**Methods**

Our sample size, hypotheses, and statistical analyses were pre-registered at the Open Science Framework (OSF): <https://osf.io/8pd2s/>.

***Participants & Design***

A sample size calculation, using GLIMMPSE ([glimmpse.samplesizeshop.org)](http://glimmpse.samplesizeshop.org/#/)), revealed that collecting data of 20 participants in each study was sufficient to detect the hypothesized interaction effect with a 80% chance at an alpha level of .05, assuming a .70 within-subjects correlation across measures. This calculation was based on the effect size reported by Craig & Lipp (2018; Study 2), as this study reported a lower effect size (partial eta: .18) than Bijlstra et al. (2010). In total, 54 Radboud University students either completed an evaluative (Study 1: happy versus sad) or a stereotype (Study 2: sad versus angry) emotion categorization task (Study 1: *N* = 29, 76% female, *M*age = 23.64, *SD*age = 3.61, range 18 – 34, one participant did not provide age-data; Study 2: *N* = 25, 64% female, *M*age = 24.84, *SD*age = 5.47, range 20 – 47). Within these studies participants categorized faces of young and old men based on their emotional expressions. Our goal was to test at least 50 participants, 25 for each study. Moreover, because the study was part of a chunk, and the experiment with the highest sample size was 60, we aimed for 60 participants in total (see Pre-registration). Assignment to each emotion categorization task was counterbalanced. Participants were compensated with €5 or course credit.

***Materials & Procedure***

The materials and procedures were kept similar to Craig and Lipp (2018). That is, the same sixteen models, eight young men (model numbers: 008, 013, 016, 031, 037, 049, 057, 072) and eight old men (model numbers: 004, 015, 033, 042, 053, 059, 065, 076), were selected from the FACES database (Ebner, 2010). For these models, we selected pictures depicting anger, sadness, and happiness, resulting in a total of 48 pictures: 3 emotional expressions x 2 age-groups x 8 targets per age-group. Per study this led to a total of 32 pictures.

Participants in the present studies were seated in individual cubicles. It was their task to categorize two emotional expressions as quickly and accurately as possible. Participants’ categorization task depended on their assigned study (i.e., either ‘happy versus sad’ or ‘anger versus sad’). Trials consisted of a fixation cross, presented for 1000 milliseconds, followed by a picture of an emotional expression. This picture was visible until a response was made or for the maximum of 3000 milliseconds. Unlike Craig and Lipp (2018), we included feedback messages after the presentation of the picture to indicate too late (“TOO LATE”) or incorrect (“WRONG”) responses. For correct responses, no feedback was included. In the task, participants were asked to categorize pictures of emotional faces, by pressing one of two keys (left or right ‘SHIFT’ key). Response mapping was counterbalanced between participants. Within the experiment participants were presented with 128 trials, in which each picture was presented four times. To achieve this, the 32 pictures per study were presented in random order without replacement and this procedure was repeated four times.

***Data reduction & Analysis***

We excluded responses that were incorrect (9.8%) or too late (0.1%), and responses that were correct but faster than 100 milliseconds (0%, 1 trial), or faster/slower than three *SD* from participants’ mean response latency (1.5%). For each study we submitted both the average response latencies and the proportion of correct trials to a 2 (Age-group: young versus old) X 2 (Emotion: ‘happy versus sad’ or ‘angry versus sad’) repeated measures ANOVA. Additionally, the categorization data regarding the sad facial stimuli were subjected to a 2 (Age-group: young versus old) X 2 (Study: ‘happy versus sad’ or ‘angry versus sad’) mixed design with repeated measures on the Age-group variable and Study as between participants variable.

**Results**

Tables 1 and 2 present an overview of the averages and standard deviations for each dependent variables of the current study. Anonymized data are stored at the OSF: <https://osf.io/8pd2s/>.

***Study 1: Happy versus Sad***

*Response latencies.* In line with Craig and Lipp (2018), the predicted two-way interaction between Age-group and Emotion was observed, *F*(1, 28) = 16.04, *p* < .001, η2 = .36, see Table 1. Happiness was categorized faster than sadness on young male faces, *F*(1, 28) = 16.80, *p* < .001, η2 = .38, but not on old male faces (*F*<1, *p* ns.). Additionally, happiness was categorized faster on young male faces than old male faces, *F*(1, 28) = 26,79, *p* < .001, η2 = .489, whereas there was no difference in categorizing sadness on young or old male faces. Furthermore, emotional expressions in general were categorized faster on young than old male faces, *F*(1, 28) = 15.04, *p* < .001, η2 = .35, and happy expressions were categorized faster than sad expressions, *F*(1, 28) = 4.99, *p* = .034, η2 = .15.

