Do you see what I see?
A Cross-cultural Comparison of Social Impressions of Faces

Amanda Song[1]
Cognitive Science
UC, San Diego
mas065@ucsd.edu

Weifeng Hu[1]
Computer Science and Engineering
UC, San Diego
weh031@eng.ucsd.edu

Devendra Pratap Yavav
Computer Science and Engineering
UC, San Diego
dpyadav@eng.ucsd.edu

Fangfang Wen & Bin Zuo
Psychology
Central China Normal University
wenff/zuobin@mail.ccnu.edu.cn

Edward Vul
Psychology
UC, San Diego
evul@ucsd.edu

Garrison Cottrell
Computer Science and Engineering
UC, San Diego
gary@eng.ucsd.edu

Abstract

Research has suggested that social impressions of faces made by Western and Eastern people have different underlying dimensionalities. However, the individual level consistency, the group-level agreement of rater groups, and the interactions between face ethnicity, rater ethnicity, and social impression traits remain largely unknown. In this paper, we perform a large-scale data-driven cross-cultural study of facial impressions, and illustrate the idiosyncrasies and similarities behind Caucasian and Asian participants in their social impressions of faces from both ethnicity groups. Our study illustrates multiple interesting findings: (1) Asians rate faces lower on most positive traits, compared with Caucasian raters, and they have more diverse opinions than Caucasians. (2) Caucasian faces receive higher average ratings on social impression traits related to warmth due to the preponderance of smiles in Caucasian images, but similar mean scores on traits related to capability, compared to Asian faces. (3) Caucasians and Asians disagree most on capability related traits, especially on "responsible" and "successful." Opinions on these two traits diverge more on Asian than on Caucasian faces. Our findings provide new insights on the nuances of cross-cultural differences in social impressions of faces.

Keywords: First impressions; cross-cultural comparison; large scale online experiment; statistical analysis; face perception

Introduction

Although we are told not to judge a book by its cover, we nonetheless do it frequently when we see people for the first time. At the first sight of a new person, our brain automatically forms impressions of them – how trustworthy are they? how kind? what is their social status? Even if these spontaneously formed social impressions are not objectively true (Olivola & Todorov, 2010) (consider the case of Ted Bundy!), they nevertheless affect important aspects of our lives including interpersonal relationships, hiring and financial decisions (Rezlescu, Duchaine, Olivola, & Chater, 2012), even legal judgments (Wilson & Rule, 2015) and electoral outcomes (Todorov, Mandisodza, Goren, & Hall, 2005; Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015).

Regardless of their dubious accuracy, people have fairly high agreement in the facial impressions they form (Falvello, Vinson, Ferrari, & Todorov, 2015). This agreement is also reflected in the image-level facial features that drive impression formation, such as the apparent age, gender, race and expressions of the face (Ebner, 2008; Adams Jr, Hess, & Kleck, 2015; Zebrowitz, Kikuchi, & Fellous, 2010). This agreement also arises in the correlation structure among the impressions of different traits, that seem to fall along three factors: warmth, competence and youthful-attractiveness (Todorov et al., 2015; Sutherland et al., 2018).

Despite these universal aspects of facial impressions, they are also influenced by the cultural background of the viewer (Todorov et al., 2015). This should be no surprise. Research suggests that culture even shapes visual perception (Nisbett & Miyamoto, 2005), and it certainly shapes our social norms, expectations, and values. For instance, East Asians have been characterized as being more collective and holistic, whereas Westerners have been more individualistic and analytic (Hofstede, 1980; Oyserman, Coon, & Kemmelmeier, 2002); perhaps this would make friendlier looking people seem more capable to Asian viewers. Moreover, culture also influences our eye movements when we look at faces (Blais, Jack, Scheepers, Fiset, & Caldara, 2008), which may mean that different facial features will be more salient to viewers from different cultures. Altogether, cultural differences in facial impressions seem quite plausible, and their social importance may be increasingly large, given the preponderance of face-to-face international interactions over video conferencing and social media.

