Replication of Corrigendum: The Crowd-Emotion-Amplification Effect

Liu Jiachen[1], Yang Qingren[1], & Zhang Yilin[1]

[1] Nanjing Normal University

Author Note

Each member of our group actively participated in every step of the replication process, but there was still a focus on division of labor. We had the following responsibilities: Liu Jiachen was primarily responsible for code verification, creating plots, and proofreading the manuscript; Yang Qingren primarily handled the Chinese manuscript and statistical verification; Zhang Yilin was primarily responsible for the English manuscript, presentation, and PowerPoint production.

Abstract

We replicated the first study of the article “The Crowd-Emotion-Amplification Effect” written by Amit Goldenberg, published in the journal “Psychological Science” in March 2021. This article investigates how people judge the emotions of a group of individuals by rapidly scanning their facial expressions. The researchers proposed that when people observe a group, they tend to focus on faces displaying strong emotions, leading to a phenomenon known as the crowd-emotion amplification effect, where the estimated average emotional response of the group is more extreme than the actual average.In the first study (N = 50), the researchers documented the crowd-emotion amplification effect. In the second study (N = 50), even with increased exposure time, the effect was successfully replicated. In the third study (N = 50), the researchers used eye-tracking technology to demonstrate that attention bias towards emotional faces drives the amplification effect. Through replicating this study, we gained a better understanding of the crowd-emotion amplification effect.

Keywords: emotions, social cognition, perception, intergroup dynamics, open data, open materials, preregistered

Replication of Corrigendum: The Crowd-Emotion-Amplification Effect

Introduction

Imaging yourself making a speech in front a group of people, you quickly scan the audience, how can you make split-second judgements about the emotions of the crowd. This article proposes that perceivers preferentially attend to faces exhibiting strong emotions, as highly emotional faces are more salient than neutral faces (Pessoa et al., 2002), which in turn generates a crowd-emotional-amplification effect—estimating a crowd's average emotional response as more extreme than it actually is. One explanation is that people preferentially attend to faces conveying strong emotions over faces with neutral expressions (Eimer & Holmes, 2007; Pessoa et al., 2002). Thus, when a person infers a crowd's emotion, attentional bias toward more emotional faces should contribute to an amplification in the average emotion estimate. When generating rapid evaluation of others' emotions, the capacity for visual representation is finite (Alvarez, 2011; Whitney & Yamanshi Leib, 2018), ensemble coding compensates for this limitation by allowing perceivers to form compressed, summary representations of visual information (Whitney et al., 2014). While how these summaries are computed is still under debate. Whether people encode all items in a set or only a subset, it is plausible they preferentially attend to the most salient items in a set (Kanaya et al., 2018; Sweeny et al., 2013). Kanaya and colleagues (2018) explored this hypothesis, their experiments provided evidence of amplification in the estimation of multiple objects, however, it is thus unclear whether a similar effect of amplification would be evident with faces. By replicating this study, we can further deepen our understanding of the crowd-emotion amplification effect. In the first study (N = 50), researchers documented the presence of the crowd-emotion amplification effect. In the second study (N = 50), even with increased exposure time, the effect was replicated. In the third study (N = 50), researchers employed eye-tracking technology to demonstrate that attentional bias toward emotional faces drives the generation of the emotion amplification effect. The results of these replicated studies enable us to gain a more comprehensive understanding of the characteristics and underlying mechanisms of the crowd-emotion amplification effect. The significance of conducting replicated studies lies in several aspects. Firstly, replication allows us to validate and confirm the findings of the original study, ensuring the reliability and robustness of the observed effects. It helps establish the generalizability of the phenomenon under investigation and strengthens the overall body of knowledge in the field.

Secondly, replication provides an opportunity to examine the consistency and stability of the effects over different samples, settings, or conditions. By replicating the study with different participants or manipulating certain variables, we can explore the boundary conditions of the phenomenon and gain a deeper understanding of its underlying mechanisms.

