Diversity in Online Advertising: A Case Study of 69 Brands on Social Media

Jisun An(✉) and Ingmar Weber

Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar
jisun.an@acm.org, iweber@hbku.edu.qa

Abstract. Lack of diversity in advertising is a long-standing problem. Despite growing cultural awareness and missed business opportunities, many minorities remain under- or inappropriately represented in advertising. Previous research has studied how people react to culturally embedded ads, but such work focused mostly on print media or television using lab experiments. In this work, we look at diversity in content posted by 69 U.S. brands on two social media platforms, Instagram and Facebook. Using face detection technology, we infer the gender, race, and age of both the faces in the ads and of the users engaging with ads. Using this dataset, we investigate the following: (1) What type of content brands put out – Is there a lack of diversity?; (2) How does a brand's content diversity compare to its audience diversity – Is any lack of diversity simply a reflection of the audience?; and (3) How does brand diversity relate to user engagement – Do users of a particular demographic engage more if their demographics are represented in a post?

Keywords: Diversity · Gender · Race · Demographics · Advertising Brand · User engagement · Social media · Instagram · Facebook

1 Introduction

In the early 1960s, the ad world in the US still had a one-size-fits-all strategy: everyone saw the same advertisement, and the advertisement typically represented White people. Since Tom Burrell [20] started working at Wade Advertising Agency in Chicago as the first African American, most companies have realized that, in his words, "If you don't target, you cannot sell.", leading to an increase in diversity of the people depicted in ads. These changes were, however, slow [7], mainly because many businessmen were afraid to lose sales by including African-American models in promotional materials [5]. Even as recent as 2016, Lloyds reported that in the UK, just 19% of people in ads are from minority groups [17].

Due to the continuing under-representation of minority groups, researchers have studied how different ethnic groups perceive culturally embedded ads, i.e., ads with African-American or Asian actors, and whether they are in favor of it or

S. Staab et al. (Eds.): SocInfo 2018, LNCS 11185, pp. 38–53, 2018.
https://doi.org/10.1007/978-3-030-01129-1_3

not. Somewhat surprisingly, the results are mixed. A few studies showed Whites responded similarly or more favorably to ads with non-White models [5,18]; in contrast, other studies showed integrated or all-Black casts in ads may or may not elicit a backlash among some White college students and adults [6]. Several studies showed that, relative to Whites, non-Whites seem to be more aware of and responded more favorably to ethnically resonant ads [8,18]. However, it has been reported that high-income Asians and Hispanics are known to prefer to see Whites in ads [16], and when choosing a doctor, Kenyans prefer to meet European-looking doctors when a condition is a serious matter [19]. These previous studies have mostly been conducted in a lab-experiment settings, and they have focused on ads on TV or in magazines, which are both forms of media with one-way interaction. Since a brand can see the users' responses on social media directly, social media has an advantage when it comes to understanding what role diversity can play in ads.

In this study, we characterize the diversity among ads of top US brands on social media. To this end, after filtering, we analyze 14,303 posts and 850,109 comments relating to 69 U.S. brands on two popular social media platforms: Instagram (IG) and Facebook (FB). We use computer vision to automatically infer the gender, race, and age of faces depicted in ads, and we apply the same technology to the profile pictures of users engaging with the ads.

With this large-scale data set, we answer the following research questions:

1. How much demographic diversity is there in online advertising in the content put out by major brands on social media?
2. How does the demographic diversity in a brand's posts compare to the diversity of their engaging audience?
3. Is there resonance between the demographics depicted in a particular post and the audience engaging with the post?

We observe that most brands over-represent White faces, compared to their users, and that there is a resonance between a post's demographics and the engaging audience.

2 Related Work

Since ads can reinforce or introduce stereotypes, how particular demographic groups are depicted in ads has long been studied. Until the late 90s, the prime-time ads clearly showed distinct racial segregation with Whites appearing in ads for upscale products, beauty products, and home products while people of color appeared in ads for low-cost, low-nutrition products and athletic and sports equipment ads [10]. In another example, Asian Americans are frequently depicted as highly educated, proficient with technology, and affluent [23].

Studies on how viewers react to ethnically resonant ads, i.e., ads including non-White actors, have shown mixed results. One theory, *in-group preference theory*, posits that in-group members on the basis of race will evaluate other in-group members more favorably than out-group members [24]. Indeed, some

researches have shown that people identify more with and respond more favorably to ads with same-race models/actors [24,26]. However, a second theory, *polarized appraisal theory*, predicts that out-group members will be evaluated more extremely (positively or negatively) than in-group members [24]. Some studies showed Whites responded similarly or more favorably to ads with non-White models [5,18]. However, other studies showed that ads with non-White models may or may not elicit a backlash among some White [6]. For other races, Appiah *et al.* reports that, compared to ads with White actors, ads with Black actors are more favored by members of all races [2]. Lastly, relative to Whites, non-Whites often were more aware of and responded more favorably to ethnically resonant ads [8,18]. In this work, we extend these theories on how viewers react to ethnically resonant ads to gender and age and examine whether users engage more with ads showing same-gender, same-race, or same-age group actors (i.e., in-group preference theory).

(a) An example brand's post on Instagram (Nike)

(b) An example brand's post on Facebook (Adidas)

Fig. 1. Example posts by brands on Instagram and Facebook with user comments. Personal profile pictures and names are blackened for privacy concerns.