Previous studies of cross-cultural facial impressions have identified similarities and differences in a number of individual traits such as attractiveness (Cunningham, Roberts, Barbee, Druen, & Wu, 1995) and intelligence (Krys, Hansen, Xing, Szarota, & Yang, 2014). Yet most prior studies used a small set of strictly controlled face stimuli, limiting the generalizability to everyday face photos with real-world variation. Furthermore, prior studies explored one trait at a time with different face stimuli, compromising any across-trait comparisons in cultural agreement levels. Bridging this gap requires large-scale cross-cultural studies of many traits using a large set of real-world facial images.

Here we compare how Chinese Asians and American Caucasians (henceforth, Asians and Caucasians, with the country understood) form impressions of 15 traits for each of thou-

---

[1]Joint first authors

sands of real-world Asian and Caucasian face images. We consider 15 social impression traits that cover three major categories: (1) warmth related traits, such as warm, happy, friendly and kind, (2) physical appearance appraised traits, such as attractive and healthy, (3) capability related traits, such as capable, diligent, high-social status, intelligent, powerful, responsible and successful. Our study shows both cross-cultural universals and differences in impression formation as a function of rater ethnicity, face ethnicity, and face gender. We should state here that the Caucasian coauthors found some of the Asian ratings to be so surprising as to be unbelievable, while the Asian coauthors agreed with these ratings and were also surprised by some of the Caucasian ratings.

Large Scale Dataset Collection

In this study, we aim to compare Western and Eastern cultural differences in the social impression perception of Caucasian and Asian faces. To this end we had Caucasian and Asian subjects rate their first impressions of thousands of Caucasian and Asian faces on 15 socially relevant traits.

Image Stimuli

We selected 1,099 Caucasian faces from the US 10K Adult Database (Bainbridge, Isola, & Oliva, 2013). For Asian faces, we followed a procedure similar to (Bainbridge et al., 2013) and collected Asian faces from the online image search engine (Microsoft Bing). We gathered the most frequently used Chinese first names and last names for both genders, and then used the combination of first and last names as the keywords to search. We then downloaded the first few face images that were associated with the name combination. After the original images were downloaded, we ran a face detector to crop the face region from the image, and removed the images if they met one of the conditions: (1) the face region resolution was lower than 200×200; (2) the face was that of a celebrity (to the best of our knowledge); (3) more than half of the face was occluded; (4) the face belonged to an infant. After preprocessing, we kept 1,638 Asian faces. Figure 1 shows a few examples of the Caucasian and Asian face stimuli.

Figure 1: Examples of Caucasian and Asian face stimuli.

Social Impression Traits

We used 15 social impression traits that align with the three key dimensions commonly found in prior research on first impressions from faces (Sutherland et al., 2018; Todorov et al., 2015): (1) warmth/approachability related traits: friendly, happy, kind, trustworthy, and warm; (2) attractive/youthful traits: attractive, healthy; and (3) competence related ones:

calm, capable, diligent, (of) high social status, intelligent, powerful, responsible and successful.

Participants' Task

The main task is to indicate their first impression of an image on a specific trait by providing a rating on a scale of 1-9, as shown in Figure 2. To avoid demand effects, we asked people how they think others would perceive the face, which we previously found reduces social desirability biases when offering potentially contentious opinions. Participants saw multiple faces in a sequence, and rated one face at a time.
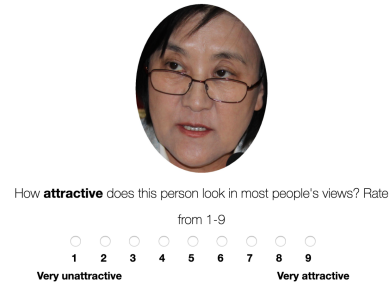

Figure 2: First impression rating task page.

