

## Supplemental Material 4

### S4. Additional studies that did not include the intergroup manipulation

#### **Additional Study 1**

##### **Method**

**Participants and design.** We recruited 279 undergraduate students with a mean age of 21 (range 18 – 70) for this study. Of the total number of participants 143 were Han Chinese students from the University of Macau (China) and 136 were Dutch students from the University of Amsterdam (The Netherlands). This study had a 2 (BS, Group Expresser: Chinese vs. European) x2 (BS, Group Perceiver: Chinese vs. European) design.

**Stimuli and procedure.** For Chinese models, we made use of the Taiwanese Facial Expression Image Database (TFEID, Yen & Chen, 2007), European models were based on the ADFES (van der Schalk, Hawk, Fischer, & Doosje, 2011). In order to match the TFEID stimuli with the ADFES stimuli we presented seven emotions per group (contempt, surprise, fear, anger, sadness, happiness, and disgust) including male and female models. We used an identical procedure to the one described in the meta-analysis, but we left out the ranking task (sorting a number of ethnic groups that are represented in the Netherlands/the United States/the United Kingdom according to size of their population). Considering that the vast majority of Chinese citizens is Han Chinese we assumed that a ranking task would have been ineffective in the Chinese context. Since it would not have been methodologically correct to administer this task only to British participants, we decided to leave it out altogether. The ethical committee of the Faculty of Social and Behavioural Sciences of the University of Amsterdam approved this study as meeting the requirements of standard survey research (2016-SP-6359).

## Results

We conducted an analysis of variance with Group Membership Expresser (Chinese vs. Dutch) and Group Membership Perceiver (Chinese vs. Dutch) as factors and intensity of emotion (aggregated average scores for all target emotions) as outcome variable. This analysis revealed that neither Group Membership Expresser ( $F(1, 275) = 3.18, p = .076, \eta_p^2 = .011$ ) nor Group Membership Perceiver ( $F(1, 275) = 2.79, p = .096, \eta_p^2 = .010$ ) had an effect on the perceived intensity of emotions. The interaction between Group Membership Expresser and Group Membership Perceiver was not significant either, ( $F(1, 275) = 0.037, p = .847, \eta_p^2 = .000$ ).

## Additional Study 2

### Method

**Participants and design.** Participants were 267 Moroccans (mean age = 28, range: 18 – 59, 130 males, 137 females) who were recruited via Qualtrics panel ([www.qualtrics.com](http://www.qualtrics.com)). The study had a 2 (BS, Group Expresser: Arab vs. European) x 2 (BS, Gender Expresser: female vs. male) design and made use of ADFES Stimuli (van der Schalk, Hawk, Fischer, & Doosje, 2011). Each participant ran a total of nine trials (contempt, embarrassment, pride, surprise, fear, anger, sadness, happiness, and disgust) with female participants judging female stimuli and male participants judging male stimuli. We used the same procedure as described in the meta-analysis but again decided to leave out the intergroup manipulation in order to keep the follow-up studies more comparable. The ethical committee of the Faculty of Social and Behavioural Sciences of the University of Amsterdam approved this study as meeting the requirements of standard survey research by (2016-SP-6558).

## Results

An analysis of variance with Group membership as factor and the perceived intensity of the target emotion (aggregated average score of all target emotions) as dependent measure, revealed that there was a main effect of Group membership ( $F(1, 265) = 6.36, p = .012, \eta_p^2 = .023$ ). Moroccan perceivers attributed more intense emotions to Europeans than to Arabs ( $M_{European} = 67.07, SE = 1.37$  vs.  $M_{Arabs} = 62.23, SE = 1.35$ ).

### Additional Study 3

#### Method

**Participants and Design.** The sample consisted of 279 participants, of which 139 were Moroccan males and 140 were White British males. Since the RaFD only offers male models of both ethnic groups, we decided to test male participants only. Their age ranged between 18 and 84 ( $M_{age} = 41$ ), after exclusion of one participant who indicated being 100 or older. This study employed a 2 (BS, Group Expresser: Arab vs. European) x 2 (BS, Group Perceiver: Arab vs. European) design.

#### Stimuli and procedure.

We chose the RaFD (Langner et al., 2010) because this is the only set, known to us, that includes both Arab and European models, and a variety of emotions. We presented all 8 displays that the RaFD provides (contempt, surprise, fear, anger, sadness, happiness, disgust, neutral) resulting in 8 trials per participant. The stimuli were stills and the faces were shown face forward. We used 8 different male models per emotion and ethnic group. Participants were recruited via Qualtrics panels ([www.qualtrics.com](http://www.qualtrics.com)). We made use of the same procedure as described in additional studies 1 and 2. The ethical committee of the Faculty of Social and Behavioural

Sciences of the University of Amsterdam approved this study as meeting all requirements of standard survey research (2016-SP-6556).

## Results

We conducted an analysis of variance with Group Membership Expresser (Arab vs. European models) and Group Membership Perceiver (Moroccan vs. British) as predictors and emotional intensity (aggregated average scores for all target emotion) as outcome variable. Neither Group Membership Perceiver ( $F(1, 275) = 0.060, p = .806, \eta p^2 = .0001$ ) nor Group Membership Expresser ( $F(1, 275) = 0.370, p = .843, \eta p^2 = .001$ ) was significant and thus did not affect perceived intensity of facial expressions. However, there was a significant interaction between Group Membership Expresser and Group Membership Perceiver ( $F(1, 275) = 6.20, p = .013, \eta p^2 = .022$ ). Pairwise comparison showed that British perceivers rate emotions by European models as more intense than Moroccan perceivers ( $F(1, 275) = 4.81, p = .029, \eta p^2 = .017, M_{British\ perceiver} = 68.64, SE = 1.62$  vs.  $M_{Moroccan\ perceiver} = 63.32, SE = 1.81$ ). However, there was no difference between British and Moroccan perceivers rating Arab models ( $F(1, 275) = 1.76, p = .185, \eta p^2 = .006, M_{Moroccan\ perceiver} = 63.94, SE = 1.80$  vs.  $M_{British\ perceiver} = 67.17, SE = 1.64$ ). Finally, British perceivers attributed marginally more intense emotions to European models than Moroccan perceivers to Arab models ( $F(1, 275) = 3.78, p = .053, \eta p^2 = .014, M_{British\ perceiver} = 68.64, SE = 1.62$  vs.  $M_{Moroccan\ perceiver} = 63.94, SE = 1.80$ ).