3 Data and Methodology

For this study, we first select a set of brands. Then, we identify their official accounts on both Instagram and Facebook. For those brands with official accounts, we collect all posts (statuses and posts – jointly referred to as posts) published and the comments they received. We then use an existing computer vision tool to analyze the pictures with faces in the brands' posts and the profile pictures of the users engaging with brands and determine their age, gender, and race.

Brands on Social Media. We use two lists of brands on the Web including BrandFinance's Global 500 2016[1] and Interbrand's Best Global Brands 2016[2].

[1] http://brandirectory.com/league_tables/table/global-500-2016.
[2] http://interbrand.com/best-brands/best-global-brands/2016.

We choose to use these two lists due to their availability at the time of data collection. We find that at the end of 2017, the list of global brands has not changed significantly compared to the earlier one. We employed the industry classification method used by BrandFinance to keep the sectors that follow a business-to-consumer model, under the assumption that these brands would be more active in using social media for user engagement. We also limit our focus to U.S. based brands that officially appear in at least one of the following two social media: Instagram and Facebook. This results in a list of 132 brands which further could be categorized into 14 sectors. Furthermore, we only select brands with official brand accounts created by the companies, that use the English language for communication.

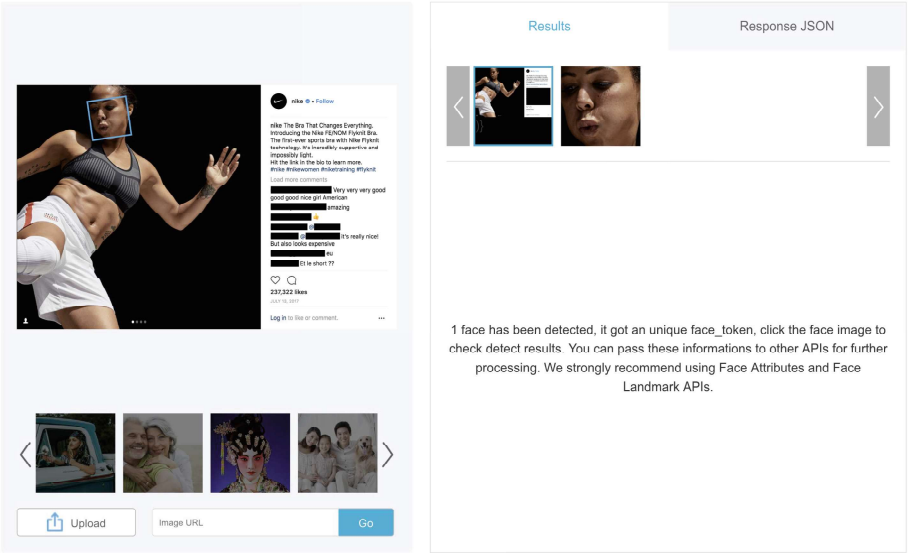
Instagram. Among the 132 brands, we find 82 brands with official Instagram accounts. We crawl 107,678 posts published by the brands from the creation of the account to November 28, 2017. Figure 1(a) shows an example post by Nike on Instagram with user comments. Then, we collect all 15.84M public comments made on the posts, which include the URLs of the profile pictures of users.

Facebook. We find 98 brands with official Facebook pages. Using Facebook Graph API, we collect 255,935 posts published from the creation of the page to November 15, 2017. Figure 1(b) shows an example post by Adidas on Facebook with user comments. We collect 1.68M comments of the posts together with the author name and ID. Using the author ID, we were able to access their profile images on Facebook. Due to timeout issues with the Graph API, we only obtained a random subset of the comments left on the 256K posts.

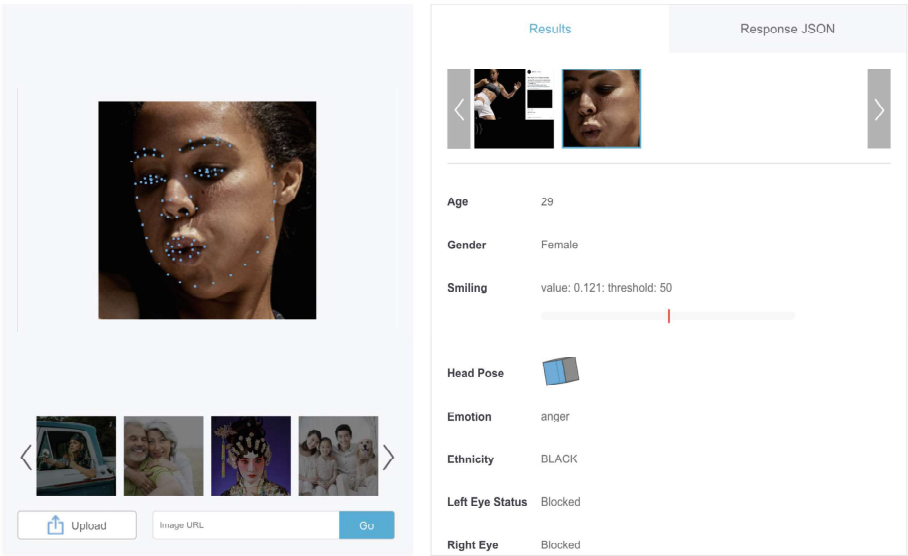
Data Cleaning. We focus our analysis on users who comment on posts by these brands and for whom Face++[3] detects a single human face in the profile image. Furthermore, we limit our analysis to (i) brand posts that show a single, human face, (ii) valid demographics, (iii) brands that have a sufficient number of posts showing faces and receiving comments, and (iv) brands where the act of commenting is generally an expression of positive emotion or implicit "liking". Next, we elaborate on these filtering steps.