Caucasian Rater Data Collection

We recruited Caucasian participants using Amazon Mechanical Turk (Litman, Robinson, & Abberbock, 2016). There were 428 Caucasian subjects (254 are female), with a median age range of 30-39 years old. Since rating social traits is a subjective task, we designed a screening mechanism to ensure participants were paying attention to the task.

The screening consisted of 20 randomly selected faces and a randomly-selected social trait to rate (the interface is shown in Figure 2). The 20 faces were presented, then they were shuffled and shown again, resulting in a 40-trial sequence. If a participant's reliability was significantly above zero, and they used at least three different scores from the 9 point scale, the participant was considered to have passed the "reliability test." Reliable participants were invited to complete as many more main tasks as they wanted. In the main task, there were 100 faces. As in the screening task, the participants rated the faces on a single trait, one face at a time. In each task, the 100 faces contained 90 unique faces of the same ethnicity, and 10 repeated faces randomly drawn from the 90 faces. Every image-trait combination was rated at least ten times by our Caucasian participants. We found the reliability was adequate for subjects that passed the first screening, so we did not analyze the 10 repeated faces further.

Asian Rater Data Collection

We recruited Chinese participants via the data100 website (https://www.data100.com.cn) as well as via online volunteer sourcing. The task instructions and all traits were translated into simplified Chinese and back-translated into English to ensure that the Asian participants were rating the same social traits as the Caucasians. While Caucasian participants

were able to participate in multiple tasks, we were unable to find a platform that allowed for this in China. Hence, Asian raters participated in the task just once due to the limitations of data100. Because of this, we integrated the screening process into the rating sequence, using the same criterion as the Caucasian subjects - 20 faces were repeated, and data from raters who were self-consistent were kept.

In total, 23,304 Asian participants were recruited; 14,338 were female and the median age range was 20-29 years old. Each image in our dataset was rated at least ten times by Asian participants on every trait.

Dataset Analysis and Results

Individual Reliability

For the screening, we computed the test/retest Spearman correlation (Zwillinger & Kokoska, 2000) on the repeated trials. Our participants were very self-consistent, with an average Spearman correlation above 0.7 for both rater ethnicities.

Group Level Consistency

We used one-way intraclass correlation coefficient (ICC) to measure group level agreement by evaluating the ratio of the variance of item random effects to the overall rating variance. Figure 3 shows the ICCs of each trait for each demographic participant group ranked by overall average ICC. Asian raters have a lower ICC than Caucasian raters; this lower group-level consistency among Asian raters may reflect more diverse opinions about how to evaluate these social traits. Within the same ethnic group, there were no statistically significant differences between male participants and female participants. Similar to previous research (Hehman, Sutherland, Flake, & Slepian, 2017), we found that there is more agreement for traits representing appearance-based appraisals (e.g., happy, warm, friendly, kind, attractive), than for competence-related traits (such as diligent, capable, intelligent, and powerful); this effect should not be too surprising as attractiveness, youth, and propensity to smile are much more evident in a picture than traits like diligence.

Group Mean Analysis

First, we examined how Caucasian and Asian participants rated faces differently on average for each trait. We divided the participants by ethnicity and subdivided the face images into four demographic groups according to the race and gender of the face. Then, for each face image group, we plot the mean ratings across all Asian raters against all Caucasian raters. The results are shown in Figure 4. A follow up ANOVA in Figure 5 further illustrates the variance explained by each single factor and the interactions among them.