Furthermore, replication allows for the identification of potential limitations or factors that may influence the observed effects. If the results of the replication study align with those of the original study, it enhances the confidence in the validity of the findings. On the other hand, if the results differ, it prompts further investigation into the factors contributing to the discrepancy and encourages critical evaluation and refinement of the original research.

Ultimately, the significance of replication lies in its contribution to the cumulative nature of scientific knowledge. It fosters transparency, rigor, and intellectual progress by encouraging researchers to build upon existing findings, challenge assumptions, and refine theories. Through replication, we can gain a more comprehensive and reliable understanding of the phenomenon under study, and its implications can extend to various fields, applications, and practical contexts. Through an in-depth exploration of the crowd-emotion amplification effect, we can develop a more accurate understanding of the dynamic changes in emotions within a group and apply it in practical contexts. The significance of this research lies in providing valuable insights into how individuals make judgments about emotions when observing a crowd, while also reminding us of important considerations in social interaction and emotional cognition.

Research hypothesis

Based on the above research background, the study proposed three hypotheses. Hypothesis 1 : Participants would show a crowd-emotion-amplification effect, estimating a crowd's average emotion as more intense than it actually is. Hypothesis 2 : The amplification effect would be stronger as the number of faces within an array increased (following Kanaya et al., 2018). Hypothesis 3 : Amplification would be more robust for negative emotions than for positive emotions.

Methods

Participants

After conducting a power analysis that suggested that a sample size of 50 participants completing 150 trials would provide power of 97% to support the first hypothesis, they recruited 50 participants (23 men, 27 women; age: M = 19.52 years, SD = 1.69), each of whom completed 150 trials.

Procedure

In each trial, participants first saw an array containing 1 to 12 faces (Fig. 1). These faces expressed different intensities of emotion from either neutral-to-angry or neutral-to-happy continua (Fig. 2). The group-average intensity, the valence of the faces, and the set size were all chosen randomly on each trial. They did not mix the happy and angry faces in the same set for two reasons. First, doing so could undermine their ability to detect an amplification effect. Second, the most-happy and most-angry faces were not equal in intensity, thus making the average between the two different from zero. They concatenated 50 modifications of a face from neutral to extremely angry and from neutral to extremely happy, effectively creating a visual rendering of 50-points anger or happiness scales. (Fig. 2). On every trial, an array containing between 1 and 12 faces was presented to participants. Faces could appear in any of 12 fixed locations on the screen. The group's mean emotional intensity was randomly set to be between 10 and 40 (on the basis of a 50-point scale of angry or happy faces, 1 being neutral and 50 being very angry or very happy). They limited the range of the group means in order to allow distributions of intensity that were as close to uniform as possible within the 50-point scale. The sets were engineered so that the standard deviation of a 12-face array was always 10, from which they randomly chose a subset of 1 to 12 faces. To prevent residual visual processing, they immediately followed each face in the array with a mask at the same location. After viewing the face array and the mask, participants were asked to evaluate the average emotion expressed in the array. They presented a single face bearing a neutral expression on the screen after the array was masked, so participants could make their evaluations. After completing the main task, participants filled out a short survey that included an abbreviated version of the Social Interaction Anxiety Scale, the Need to Belong Scale, and demographic questions including age, gender, race, and education level. Figure 2 A sample of three faces from the neutral-to-angry scale (top) and from the neutral-to-happy scale (bottom) that were used in the studies. Results To measure amplification in estimation of the face sets, they conducted a mix-model analysis of repeated measures, comparing the actual mean emotion expressed in each set with participant estimated mean emotion. The results show that the estimated mean crowd emotion was 1.87 points higher (scale from 1 to 50) than the actual mean crowd emotion, b = 2.87, 95% confidence interval (CI) = [2.53, 3.21], SE = 0.17, t(14533) = 16.8, $p < .001$, R2 = .05, which support the crowd-emotion-amplification hypothesis (Hypothesis 1). They used a single mode to test whether an increase in the number of faces (Hypothesis 2) or the type of emotions expressed by the faces (Hypothesis 3) influenced the crowd-emotion-amplification effect. Results were in line with the second and the third hypothesis. Number of faces significantly predicted an increase in amplification, b = 0.42, 95% confidence interval (CI) = [0.21, 0.64], SE = 0.03, t(7233) = 3.86, $p < .001$, R2 = .08 (Fig 3). For hypothesis 3, amplification was stronger for crowds expressing negative emotions than crowds expressing positive emotions, b = 0.24, 95% confidence interval (CI) = [0.02, 0.45], SE = 0.11, t(7253) = 2.18, p = .02, R2 = .08. The interaction between number of faces and the valence expressed by the faces was not significant, b = 0.08, 95% confidence interval (CI) = [-0.13, 0.30], SE = 0.11, t(7226) = 0.76, p = .44, R2 = .08.