Single Faces Only. As a first filter, we use Face++ to detect if a post contains a single, human face. Figure 2(a) shows an example of face detection by Face++ based on a post by Nike. For this, only posts with non-video images are considered. Posts where Face++ detects no or multiple faces are discarded. This leaves us with 14,852 posts for Instagram and 17,196 posts for Facebook. As we observed a false positive rate of around 27% for brand posts, e.g., due to cartoon images, such as the face in the Starbucks logo and other face-like patterns, we then manually filter all of the 32,048 posts, leaving us with 11,944 and 14,068 posts with single, human faces for Instagram and Facebook respectively. Note that for the profile pictures of users commenting on the posts, Face++ performs better as profile pictures tend to have a clean headshot, making manual post-filtering unnecessary.

[3] https://www.faceplusplus.com/.



(a) Face detection by Face++



(b) Demographic inferring by Face++

Fig. 2. An example of face detection and demographic inference by Face++ with a post by Nike on Instagram. We note that personal profile pictures and names are blackened for privacy concerns; however, we use unobfuscated images for Face++ when inferring the demographics of users.

Valid Demographics. Experiments on the accuracy of Face++ race inference, reported further down, show that the "Asian" category has many false positives. Hence, we limit our analysis to brand posts and user profiles detected as either White or African-American (AF-AM).

Sufficient Posts and Comments. To be able to perform meaningful statistical analysis, we ignore brands that do not have at least 50 posts with single faces, as well as a total of at least 100 distinct users with single face profile pictures commenting on these posts. This filter is applied separately to Instagram and Facebook.

Positive Engagement. By and large, we observe that the act of commenting on a brand's post is an expression of positive emotion toward the brand. However, for airlines and telecommunication providers, a large fraction of comments are negative or complaints. In our manual inspection, we observed that positive comments were often directly connected to the content of the brand's post (e.g., "I love this pumpkin latte!"), whereas negative comments were posted regardless of what the brand's post is about (e.g., "I tried to call the service center several times, but no one answered."). Hence, we exclude airline and telecommunication provider brands from our analysis as we expect different engagement dynamics in these cases.

At the end of this final filtering step, we are left with a total of 69 distinct brands, 40 on Instagram and 46 on Facebook with their 4,877 and 7,426 posts, respectively.

Demographic Inference. To automatically infer the demographics of the faces in the images posted by brands or in the profile pictures by users, we again use Face++, a deep learning-based image analysis tool, to detect faces given an image and to infer demographics of the faces in the image. When a face is detected, the Face++ API returns information, including a gender, an age estimate, and a race (White, African American (AF-AM) and Asian). Figure 2(b) shows an example of demographic inferring by Face++. Face++ returns other facial features such as whether the face is smiling or not and the emotion shown in the face, which could be relating to user engagement. In this study, we focus on the demographics of the faces in the brands' posts. Face++ has been validated and previously used in several lines of research looking at the demographics of Twitter users [1,3,25,30] and of Instagram Users[4]. In a recent comparative analysis, Face++ performed as good as other state-of-the-art tools such as Microsoft Azure Face API [12]. Moreover, Face++ showed more than 90% accuracy in gender and race detection with high-quality images [12].

Accuracy of Face Detection. To examine the accuracy of the demographic inference by Face++ in our setting, we create a crowdsourcing task on Figure Eight, formerly known as CrowdFlower. For each of the images, three workers first tag whether the image includes a single person's face. If a real face appears, they were asked to infer the gender (Female or Male), race (White, AF-AM, or

[4] http://selfiecity.net.

Asian), and age group (Minor [<= 17], Middle, Elders [>= 60] of the face. We limit access to our task to workers living in the US to reduce the cultural differences in inferring the race of a face in the image as a recent study has reported that the concepts and definitions of race are not uniform between countries [28]. We find that 29% of the images have either no real person's face or multiple faces. Among the images with a single face, the accuracy of inferring gender and age group by Face++ is 86.7% and 87.3%, respectively, while for race detection, the accuracy is lower at 71.5%. 84.22% (77.78%) of the posts with White (AF-AM) faces are correctly inferred by Face++. However, we find that the accuracy of the posts with Asian faces is only 21.77% — 60.48% were classified as White by Face++. Given the results, we decide to consider only the posts with White and AF-AM faces for this study. We note that only 2.0% (1.4%) of white (AF-AM) are misclassified as Asian.