We observe that Asian raters give overall lower ratings than Caucasian raters. All of the ratings in Figure 4, including happy, are significantly higher for Caucasians over Asians ($p < 0.01$). This trend aligns with prior results arguing that compared to Chinese participants, European Americans tend to emphasize the positive, and downplay the negative (Sims et al., 2015).
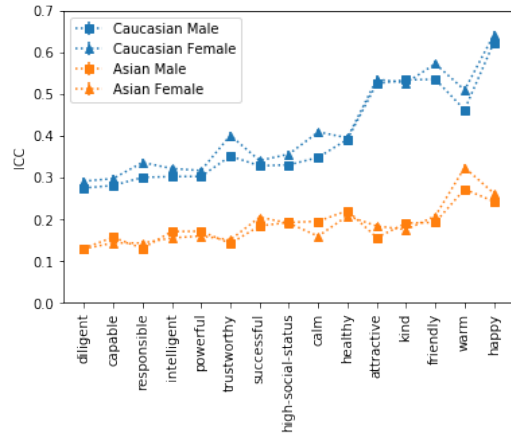

Figure 3: ICC for Caucasian and Asian participants separated by rater ethnicity and gender, sorted from low to high based on average ICC.

Second, we find that on average, images of Caucasians are rated higher than images of Asians across all traits ($\beta = 0.22$, se $= 0.008$), in particular for **warmth related traits** ($\beta = 0.41$, se $= 0.014$). However, smiling seemed more common among the Caucasian faces than Asian faces in our pseudo-randomly sampled image set. To correct for this we tagged whether a facial image is smiling using AWS Rekognition. We found that 75% of Caucasian images were smiling, while only 31% of Asian images were. Correcting for the effect of smiling reverses the image ethnicity effect, such that warmth related traits are rated lower for Caucasian smiling images than Asian smiling images ($\beta = -0.14$, se $= 0.017$), and lower for Caucasian non-smiling images than Asian non-smiling images ($\beta = -0.56$, se $= 0.019$). Table 1 shows the dramatic disparities in smiling rates and the reversal of the Caucasian advantage when smiling is controlled. This pattern of results is suggestive of raters implicitly correcting for the different baserate of smiles among Asian and Caucasian faces; thus making a smile more diagnostic for Asian faces, and a lack of smile more diagnostic for Caucasian faces. Regardless of the specific reason, the direction and magnitude of the mean difference in ratings for Caucasian images appears to be driven entirely by the preponderance of smiles in Caucasian images, not due to differences in how Asians and Caucasians are perceived.

Table 1: Average ratings across all warmth related traits when separating images by ethnicity and whether they are smiling.

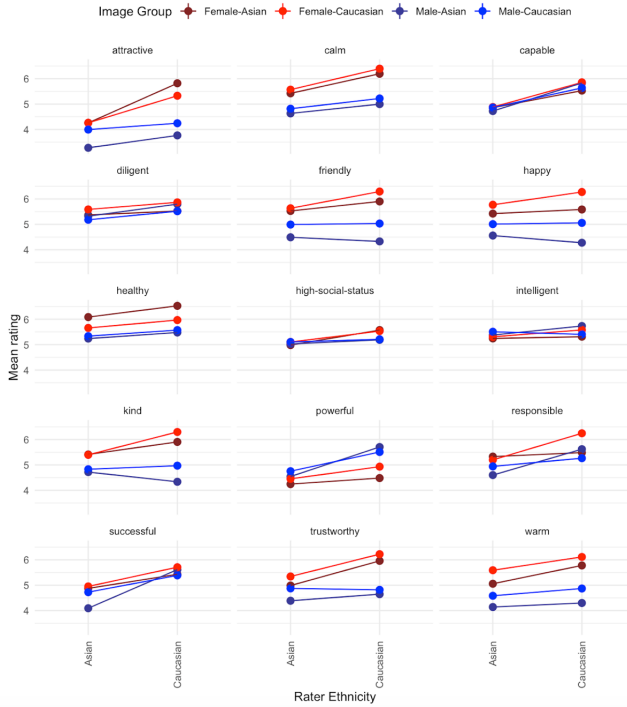|		Asian Raters	Caucasian Raters	%
Non-smiling	Asian	4.56	4.34	69%
Non-smiling	Caucasian	4.13	3.65	29%
Smiling	Asian	5.52	6.72	31%
Smiling	Caucasian	5.60	6.35	71%

Figure 4: For each trait, we split the images based on the gender and ethnicity of the face, and assessed Caucasian and Asians raters' mean ratings and standard errors for the four image groups. Overall, Caucasian raters give higher mean ratings on faces and Caucasian faces in general receive higher ratings. Interaction patterns of specific traits are elaborated in the main text.