Conclusion

In sum, the study 1 support crowd-emotional-amplification effect, participants estimated that face sets were more emotional than they actually were, amplification increased with set size, and amplification was stronger for negative compared with positive emotions.

Study Replication Ideas and Process

Our approach to replicating the study involved several steps. First, we followed the open practices checklist provided in the article to locate the necessary raw data and code. We carefully examined the data to verify if any preprocessing had been applied and checked the compatibility of the code with our platform. Second, we conducted a thorough review of the code, cross-referencing it with the article to understand the tasks it performed, verifying the steps involved, and comprehending its logical flow. Third, we attempted to run the original code, identifying any issues that arose, making necessary corrections, and continuing the execution. Fourth, we explored alternative statistical approaches to test whether different results would be obtained. Finally, in sections where visualizations were not provided, we supplemented the analysis with appropriate graphics to enhance the understanding and visual representation of the data obtained.

Republication Results

In terms of data, our obtained results are basically consistent with those in the text, but in the results of the analysis of hypothesis 2 and hypothesis 3, there are some discrepancies with the data in the text. Based on the principle of scientific notation, the b for hypothesis 2 ought to be 0.43, but not 0.42, and the p for hypothesis 3 ought to be 0.03 but not 0.02. What's more, the SE for hypothesis 2 should be 0.11, but the article puts 0.03 (Fig. 4). Also, the same problem exists with the reporting of R2 in the results of assumptions 2 and 3, which should be 0.09 instead of 0.08 (Fig. 5). However, during the process of replication, we encountered several issues. Firstly, there were problems with data cleaning. Even though the data had already undergone preprocessing, there were still some issues that affected the overall functioning of the code. The most common issue was related to data type conversion. The original code often required specific data types, such as factor variables or other types, for performing statistical operations. However, the data had default types, which prevented the statistical analyses from being conducted. Therefore, we redefined the data preprocessing steps. We carefully selected appropriate data types for different scenarios to ensure the accuracy of the results and the smooth execution of the code. Secondly, there were errors in variable names. The column names of the encoded data were incorrect, and the variable names used in the code were also erroneous. We made the necessary corrections to address these issues. Thirdly, the existing data visualizations in the original article were not clear, and the plotting code was incomplete. We filled in the missing code and enhanced the existing images to improve their readability. Moreover, we supplemented the analysis with additional plots using the "gplot" library to provide more intuitive insights.

Discussion and Conclusion

According to the article on open practices, we obtained the data and analysis code for the first study in the article. By understanding the content of the code, we ran the provided code in our local R environment and made necessary corrections, ultimately obtaining results similar to those in the article.