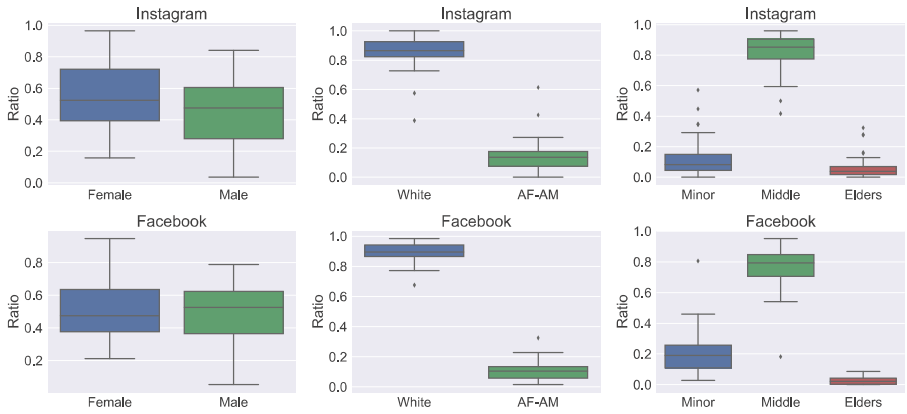


Fig. 3. Demographic distributions by gender, race, and age groups on Instagram and Facebook. Each box and whisker element summarizes the values across different brands. Most brands produce roughly gender balanced content, with White (˜80%) and 18–60 (˜80%) demographics representing the majority of ad images with individual faces.

Users with Faces. We use Face++ to infer the demographics of the faces in the profile pictures of the users. For this work, we only consider users with a single, not Asian, face in their profile picture. Furthermore, we focus on users who leave comments on those 14,303 validated posts, which are 171,179 and 8,841 users for Instagram and Facebook, respectively. Since Face++ performs well for user profile pictures, which tend to be clean, head-shot, we use the results by Face++ [12]. On Facebook, we find 48% males and 52% females in our user dataset. 86.8% of the users are detected as "White" and 13.2% as "African American (AF-AM)." We group users by their age in three categories: those with an inferred age below or equal to 17 (denoted as "Minor"), 18 to 60 ("Middle"), and over or equal to 60 ("Elders"). The majority group is Middle which entails 78.4% of users, then Minors with 16%, and Elders with 5.6%.

Compared to FB, Instagram has more female users (60.1%), less white users (85%), more AF-AM (15%), less minors (3.6%), more middle-age users (92.3%), and less elders (4.1%).

Data Ethics. All the data collected comes from publicly accessible data sources. The collected data is securely stored and processed on a password-protected machine. Results are reported for anonymous, aggregate results. All examples of individual-level data, such as comments, are obfuscated, with the exception of posts from brands, featuring professional models.

4 Diversity in Online Ads

We begin by characterizing what faces brands are putting out on their social media. Figure 3 shows gender, race, and age distributions for Instagram and Facebook as boxplots. For Facebook, across all posts with faces, the detected gender is fairly balanced, with 50.6% females. Notably, 89.6% of the users are

Table 1. Top 3 brands for the fraction of each demographic in their posts on Instagram and Facebook. Examples: 81% (79%) of the faces in Under Armour Instagram (Facebook) posts with a single face are males.

Group	Rank	Instagram		Facebook	
		Brand	Ratio	Brand	Ratio
Female	1	Clinique	0.96	Victoria's Secret	0.95
Female	2	Walgreens	0.93	Kohl's	0.84
Female	3	Neutrogena	0.92	Tiffany & Co	0.82
Male	1	Gatorade	0.84	Under Armour	0.79
Male	2	Sprite	0.82	Oracle	0.79
Male	3	Under Armour	0.81	Gillette	0.77
White	1	Ford	1.00	Target	0.99
White	2	Domino	1.00	Victoria's Secret	0.98
White	3	Cisco	0.97	Lowe's	0.98
AF-AM	1	Sprite	0.61	Allstate	0.32
AF-AM	2	Gatorade	0.42	JP Morgan	0.23
AF-AM	3	Google	0.27	McDonald's	0.19
Minor	1	Kroger	0.57	Pampers	0.81
Minor	2	Costco	0.45	Dollar Tree	0.46
Minor	3	Amazon.com	0.35	The Home Depot	0.41
Middle	1	Neutrogena	0.96	Oracle	0.95
Middle	2	Nordstrom	0.95	Uber	0.93
Middle	3	Tiffany & Co	0.94	Xbox	0.93
Elders	1	KFC	0.32	Lowe's	0.09
Elders	2	Whole Foods	0.28	Costco	0.08
Elders	3	Ford	0.28	Kroger	0.07

detected as "White", with 10.4% "African American". The majority age group is middle age (76.8%), followed by minors at 20.5%, and elders at 2.7%. Compared with FB, Instagram has more females (55.3%), less White users (85.6%), more AF-AM users (14.4%), less minors (12.1%), more middle-age users (81.5%), and more elders (6.4%).

Table 1 lists the top three brands ranked by the fraction of faces of a particular demographic attribute in their posts on Instagram and Facebook. Those brands with skewed fraction of a demographic group are ones with a particular target group. For example, cosmetics and sports brands tend to have the highest male or female ratio as they are the one who would consume their product. Ford, for example, has the highest number of White elders models in their posts, revealing their target group. We note that our analysis is based on single-face brand posts.

5 Gap Between Demographics of Ads and Users

Observing predominantly White faces in online advertising could simply be a reflection of the brand's user demographics. Here, we compare the alignment of the demographic distribution in a brand's posts to that of their user base. To this end, we compute the difference between the ratio of a certain demographic group in the ads and that of the engaging users for each brand. A positive value indicates that the corresponding group is over-represented in the ads.