Besides warmth related traits, we can see interesting cross-cultural similarities/differences and interaction patterns in the following traits by examining Figure 4 and 5 closely. For each effect in each trait we report the Tukey HSD/Range corrected 95% confidence interval on the relevant pairwise difference.

Physical Appearance Related Traits

Attractive: Images of Caucasian females are rated as less attractive than those of Asian females by both Caucasian raters $[-0.58, -0.38]$, and Asian raters $[-0.25, -0.08]$.

Capability Related Traits

High social status: Caucasians rate males as lower in social status than females; this holds true for both Caucasian images $[-0.43, -0.23]$, and Asian images $[-0.43, -0.27]$. In contrast, Asians rate Asian males as higher in social status than Asian females $[0.02, 0.17]$ (with no significant male-female difference for Caucasian images $[-0.19, 0.03]$).

Powerful: Both Asian and Caucasian raters rate males of the *other* ethnicity as more powerful than males of their own ethnicity (i.e., Asians rate Caucasian males as more powerful than Asian males $[0.04, 0.26]$; Caucasians rate Asian males as more powerful than Caucasian males $[0.14, 0.34]$).

Successful: Asian raters give the lowest ratings to Asian male images (lower than images of Asian females $[-0.728, -0.5792]$, Caucasian males $[-0.6852, -0.4755]$,

and Caucasian females $[-0.9, -0.75]$). No such effect appears for Caucasian raters.

Responsible: Both Asians and Caucasians rate male images of their own ethnicity to be the least responsible. Specifically, Caucasians rate images of Caucasian males as less responsible than images of Caucasian females $[-1.11, -0.88]$, Asian males $[-0.48, -0.26]$, and Asian females $[-0.35, -0.12]$, while Asians rate Asian males as less responsible than Asian females $[-0.61, -0.45]$, Caucasian females $[-0.76, -0.59]$, and Caucasian males $[-0.36, -0.13]$.

Inter-group Correlation Analysis

How consistently do Caucasians and Asians rate various traits? What traits do they agree on? Are there differences in their agreement levels regarding Asian faces versus Caucasian faces? To address these questions, we separated our images into two groups by ethnicity. Since we used the back-translation process for translating traits from English to Chinese, we are confident that the differences here are due to culture disagreement. For each image group, we computed the average ratings by Asians and Caucasians for all traits, and then calculated their Spearman correlation. The results are shown in Fig 6. Here, the dots represent the Spearman correlation between Caucasian and Asian participants. All correlations are statistically significant. We can see that for traits like responsible and successful, there is a large disagreement between Caucasian and Asian raters, especially on Asian face images. For the attractive trait, the two rater groups agree more on Asian faces than on Caucasian ones by a relatively large margin.

To qualitatively examine the differences in the ratings on responsible, successful, and attractive, we selected facial images that are rated most differently by Caucasian and Asian raters, i.e., outliers in the plots in Figure 7. For each trait, the images on the left are rated higher by Caucasian participants and the images on the right are rated higher by Asian participants. The center panel shows the z-scored average rating by Asian (x) and Caucasian (y) participants for every face image. The red dots represent the selected outliers.

We see that for the responsible and successful outliers, Asians give much lower ratings to middle-aged Asian males compared to Caucasian raters. We suspect that Asian participants tend to have negative stereotype of government officials, who are usually middle-aged males. This stereotype makes Asian participants believe that they are irresponsible and they also gave them low ratings on successful. This stereotype can stem from news regarding the anti-corruption campaign, in which the photos of corrupt bureaucrats - usually middle-aged males - are usually shown to the public.