*Proportion correct.* Inconsistent with Craig and Lipp (2018), but consistent with the evaluative congruency hypothesis, we found a two-way interaction between Age-group and Emotion, *F*(1, 28) = 14.83, *p* < .001, η2 = .35. Happiness was categorized more accurately than sadness on young male faces, *F*(1, 28) = 12.79, *p* < .001, η2 = .31, whereas the reverse was true for old male faces, *F*(1, 28) = 8.02, *p* = .008, η2 = .22. Moreover, happiness was categorized more accurately on young than old male faces, *F*(1, 28) = 30.31, *p* < .001, η2 = .52. The data did not reveal a difference in categorizing sadness on young versus old male faces. Furthermore, in line with Craig and Lipp (2018), emotional expressions in general were more accurately categorized on young than old male faces, *F*(1, 28) = 23.83, *p* < .001, η2 = .46. No main effect of Emotion was observed (*F* = 1.052, *p* = .31).

***Study 2: Sad versus Angry***

*Response latencies.* In contrast to our predictions, we did not observe a two-way interaction between Age-group and Emotion (*F* < 1, *p* ns). However, we did find that emotional expressions were categorized faster on young than old male faces, *F*(1, 24) = 38.54, *p* < .001, η2 = .62, and that angry expressions were categorized faster than sad expressions, *F*(1, 24) = 9.13, *p* = .006, η2 = .28.

*Proportion correct.* Consistent with Study 1, we observed that participants categorized emotional expressions on young male faces than old male faces more accurately, *F*(1, 24) = 87.02, *p* < .001, η2 = .78. No main effect of Emotion or interaction between Age-group and Emotion was observed (*F* < 1, *p* ns).

***Analysis of sad facial stimuli between Studies 1 & 2***

*Response latencies.* Participants were in general faster in categorizing sadness on young male faces than old male faces, *F*(1, 52) = 18.92, *p* < .001, η2 = .22. However, this main effect was qualified by the two-way interaction between Age-group and Study, *F*(1, 52) = 14.51, *p* < .001, η2 = .17. In Study 2, sad faces were categorized faster on young male faces than old male faces, *F*(1, 52) = 30.99, *p* < .001, η2 = .37, but not in Study 1 (*F* < 1, *p* ns). Moreover, the old sad faces were faster recognized when accompanied by happy faces than angry faces, *F*(1, 52) = 30.99, *p* < .001, η2 = .37, for young sad faces no difference between studies was observed (*F* = 2.51, *p* = .12). Overall, the participants were faster in categorizing sad faces when accompanied by happy faces than angry faces, *F*(1, 52) = 25.10, *p* < .001, η2 = .33.

*Proportion correct.* Overall, participants were more accurate in categorizing sadness on young male faces than old male faces, *F*(1, 52) = 8.60, *p* = .005, η2 = .11. This main effect was qualified by the two-way interaction between Age-group and Study, *F*(1, 52) = 14.69, *p* < .001, η2 = .20. In Study 2, we observed a difference in categorization of sad faces between the Age-groups, *F*(1, 52) = 21.31, *p* < .001, η2 = .29, but not in Study 1 (*F* < 1, *p* ns). Moreover, participants were in general less accurate in categorizing sad faces when accompanied by angry faces than happy faces, *F*(1, 52) = 20.67, *p* < .001, η2 = .28.

**Tables**

*Table 1.* Average response latencies, proportion correct and their standard deviations for Study 1.

|  | Speed | | Accuracy | |
| --- | --- | --- | --- | --- |
| **Age-Group** | Happy | Sad | Happy | Sad |
| Young | 499 (76) | 533 (88) | 0.97 (0.04) | 0.93 (0.07) |
| Old | 537 (94) | 536 (101) | 0.89 (0.10) | 0.94 (0.05) |

*Table 2.* Average response latencies, proportion correct and their standard deviations for Study 2.

|  | Speed | | Accuracy | |
| --- | --- | --- | --- | --- |
| **Age-Group** | Angry | Sad | Angry | Sad |
| Young | 627 (98) | 658 (125) | 0.92 (0.07) | 0.90 (0.07) |
| Old | 672 (145) | 705 (125) | 0.81 (0.08) | 0.82 (0.12) |