Figure 4 shows the gap between demographics of ads and users for different demographic attributes and by platform. On both Instagram and Facebook, males are over-represented in the brands' ads as the users skew female. Generally, the gender gap is small with a few exceptions, such as Victoria's Secret, which, compared to the many men engaging with the brand, surprisingly over-represents women in their models. On Facebook, AF-AM are under-represented – on average, 10.4% of the posts include AF-AM models, while 13.2% of the FB users are AF-AM. On Instagram, minors and elders are slightly over-represented as Instagram is skewed toward middle-age users. The overall differences across all demographic groups are small.

To look into deeper into the diversity gap of each brand, Table 2 lists the top five brands for the difference of a fraction of each demographic between ads and users. Here, we present a few noteworthy examples.

Several tech companies such as Intel (FB) over-represent female faces. Or, rather, the content they put out is gender balanced but their audience skews toward male. Victoria's Secret (IG), on the other hand, depicts mostly women but attracts many comments from men. Uber (FB) and Paypal (FB, IG) over-represent White faces in their single-face posts, compared to users engaging with the brands. In fact, they do present diverse actors on their social media page. However, White faces are more common in the their images with *single* faces. Sprite and Gatorade (IG) over-represent AF-AM in their posts and many Whites engage.

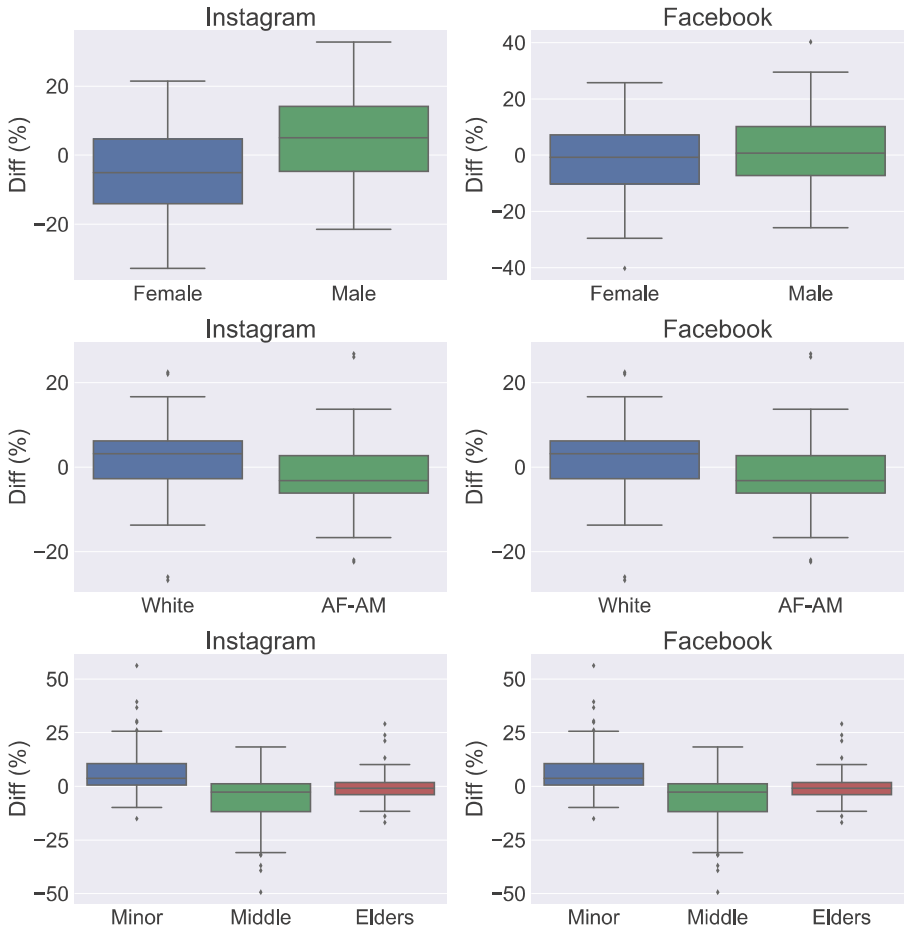


Fig. 4. Difference between the distribution of the demographics of content produced by a brand compared to the users engaging with the brand. Each box and whisker element summarizes the values across different brands. On both Instagram and Facebook, men are over-represented (+1.4% and +5.6%, respectively) compared to a (more female-skewed) user base. AF-AM are under-represented (−2.8% on FB) and the youngest demographics (+4.6% on FB and +8.5% on IG) are over-represented compared to users engaging with the brands.

There are, however, brand-specific outliers with Pampers (IG) most strongly over-representing infants in its posts. Pampers mostly uses infants wearing its products as models for its posts while the engaging users are the Middle group. Another entertaining outlier is KFC (IG), which over-represents the older and White (IG) demographics compared to its user base. The reason is the actor representing a white-haired, white-bearded Colonel Sanders, KFC's figurehead.

Table 2. Top 5 brands for the difference of fraction of each demographic between Brand posts and Users. N(posts) is the number of posts used to compute the percentage difference for each brand. Examples: faces in Victoria's Secret Instagram posts are 18% more female than the (more male) users engaging with the brand. The models used by Pampers on Facebook – infants – are 37% more likely to be in the youngest age group compared to the engaging users.