Another surprising finding is that Asian participants give higher ratings for responsible and successful to young people and even children. Since our Asian participants' average age range is 20-30, younger than Caucasian groups (average age range 30-40), it might be related to the phenomena reported by a large number of studies that people tend to like
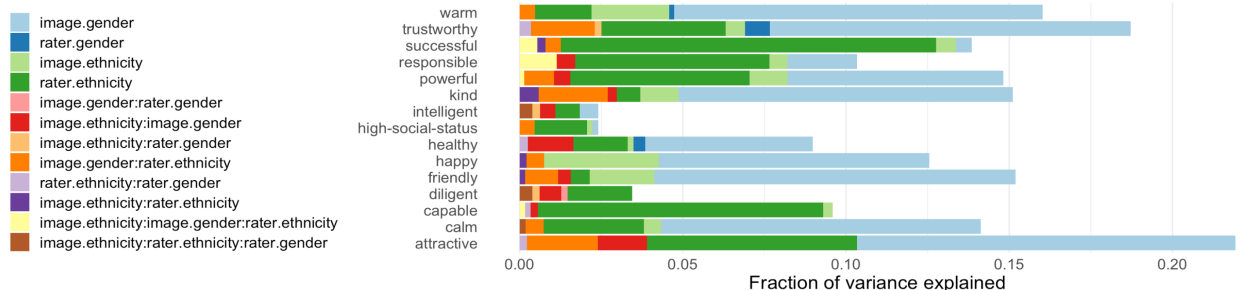
Figure 5: ANOVA analysis. For each trait we assessed what fraction of the overall variance in ratings can be explained by attributes of the rater (ethnicity, gender) and attributes of the image (ethnicity, gender), and their interactions. Here we summarize these ANOVAs as the overall variance each term explains. For most features, the dominant explanatory factors are image gender (light blue; reflecting that females are rated as more attractive, warm, and friendly), and rater ethnicity (dark green, reflecting that Asians tend to give less positive ratings overall).
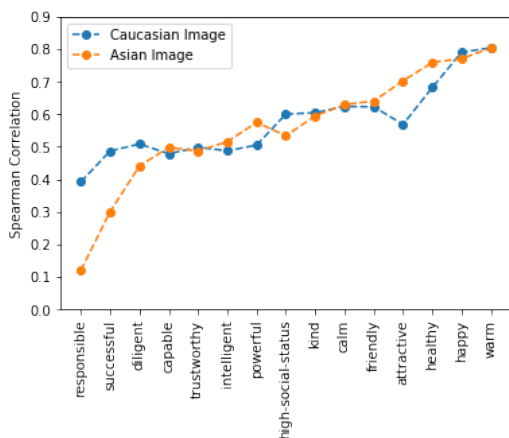

Figure 6: Spearman correlation between Caucasian and Asian raters on the 15 traits on Asian and Caucasian faces.

people who are similar to them (Montoya, Horton, & Kirchner, 2008). However, in an analysis breaking this out by rater age, this was not the case. Asians across age groups consider younger people to be slightly more responsible and successful, while Caucasians strongly rate older people as higher on these attributes.

We also examined the attractive trait. In Figure 7, we find it surprising that Caucasian participants give high ratings usually to young females, whereas Asian participants give higher ratings to senior people. While this could be due to a cultural differences in understanding the term "attractive," as in Chinese, the word also means how good and kindly a person looks. However, we ran a second experiment using just a word that meant "good-looking" and the phenomenon still held, so this is a puzzle for future work to investigate.

Discussion

We compared how Chinese and Asian viewers estimate 15 social traits from each of thousands of real-world face images varying in ethnicity, gender, and age of the person pictured.

These data revealed a number of similarities and differences in facial impression formation across these two cultures.

First, although Caucasian and Asian raters were similarly self-consistent, they differed in their group agreement levels, with Caucasians having markedly larger across-rater consistency scores. This suggests that Caucasian participants tend to judge most traits similarly, whereas for Asian participants there are diverse opinions on most of the traits in facial images.