We conducted tests for three hypotheses. To measure the amplification effect in facial ensemble estimation, we performed a repeated-measures mixed model analysis comparing the difference between the actual average emotion expressed in each ensemble and the participants' estimated average emotion. As each participant encountered four facial identities, we included two random intercepts: facial identity and participant. The results supporting the hypothesis of group emotion amplification showed that the estimated average group emotion was 2.87 points higher (on a scale of 1 to 50) than the actual average group emotion, with b = 2.87, 95% confidence interval (CI) = [2.53, 3.21], SE = 0.17, $t(14533) = 16.8$, $p < .001$, R2 = .05.1.

We used a model to test whether increasing the number of faces (hypothesis 2) or the emotional expression type of faces (hypothesis 3) influenced the group emotion amplification effect. This not only reduced the number of comparisons but also allowed testing for interaction effects between the two variables. For our dependent variable, we created difference scores between participants' estimates of average group emotion and the actual average group emotion; positive scores indicated an amplification effect. We then performed a repeated-measures mixed model analysis where facial quantity, facial emotional value, and their interaction predicted the degree of difference between estimated and actual average emotions. Similar to the previous analysis, we included random intercepts for facial identity and participant.

The results were consistent with our second hypothesis: facial quantity significantly predicted an increase in the amplification effect, with b = 0.42, 95% CI = [0.21, 0.64], SE

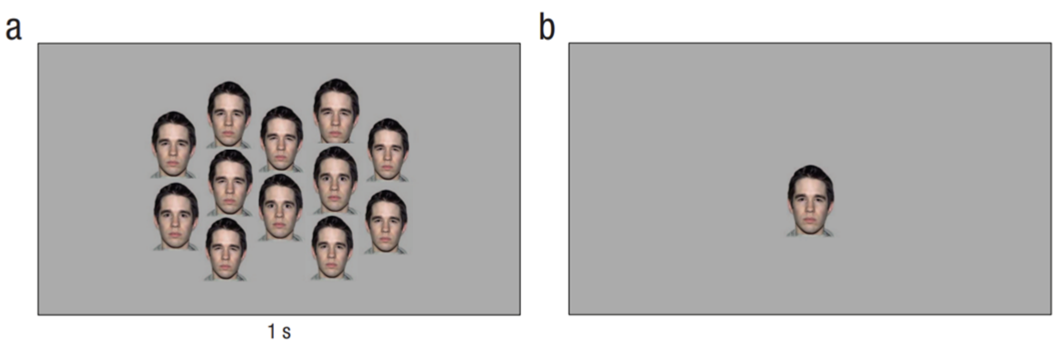
= 0.03, t(7233) = 3.86, $p < .001$, R2 = .08 (Figure 3). Supporting our third hypothesis, the results showed that the amplification effect was stronger for groups expressing negative emotions compared to groups expressing positive emotions, with b = 0.24, 95% CI = [0.02, 0.45], SE = 0.11, t(7253) = 2.18, p = .02, R2 = .08. The interaction effect between facial quantity and facial emotional value was not significant, with b = 0.08, 95% CI = [-0.13, 0.30], SE = 0.11, t(7226) = 0.76, p = .44, R2 = .08.