Group	Rank	Instagram			Facebook		
		Brand	N(posts)	Diff(%)	Brand	N(posts)	Diff(%)
Female	1	Walgreens	101	+21%	Intel	106	+26%
Female	2	Victoria's Secret	532	+18%	Purina	34	+25%
Female	3	Progressive	41	+15%	Victoria's Secret	1443	+21%
Female	4	Pantene	131	+11%	Paypal	71	+19%
Female	5	Google	33	+8%	Domino	55	+16%
Male	1	Starbucks	31	+33%	Sysco	72	+40%
Male	2	Netflix	74	+33%	Allstate	77	+30%
Male	3	Whole Foods	50	+32%	Netflix	216	+24%
Male	4	Sprite	62	+24%	The Home Depot	44	+23%
Male	5	UPS	47	+21%	Subway	67	+20%
White	1	Progressive	41	+22%	Kroger	71	+22%
White	2	Paypal	66	+16%	Uber	60	+17%
White	3	Domino	28	+11%	Paypal	71	+13%
White	4	KFC	37	+11%	Western Digital	91	+11%
White	5	Ford	40	+10%	Target	68	+11%
AF-AM	1	Sprite	62	+27%	Allstate	77	+14%
AF-AM	2	Gatorade	120	+26%	Pizza Hut	63	+12%
AF-AM	3	Costco	38	+11%	McDonald's	31	+8%
AF-AM	4	Starbucks	31	+9%	The North Face	103	+6%
AF-AM	5	Google	33	+8%	The Home Depot	44	+6%
Minor	1	Kroger	84	+56%	Pampers	154	+37%
Minor	2	Costco	38	+39%	Dollar Tree	37	+26%
Minor	3	Amazon.com	23	+30%	Domino	55	+21%
Minor	4	Walmart	93	+30%	Aflac	56	+17%
Minor	5	Target	48	+26%	CVS	68	+17%
Middle	1	Neutrogena	24	+3%	Uber	60	+18%
Middle	2	Electronic Arts	29	+3%	Purina	34	+15%
Middle	3	Nordstrom	212	+1%	Cisco	165	+12%
Middle	4	Victoria's Secret	532	+0%	Intel	106	+11%
Middle	5	Tiffany & Co.	234	−0%	Oracle	210	+10%
Elders	1	KFC	37	+29%	Pizza Hut	63	+5%
Elders	2	Whole Foods	50	+24%	JP Morgan	57	+4%
Elders	3	Ford	40	+21%	Sysco	72	+4%
Elders	4	Netflix	74	+13%	Thermo Fisher Sci.	48	+4%
Elders	5	Starbucks	31	+10%	Lowe's	46	+3%

6 Impact of Gender, Race, and Age on Users' Engagements with Ads

We now turn our focus to how users are *reacting* to the demographics represented in online ads. In particular, we examine whether the users' reactions provide evidence for in-group or out-group preference. We bucket each post by the demographic group it shows, say, a woman. We then compute the fraction of women among the users commenting on the post. This fraction is then compared to the fraction of women commenting on posts that do *not* show a woman (i.e., those that show a man). If showing a woman is linked to observing a higher fraction of women among the engaged users, then we take that as evidence for in-group preference. We then run a two-sided t-test on the two sets of values, (a) fractions of the engaged "female" users on posts showing a female model and (b) fractions of the engaged "female" users on posts showing a non-female (e.g., male) to test whether the differences of those fractions are statistically significant. For the age, we use the following two sets of values, (a) fractions of the engaged "minor" users on posts showing a minor and (b) fractions of the engaged "minor" users on posts showing a non-minor (e.g., middle or elders) to run a two-sided t-test.

Table 3 shows the mean of (absolute) percentage differences between the two fractions. For all cases shown, group X is linked to a larger fraction of users of group X engaging with the post. For this analysis, we use brands with more than 20 posts having at least 20 face-detected users, resulting in examining 17 brands on Instagram. When using 10 and 15 as a threshold, we find the trends are consistent.

Table 3. In-group Preference Theory test result. Values indicate the mean absolute difference between the engaging user demographics when a post does or does not have the corresponding demographic attribute. For example, for Instagram posts showing a female face, the fraction of female users commenting on this post is 13.3% higher compared to posts not showing a female face (i.e., those showing a male face). For most demographic groups we observe a statistically significant "demographic resonance".

Hypothesis	Female	Male	White	AF-AM	Minor	Middle	Elders
Instagram	+13.3%***	+13.3%***	+4.2%***	+4.2%***	+1.0%***	0.5%	+1.4%***

Significance code: <0.001 ***, <0.01 **, <0.05 *

Our results provide evidence for in-group preference for gender – when the image shows a female/male, female/male users engage more. However, this might be an artifact as this analysis was done *across* brands, and a brand with more female models generally also has a more female user base. Thus, we repeat the analysis *within* each brand, i.e., we look at whether the same brand sees variance in the composition of the engaging users, depending on the demographics of the model in their post.