Second, we found that Asian raters give lower ratings on average to almost every positive social trait compared to Caucasian raters. We suspect it is due to the fact that Chinese Asian participants tend to emphasize the negative more (Sims et al., 2015).

Third, we find that the ethnicity of raters and ethnicity & gender of the face images strongly influences ratings. Asian faces on average receive lower ratings in warmth-related traits such as happy, trustworthy and warm, because in our image set, Asian faces are less likely to be smiling than Caucasian images.

Last, Asian and Caucasian raters tend to disagree on traits like responsible and successful, in particular on Asian images. Upon further investigation of faces with extremely disparate ratings, this effect appears to reflect Asian and Caucasian impressions of middle aged Asian males: Caucasians tend to see them as quite responsible and successful, while Asian raters do not. Given that both Caucasian and Asian participants have high self-consistency, this suggests very different attitudes between Caucasian and Asian participants concerning which people are responsible or successful.

Our dataset and analyses provide new perspectives for cross-cultural studies of facial impressions. They highlight interesting observations on how Caucasian and Asian participants view certain facial impression traits differently. They also open the door to further studies such as building computational models to predict the ratings of faces by Caucasian and Asian raters.
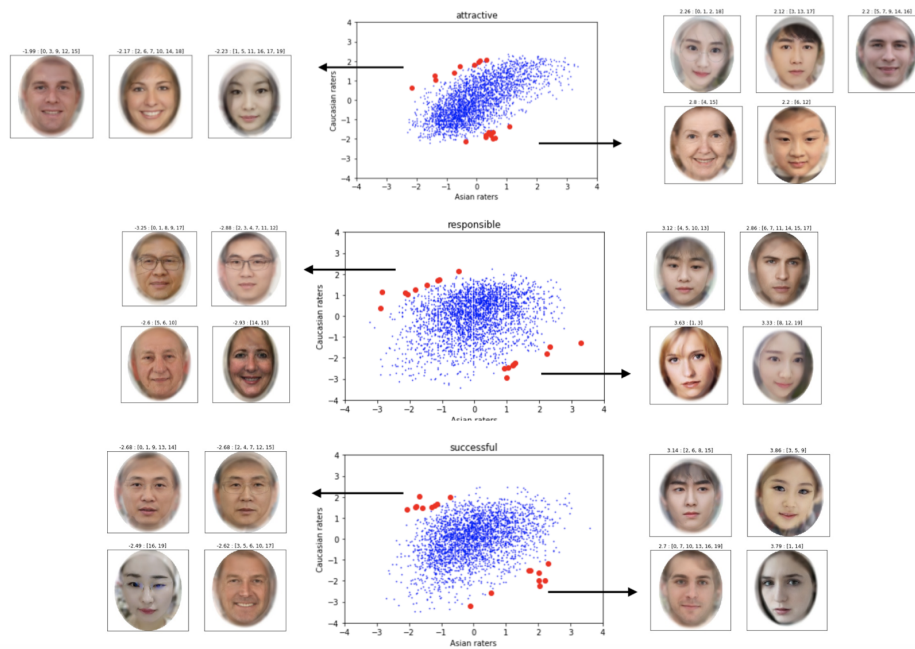
Figure 7: Images that are rated most differently by Caucasians and Asians in responsible, successful and attractive (from top to bottom). Images on the left side are rated lower by Asians than Caucasians, whereas images on the right side are rated higher by Asians than Caucasians. We morphed images in order to preserve privacy while still showing the facial features that are rated most differently.

Acknowledgments

References

Adams Jr, R. B., Hess, U., & Kleck, R. E. (2015). The intersection of gender-related facial appearance and facial displays of emotion. *Emotion Review*, *7*(1), 5–13.

Bainbridge, W. A., Isola, P., & Oliva, A. (2013). The intrinsic memorability of face photographs. *Journal of Experimental Psychology: General*, *142*(4), 1323-1334.