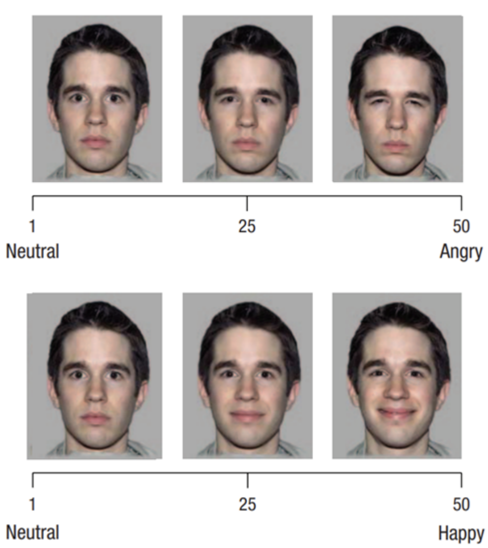
In summary, participants' estimations of facial ensembles' emotions were stronger than the actual emotions, and the amplification effect increased with the size of the ensemble, with a stronger effect for negative emotions compared to positive emotions. However, it is important to note that the third effect was relatively weak, consistent with mixed results in the relevant literature. What mechanisms contribute to our amplification effect? Could it be due to participants not having enough time for more than one or two fixations, resulting in a weakened effect that would diminish if they had more time? One possibility is that increasing exposure time would allow participants to view more faces, providing them with a larger sample to estimate the mean and potentially reducing or eliminating the amplification effect. However, another possibility is that increasing exposure time would allow more time to fixate on stronger facial expressions, thereby increasing the magnitude of the amplification effect. We were fortunate that the code was able to produce results consistent with those of the original authors. However, during the process of replication, we still encountered many aspects that could not be perfectly reproduced. There may be bugs in the code, unclean data preprocessing, lack of visualizations resulting in poor readability, and so on. This highlights the importance of advancing open science, where we can learn from the knowledge of others and also identify minor flaws in their research, prompting researchers to make improvements. This plays a crucial role in promoting the overall healthy development of science.

References

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. Trends in Cognitive Sciences, 15(3), 122–131. https://doi.org/10.1016/j.tics.2011 .01.003 Eimer, M., & Holmes, A. (2007). Event-related brain potential correlates of emotional face processing. Neuropsychologia, 45(1), 15–31. https://doi.org/10.1016/j.neuropsycholo gia.2006.04.022 Kanaya, S., Hayashi, M. J., & Whitney, D. (2018). Exaggerated groups: Amplification in ensemble coding of temporal and spatial features. Proceedings of the Royal Society B: Biological Sciences, 285(1879), Article 20172770. https:// doi.org/10.1098/rspb.2017.2770 Pessoa, L., McKenna, M., Gutierrez, E., & Ungerleider, L. G. (2002). Neural processing of emotional faces requires attention. Proceedings of the National Academy of Sciences, USA, 99(17), 11458–11463. https://doi.org/10.1073/pnas.172403899 Sweeny, T. D., Haroz, S., & Whitney, D. (2013). Perceiving group behavior: Sensitive ensemble coding mechanisms for biological motion of human crowds. Journal of Experimental Psychology: Human Perception and Performance, 39(2), 329–337. https://doi.org/10.1037/a0028712 Whitney, D., Haberman, J., & Sweeny, T. D. (2014). From textures to crowds: Multiple levels of summary statistical perception. In J. S. Werner & L. M. Chalupa (Eds.), The new visual neurosciences (pp. 695–709). MIT Press. Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. Annual Review of Psychology, 69(1), 105–129. https://doi.org/10.1146/annurev-psych-010416-044232