For four brands, we find statistically significant in-group preferences for gender among the engaging users ($p < 0.05$). For Electronic Arts, Harley Davidson, Polo Ralph Lauren, and Under Armour, female users comment more when posts show a female face (+3.1%, +6.7%, +10.3%, and +6.7%, respectively). Furthermore, we find in-group preference for racial group for the five brands: Michael Kors, Netflix, Polo Ralph Lauren, Victoria's Secret, and YouTube (+2.7%, +5.3%, +14%, +3%, and +2.1%, respectively). We only applied this analysis to binary groups and not to the tertiary age categories. We also note that we do not find any significant cases for out-group preferences (or polarized appraisal theory).

7 Limitations and Discussions

As with any machine inference method, Face++ does occasionally misclassify users. This noise does not, however, generally affect the qualitative conclusions drawn for the following reasons. First, we focus on comparing *relative* differences between distributions, e.g., between the inferred demographics in ad images and in user profiles. Most biases would not affect conclusions of the type "there are more male-looking faces in ads compared to user profiles on Instagram", even if individual faces are misclassified. Second, noisy inference actually *weakens* the results for in-group preference. In the most extreme case of noise, if all demographic labels were assigned independently and at random, it would be impossible to observe, as we do, a systematic link between the demographics in ad images and in the profile images of engaging users. In the future, we plan to correct for this attenuation by modifying the existing methods [27].

For a more focused analysis, we deliberately decided only to consider images with a *single* face. However, it would be interesting to study the engagement with ad images showing several faces, particularly mixed race couples. Occasionally, such images spark an outcry from far-right user groups [21]. We also observed that Uber and Paypal, two brands that over-represent White faces in their single-face posts, have higher levels of diversity in posts with several faces.

All of our analyses only considered the biggest U.S.-based business-to-consumer brands. Such brands are likely to have professional staff highly attuned to potential sensitivities surrounding the representation of minorities. Hence, broadening the scope to include smaller brands might yield different results. Similarly, broadening the scope to include other countries would create a less U.S.-centric view on the topic.

We use the number of comments as a signal for engagement. However, the number of comments is not the same as the number of likes. "Likes" signify how interesting the content is to users, whereas "Comments" quantify the level of discussion on the social media [3]. Thus, our results should not be generalized to the user engagement based on the number of likes. We did, however, remove brands where we observed that commenting was often a sign of *dislike*.

Our current work looks at three important demographic attributes: age, gender, and race. However, the problem of lack of diversity in advertising extends well beyond these three attributes. A case in point is that members of the LGBTQ community are both under- and misrepresented in mainstream advertising [29]. The same holds true for people with disabilities[5]. It might be possible to study some of these issues using similar methods to those used in this work. For example, computer vision could be used to detect wheelchairs or crutches. Similarly, computer vision could detect both a model's and a commenting user's weight status [14] to study if, say, the body shapes of advertising models are representative of the distribution of body shapes in the general population. Prior work looking at the social network structure found evidence for weight-based homophily [15], and one might expect a similar in-group preference for the models used in advertising, where users might be more inclined to engage with an ad that depicts a person with a similar weight status. Computational methods previously developed for the Bechdel Test [9] could potentially also be applied to study the importance members of a particular minority group are given in advertising videos.

Another application of our work beyond traditional advertising is political campaigning. In politics, researchers have studied how the physical appearance of candidates is related to the vote choice and found that attractiveness, familiarity, babyfacedness and age are predictors of vote choice, but after controlling for competence, the effects remained marginal [22]. Our computational approach could expand such studies to understand better how much candidates' physical appearance determines the engagement from the potential voters they attract.

As online advertising is becoming more and more targeted and more and more personalized, it is worth contemplating how recent advances in using computers to generate fake but realistic looking faces will change this industry [13]. For example, would users be more likely to buy a certain product if the model in the ad was a virtual digital twin of themselves? In studies on delayed gratification and how to increase saving behavior, Hershfield et al. observed that people were more likely to accept later monetary rewards when they interact with realistic computer renderings of their future selves [11]. Similarly, showing people a picture of their own aged face under continued exposure to sunlight without adequate protection can increase the use of sunscreen and, potentially, help lower skin cancer rates [4].

8 Conclusion

This paper gives a summary of demographic diversity in online advertising on Instagram and Facebook. Using computer vision to infer the demographics of faces in posts by major brands and in the profile images of users engaging with

[5] See https://www.indy100.com/article/disability-adverts-tv-uncomfortable-study-maltesers-wheelchair-8253546 or https://www.campaignlive.co.uk/article/invisibles-why-portrayals-disability-so-rare-advertising/1407945 for two opinion articles on the topic.

these posts, we observe the following. Most brands come close to gender parity in terms of the faces in their posts. However, the majority of faces are White, and only a few older faces are found. Comparing what the brands put out with the demographics of the engaging users, we observe that (i) on both Instagram and Facebook, women are underrepresented in the brands' posts; (ii) on Facebook, White faces are slightly over-represented; and (iii) on both Instagram and Facebook, young faces are slightly over-represented, and middle-age faces are slightly under-represented. There are, however, brand-specific outliers with Pampers most strongly over-representing infants in its posts, whereas KFC over-represents a particular older face – that of Colonel Sanders, the brand's figurehead. Importantly, we provide evidence for resonance between the demographics depicted in a particular post and that of the engaging users. For example, brand posts with an African-American face have, on average, a 4.2% higher percentage of African-Americans among the users engaging with these posts compared to posts without an African-American face.