Blais, C., Jack, R. E., Scheepers, C., Fiset, D., & Caldara, R. (2008). Culture shapes how we look at faces. *PloS one*, *3*(8).

Cunningham, M. R., Roberts, A. R., Barbee, A. P., Druen, P. B., & Wu, C.-H. (1995). Their ideas of beauty are, on the whole, the same as ours: Consistency and variability in the cross-cultural perception of female physical attractiveness. *Journal of Personality and Social Psychology*, *68*(2), 261.

Ebner, N. C. (2008). Age of face matters: Age-group differences in ratings of young and old faces. *Behavior research methods*, *40*(1), 130–136.

Falvello, V., Vinson, M., Ferrari, C., & Todorov, A. (2015). The robustness of learning about the trustworthiness of other people. *Social Cognition*, *33*(5), 368–386.

Hehman, E., Sutherland, C. A., Flake, J. K., & Slepian, M. L. (2017). The unique contributions of perceiver and target characteristics in person perception. *Journal of Personality and Social Psychology*, *113*(4), 513–539.

Hofstede, G. (1980). Culture and organizations. *International Studies of Management & Organization*, *10*(4), 15–41.

Krys, K., Hansen, K., Xing, C., Szarota, P., & Yang, M.-m. (2014). Do only fools smile at strangers? Cultural differences in social perception of intelligence of smiling individuals. *Journal of Cross-Cultural Psychology*, *45*(2), 314–321.

Litman, L., Robinson, J., & Abberbock, T. (2016). Turkprime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, *49*, 433–442. doi: 10.3758/s13428-016-0727-z

Montoya, R. M., Horton, R. S., & Kirchner, J. (2008). Is actual similarity necessary for attraction? A meta-analysis of actual and perceived similarity. *Journal of Social and Personal Relationships*, *25*(6), 889–922.

Nisbett, R. E., & Miyamoto, Y. (2005). The influence of culture: Holistic versus analytic perception. *Trends in Cognitive Sciences*, *9*(10), 467–473.

Olivola, C. Y., & Todorov, A. (2010). Fooled by first impressions? Reexamining the diagnostic value of appearance-based inferences. *Journal of Experimental Social Psychology*, *46*(2), 315–324.

Oyserman, D., Coon, H. M., & Kemmelmeier, M. (2002). Rethinking individualism and collectivism: Evaluation of theoretical assumptions and meta-analyses. *Psychological Bulletin*, *128*(1), 3–72.

Rezlescu, C., Duchaine, B., Olivola, C. Y., & Chater, N. (2012). Unfakeable facial configurations affect strategic choices in trust games with or without information about past behavior. *PloS One*, *7*(3), e34293.

Sims, T., Tsai, J. L., Jiang, D., Wang, Y., Fung, H. H., & Zhang, X. (2015). Wanting to maximize the positive and minimize the negative: Implications for mixed affective experience in American and Chinese contexts. *Journal of Personality and Social Psychology*, *109*(2), 292–315.

Sutherland, C. A., Liu, X., Zhang, L., Chu, Y., Oldmeadow, J. A., & Young, A. W. (2018). Facial first impressions across culture: Data-driven modeling of Chinese and British perceivers' unconstrained facial impressions. *Personality and Social Psychology Bulletin*, *44*(4), 521–537.

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science*, *308*(5728), 1623–1626.

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology*, *66*, 519–545.

Wilson, J. P., & Rule, N. O. (2015). Facial trustworthiness predicts extreme criminal-sentencing outcomes. *Psychological Science*, *26*(8), 1325–1331.

Zebrowitz, L. A., Kikuchi, M., & Fellous, J.-M. (2010). Facial resemblance to emotions: group differences, impression effects, and race stereotypes. *Journal of Personality and Social Psychology*, *98*(2), 175–189.

Zwillinger, D., & Kokoska, S. (2000). *Standard probability and statistics tables and formulae*. Chapman & Hall/CRC.