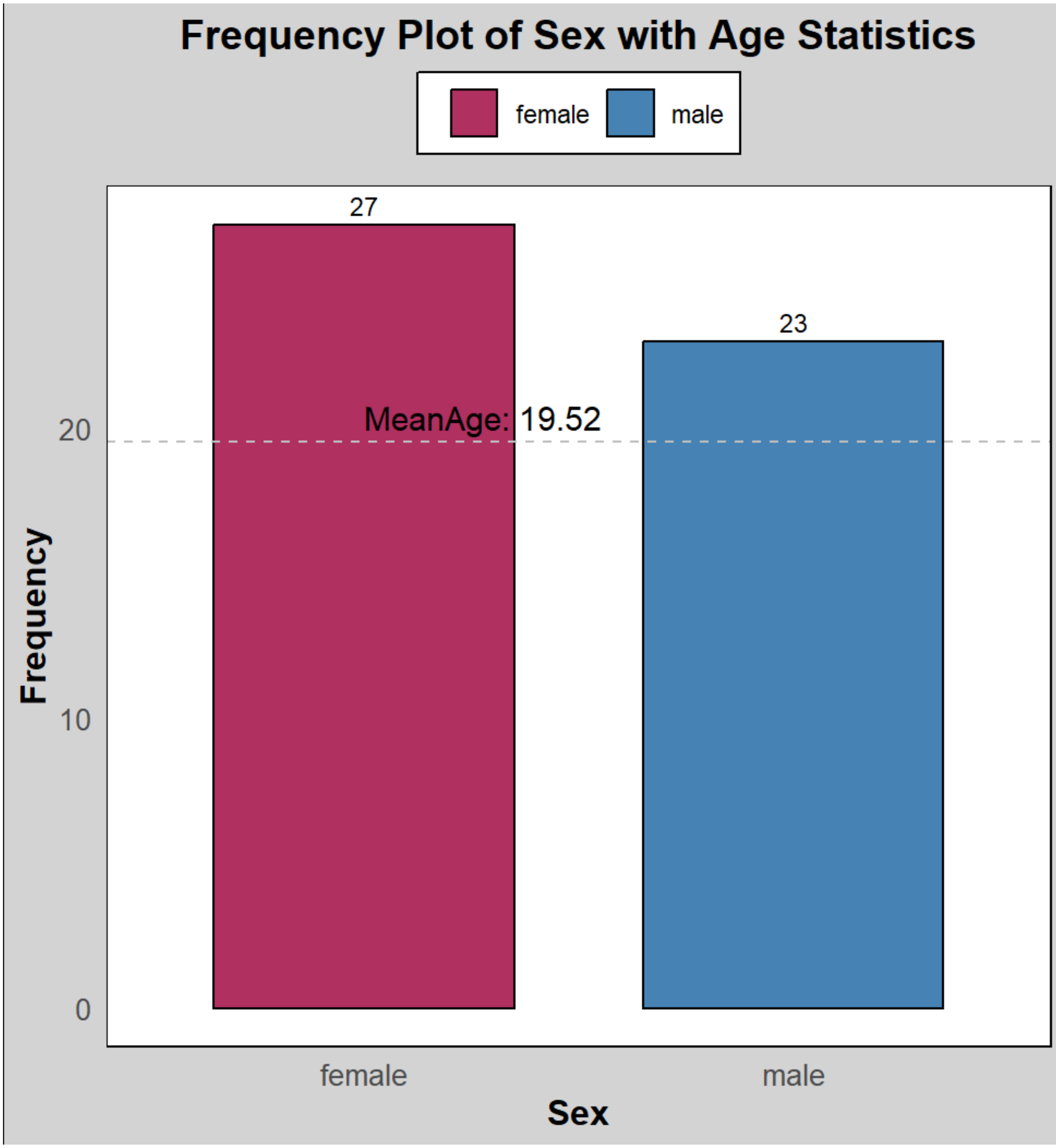
Appendix

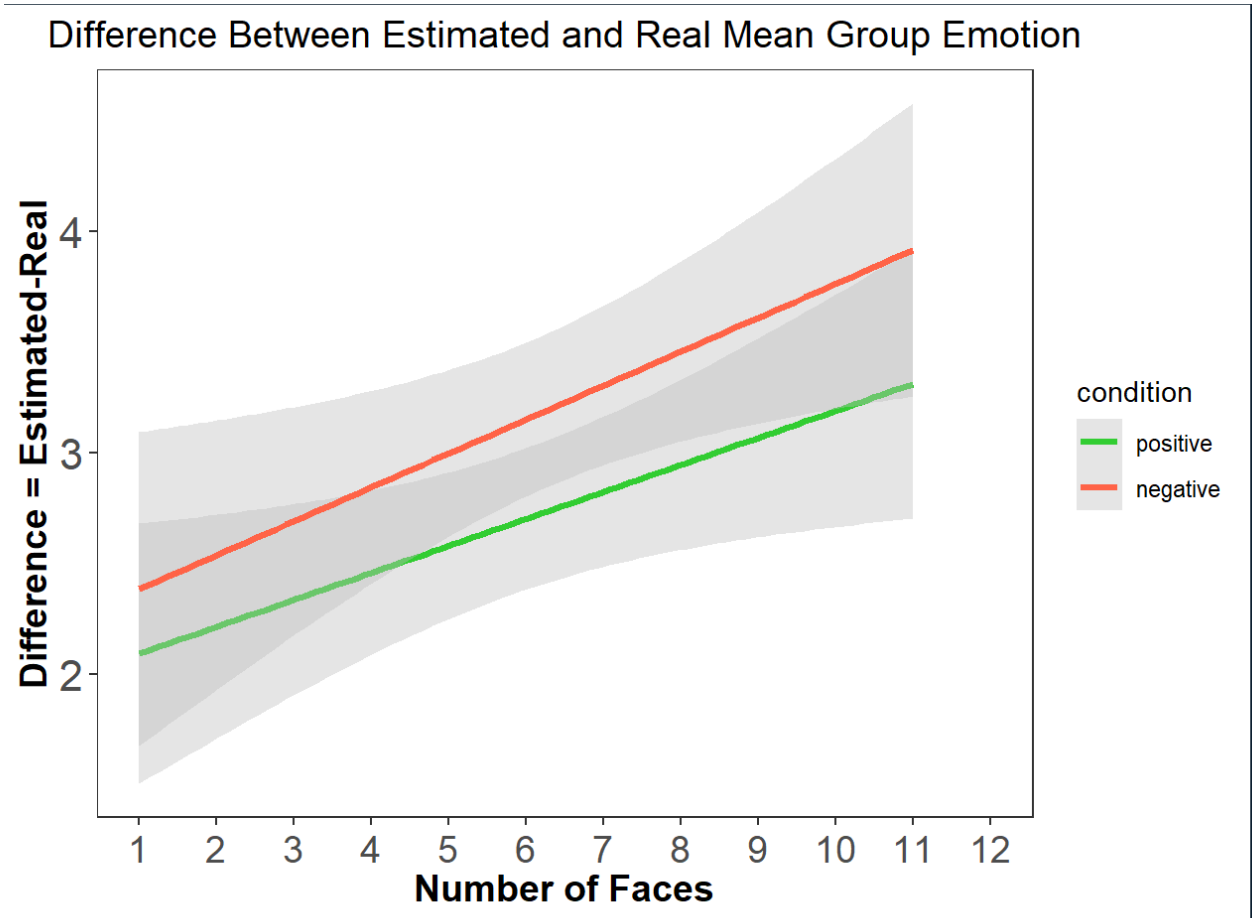


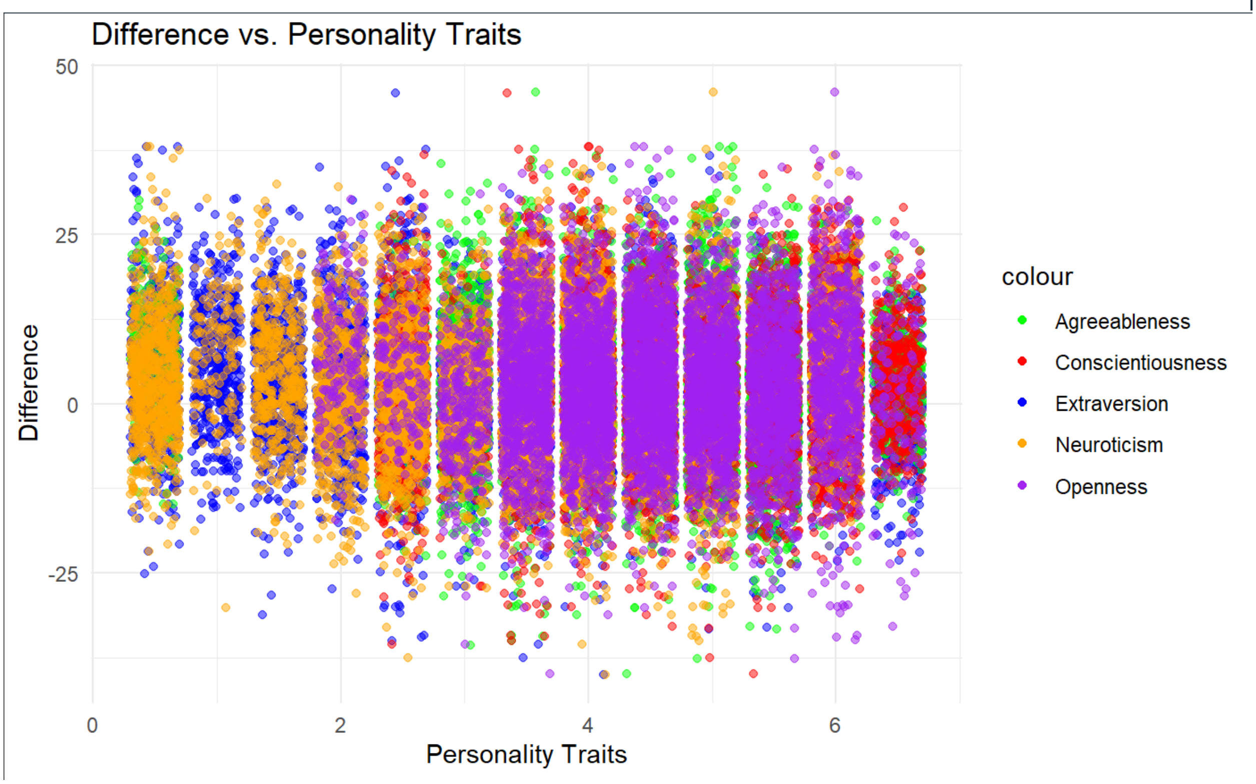


```
> # 对假设1进行验证
> r = lmer (value~ratingType+(1|id)+(1|faceIdentity),
+           dif);summary(r)
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: value ~ ratingType + (1 | id) + (1 | faceIdentity)
   Data: dif

REML criterion at convergence: 109866

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.8947 -0.7168  0.0023  0.6852  2.5487

Random effects:
 Groups       Name        Variance Std.Dev.
 id           (Intercept)   2.63    1.621
 faceIdentity (Intercept)   0.84    0.916
 Residual                 107.25   10.356
