17. Erving Goffman and others. 1959. The presentation of self in everyday life. (1959).
18. Alan J Gow, Martha C Whiteman, Alison Pattie, and Ian J Deary. 2005. Goldberg's 'IPIP' Big-Five factor markers: Internal consistency and concurrent validation in Scotland. *Personality and Individual Differences* 39, 2 (2005), 317–329.
19. Rosanna E. Guadagno, Nicole L. Muscanell, Bradley M. Okdie, Nanci M. Burk, and Thomas B. Ward. 2011. Even in virtual environments women shop and men build: A social role perspective on Second Life. *Computers in Human Behavior* 27, 1 (2011), 304–308. DOI: http://dx.doi.org/10.1016/j.chb.2010.08.008
20. D. Fox Harrell. 2009. Computational and Cognitive Infrastructures of Stigma: Empowering Identity in Social Computing and Gaming. *Proceedings of the 7th ACM Conference on Cognition and Creativity* (2009), 49–58.
21. D. Fox Harrell. 2010. Toward a Theory of Critical Computing. CTheory, ctheory.net/articles.aspx?id=641. (2010).
22. D Fox Harrell and S Veeragoudar Harrell. 2012. Imagination, computation, and self-expression: Situated character and avatar mediated identity. *Leonardo Electronic Almanac* 17, 2 (2012).
23. Rassed: Exploring Digital Impacts in Society. 2015. Understanding emerging social media platforms in Qatar. (2015). http://www.ictqatar.qa/sites/default/files/understanding_emerging_social_media_platforms_in_qatar.pdf.
24. Dominic Kao and D Fox Harrell. 2015. Toward Evaluating the Impacts of Virtual Identities on STEM Learning. In *Proceedings of the 10th International Conference on the Foundations of Digital Games*.
25. Tracy Karen. 1995. Action-implicative discourse analysis. *Journal of Language and Social Psychology* 14, 1-2 (1995), 195–215.
26. Nigel King and Christine Horrocks. 2010. *Interviews in qualitative research*. Sage Publishing.