We believe that our methodology of computationally studying diversity in online advertising could lead to more in-depth analyses, such as sector-specific break-downs or taking the actual content of the comments made by the engaging users into account. It would also be interesting to go beyond single-face images and, in particular, look at the reactions to mixed race images in advertising.

References

1. An, J., Weber, I.: #greysanatomy vs. #yankees: demographics and hashtag use on Twitter. In: ICWSM (2016)
2. Appiah, O.: Black, White, Hispanic, and Asian American adolescents' responses to culturally embedded ads. Howard J. Commun. **12**(1), 29–48 (2001)
3. Bakhshi, S., Shamma, D.A., Gilbert, E.: Faces engage us: photos with faces attract more likes and comments on instagram. In: CHI (2014)
4. Brinker, J.T., Schadendorf, D., Klode, J., Cosgarea, I., Rösch, A., Jansen, P., Stoffels, I., Izar, B.: Photoaging mobile apps as a novel opportunity for melanoma prevention: pilot study. JMIR Mhealth Uhealth **5**(7), e101 (2017)
5. Bush, R.F.: White consumer sales response to black models. J. Mark. **38**(2), 25–29 (1974)
6. Cagley, J.W., Cardozo, R.N.: White response to integrated advertising. J. Advert. Res. **10**(2), 35–39 (1970)
7. Estes, A.C.: Brief History Racist Soft Drinks (2013). https://goo.gl/XvYGD7
8. Forehand, M.R., Deshpandé, R., Reed II, A.: Identity salience and the influence of differential activation of the social self-schema on advertising response. J. Appl. Psychol. **87**(6), 1086–1099 (2002)
9. Garcia, D., Weber, I., Garimella, V.R.K.: Gender asymmetries in reality and fiction: the Bechdel test of social media. In: ICWSM (2014)
10. Henderson, J.J., Baldasty, G.J.: Race, advertising, and prime-time television. Howard J. Commun. **14**(2), 97–112 (2003)
11. Hershfield, H.E., et al.: Increasing saving behavior through age-progressed renderings of the future self. J. Mark. Res. **48**, S23–S37 (2011)

12. Jung, S., An, J., Kwak, H., Salminen, J., Jansen, B.: Assessing the accuracy of four popular face recognition tools for inferring gender, age, and race. In: ICWSM (2018)
13. Karras, T., Aila, T., Laine, S., Lehtinen, J.: Progressive growing of GANs for improved quality, stability, and variation. In: ICLR (2018)
14. Kocabey, E., et al.: Face-to-BMI: using computer vision to infer body mass index on social media. In: ICWSM (2017)
15. Kocabey, E., Ofli, F., Marin, J., Torralba, A., Weber, I.: Using computer vision to study the effects of BMI on online popularity and weight-based homophily. In: SocInfo (2018)
16. Lee, Y.J., Kim, S.: How do racial minority consumers process a model race cue in CSR advertising? A comparison of Asian and White Americans. J. Mark. Commun. 1–21 (2017)
17. Lloyds: Lloyds Diversity Report (2016). https://goo.gl/ehck2D
18. Martin, B.: The influence of ad model ethnicity and self-referencing on attitudes: evidence from New Zealand. J. Advert. **33**(4), 27–37 (2004)
19. Miller, A.N., Kinya, J., Booker, N., Kizito, M., wa Ngula, K.: Kenyan patients attitudes regarding doctor ethnicity and doctor-patient ethnic discordance. Patient Educ. Couns. **82**(2), 201–206 (2011)
20. NPR: This Ad's For You (2015). https://goo.gl/z8jQTB
21. NYTimes: Upbeat Interracial Ad for Old Navy Leads to Backlash. Twice (2016). https://goo.gl/RHkjsS
22. Olivola, C.Y., Todorov, A.: Elected in 100 milliseconds: appearance-based trait inferences and voting. J. Nonverbal Behav. **34**(2), 83–110 (2010)
23. Paek, H., Shah, H.: Racial ideology, model minorities, and the "not-so-silent partner:" stereotyping of Asian Americans in U.S. magazine advertising. Howard J. Commun. **14**(4), 225–243 (2003)
24. Qualls, W.J., Moore, D.J.: Stereotyping effects on consumers' evaluation of advertising: Impact of racial differences between actors and viewers. Psychol. Mark. **7**(2), 135–151 (1990)
25. Reis, J., Kwak, H., An, J., Messias, J., Benevenuto, F.: Demographics of news sharing in the U.S. Twittersphere. In: HT (2017)
26. Sierra, J.J., Hyman, M.R., Torres, I.M.: Using a model's apparent ethnicity to influence viewer responses to print ads: a social identity theory perspective. J. Curr. Issues Res. Advert. **31**(2), 41–66 (2009)
27. Spearman, C.: The proof and measurement of association between two things. Am. J. Psychol. **15**(1), 72–101 (1904)
28. Travassos, C., Williams, D.R.: The concept and measurement of race and their relationship to public health: a review focused on Brazil and the United States. Cadernos de Saúde Pública **20**, 660–678 (2004)
29. Tsai, W.H.S.: Assimilating the queers: representations of lesbians, gay men, bisexual, and transgender people in mainstream advertising. Advert. Soc. Rev. **11**(1) (2010)
30. Zagheni, E., Garimella, V.R.K., Weber, I.: Inferring international and internal migration patterns from twitter data. In: WWW (2014)