Number of obs: 14608, groups:  id, 50; faceIdentity, 4

Fixed effects:
                          Estimate Std. Error        df t value              Pr(>|t|)
(Intercept)                 24.204      0.527     4.865    46.0            0.00000013 ***
ratingTypeEstimated Mean     2.874      0.171 14553.558    16.8 < 0.0000000000000002 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
            (Intr)
rtngTypEstM -0.163
> #混合线性模型对假设1进行检验，基于dif数据集，以value为响应变量，ratingtype为预测变量，id和faceidentity
为随机变量，进行分析
>
> rsquare(r)
[1] 0.04891
> # 计算回归模型“r”的R方值
```

```
>
> confint(r, level =.95)
Computing profile confidence intervals ...
                           2.5 % 97.5 %
.sig01                    1.2980  2.045
.sig02                    0.4388  2.039
.sigma                   10.2380 10.476
(Intercept)              23.0742 25.333
ratingTypeEstimated Mean  2.5381  3.210
> # 计算模型“r”的95%的置信区间
```


Frequency Plot of Sex with Age Statistics
female
male
27
23
MeanAge: 19.52
20
10
0
Frequency
female
male
Sex


Difference Between Estimated and Real Mean Group Emotion
Difference = Estimated-Real
2
3
4
1 2 3 4 5 6 7 8 9 10 11 12
Number of Faces
condition
positive
negative


Difference vs. Personality Traits
Difference
50
25
0
-25
0 2 4 6
Personality Traits
colour
Agreeableness
Conscientiousness
Extraversion
Neuroticism
Openness