27. Thomas Kohler, Johann Fueller, Daniel Stieger, and Kurt Matzler. 2011. Avatar-based innovation: Consequences of the virtual co-creation experience. *Computers in Human Behavior* 27, 1 (2011), 160–168.
28. Thomas Kohler, Kurt Matzler, and Johann FÃijller. 2009. Avatar-based innovation: Using virtual worlds for real-world innovation. *Technovation* 29, 6-7 (2009), 395–407.
29. Richard A. Krueger and Mary Anne Casey. 2000. *Focus groups : a practical guide for applied research*. Sage Publishing.
30. George Lakoff. 1998. *Cognitive Semantics*. Indiana University Press, 119–154.
31. Shan Li, Paul Wang, Jordan Louviere, and Richard Carson. 2003. Archetypal analysis: A new way to segment markets based on extreme individuals. In *A Celebration of Ehrenberg and Bass: Marketing Knowledge, Discoveries and Contribution. Proceedings of the ANZMAC 2003 Conference*. 1674–1679.
32. Chong-U Lim. 2016. *Me, my Avatar (s), and AI: computational models of users and virtual identities for analysis, design, and development*. Ph.D. Dissertation. Massachusetts Institute of Technology.
33. Chong-U Lim and D. Fox Harrell. 2013. Modeling Player Preferences in Avatar Customization Using Social Network Data. In *Proceedings of the IEEE Computational Intelligence and Games Conference*. 153–160.
34. Chong-U Lim and D. Fox Harrell. 2015a. Comparing clustering approaches for modeling players' values through avatar construction. In *Proceedings of the 1st AIIDE Workshop on Player Modeling*.
35. Chong-U Lim and D. Fox Harrell. 2015b. A data-driven approach for computationally modeling players' avatar customization behaviors. In *Proceedings of the 11th AAAI Artificial Intelligence and Interactive Digital Entertainment Conference*.
36. Chong-U Lim and D Fox Harrell. 2015c. Developing computational models of players' identities and values from videogame avatars. In *Proceedings of the 10th International Conference on the Foundations of Digital Games*.
37. Chong-U Lim and D. Fox Harrell. 2015d. The marginal: A game for modeling players' perceptions of gradient membership in avatar categories. In *Proceedings of the 2nd AIIDE Workshop on Experimental AI and Games*.
38. Chong-U Lim and D. Fox Harrell. 2015e. Revealing social identity phenomena in videogames with archetypal analysis. In *Proceedings of the 6th International AISB Symposium on AI and Games*.
39. Chong-U Lim and D. Fox Harrell. 2015f. Toward telemetry-driven analytics for understanding players and their avatars in videogames. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '15)*. New York, NY, USA, 1175–1180. DOI: http://dx.doi.org/10.1145/2702613.2732783
40. Chong-U Lim and D. Fox Harrell. 2015g. Understanding Players' Identities and Behavioral Archetypes from Avatar Customization Data. In *Proceedings of the IEEE Conference on Computational Intelligence and Games*.
41. Marco Lui and Timothy Baldwin. 2012. langid. py: An off-the-shelf language identification tool. In *Proceedings of the ACL 2012 system demonstrations*. Association for Computational Linguistics, 25–30.
42. Shuyo Nakatani. 2010. Language detection library for Java. (2010). https://github.com/shuyo/language-detection
43. Northwestern University of Qatar. 2014. Media in the Middle East: research by Northwestern University in Qatar. (2014). http://www.mideastmedia.org/survey/2014/.
44. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.
45. Giovanni C Porzio, Giancarlo Ragozini, and Domenico Vistocco. 2008. On the use of archetypes as benchmarks. *Applied Stochastic Models in Business and Industry* 24, 5 (2008), 419–437.
46. Eleanor Rosch. 1999. Principles of categorization. *Concepts: Core readings* (1999), 189–206.
47. Irving Seidman. 1998. *Interviewing as qualitative research : A guide for researchers in education and the social sciences*. Teachers College Press.
48. Sercan Sengun. 2014. A semiotic reading of digital avatars and their role of uncertainty reduction in digital communication. *Journal of Media Critiques* 1, 1 (2014), 149–162.
49. Sercan Sengun. 2015. Why do i fall for the elf, when i am no orc myself? The implications of virtual avatars in digital communication. *Comunicacao e Sociedade* 27 (2015), 181–193.
50. Rafet Sifa, Christian Bauckhage, and Anders Drachen. 2014. Archetypal game recommender systems. *Proc. KDML-LWA* (2014).
51. Davis Silverman. 1993. *Interpreting qualitative data: methods for analyzing talk, text and interaction*. Sage Publishing.
52. Dick Sites. 2013. Compact language detector 2. (2013). https://github.com/CLD2Owners/cld2.
53. Shoshannah Tekofsky, Pieter Spronck, Martijn Goudbeek, and Jan Broersen. 2013a. Towards a Player Age Model. In *Ninth Artificial Intelligence and Interactive Digital Entertainment Conference*.

54. Shoshannah Tekofsky, Jaap Van Den Herik, Pieter Spronck, and Aske Plaat. 2013b. Psyops: Personality assessment through gaming behavior. In *Proceedings of the Foundations of Digital Games Conference*. 12–65.

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

56. Sarah Vieweg and Adam Hodges. 2016. Surveillance & modesty on social media: How Qataris navigate modernity and maintain tradition. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 527–538.

57. Nick Yee. 2006. Motivations for play in online games. *CyberPsychology & behavior* 9, 6 (2006), 772–775.

58. Nick Yee and Jeremy Bailenson. 2007. The Proteus effect: The effect of transformed self-representation on behavior. *Human communication research* 33, 3 (2007), 271–290.

59. Nick Yee, Nicolas Ducheneaut, Les Nelson, and Peter Likarish. 2011. Introverted Elves & Conscientious Gnomes : The Expression of Personality in World of Warcraft. *In Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems* (2011), 753–762.

60. Nick Yee, Nicolas Ducheneaut, Han-Tai Shiao, and Les Nelson. 2012. Through the Azerothian looking glass: Mapping in-game preferences to real world demographics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. 2811–2814.