Is there a difference in amusement ratings between the three facial expression conditions?

Group Member 1: Sezin Mumcu
Student Number: s3514617

Group Member 2: Amelie Gross
Student Number: s3513858

Group Member 3: Ekin Asci
Student Number: s3483231

Group Member 4: Lorant Kutnyanszky
Student Number s3534685

Link to the OSF page: https://osf.io/br2kx/

Author Contribution Statement

Drafting pre-analysis plan: Ekin, Sezin, Amelie, Lorant
Drafting presentation: Ekin, Sezin, Amelie, Lorant
Recording the presentation: Lorant
Making figures: Ekin
Performing analysis: Ekin, Sezin, Amelie, Lorant
Report: Ekin, Sezin, Amelie, Lorant

Introduction

The facial feedback hypothesis states that facial expression not only serves to express emotion but also influences the subjective experience of emotion (Soderkvist et al., 2018). Strack et al. (1988) initially tested this using a pen-in-mouth paradigm. Participants were assigned to one of three conditions: smiling, neutral, or pouting. Then, they had to watch amusing video clips while holding a pen with their teeth to elicit a smile in the smiling condition or with their lips to elicit a pouting face in the pouting condition. The original study found that the people in the smiling condition rated amusement clips consistently higher than pouting participants, which supports the hypothesis. However, a lot of replications did not come to the same conclusion. Therefore, we also recreated the study to see if there is a difference in amusement ratings between the smiling, pouting, and neutral facial expression conditions. We did this by analysing a pre-existing dataset for this experiment.

We expected our analysis to support the original study by Strack et al. (1988). Therefore, we believed we would find a significant difference between the three conditions. With amusements rated on a 7-point scale, we expected an increase of 0.64 points in the mean amusement ratings from pouting to smiling conditions and 0.37 points in the mean amusement ratings from neutral to smiling, just like in the original study. Our hypotheses specifically were:

H_0: There is no significant difference in the amusement ratings between the three conditions.

H_A: The amusement ratings are significantly higher in the smiling face condition than in the pouting face condition by 0.64 points.

This report aims to describe the process of analysing the dataset and our experience with this process. First, we will go through all the individual steps of executing the analysis and report the outcomes. Then, we will discuss what we found, what it means, and interpret the results. Lastly, we will conclude by reflecting on the entire project. We will write about what went well and what didn’t, what we found surprising, what we should have done differently, and our experiences compared to our expectations.

Method and Results

IBM SPSS 27 was used to conduct the statistical analysis. We started by preparing the variables by indicating the value labels for categorical variables Age, Sexe and AttentionCheck1-2. Then, we scanned the data for missing values by making a frequency table of all variables. There were no missing values. We also checked data quality by creating a descriptive statistics table and checking the minimum and maximum values for each variable. All variables were in the predetermined range. Then, we searched for participants who scored 0 on both AttentionCheck1 and AttentionCheck2, indicating they did not pay sufficient attention to the stimulus material. For this, we calculated the mean of AttentionCheck1-2. We filtered cases where AttentionCheckMean = 0. Accordingly, we removed participants with ID 11, 21, 27, 31, 89, 105 and 113. Although we stated in our pre-analysis plan that listwise deletion based on AttentionCheck1 and AttentionCheck2 would be done at a later stage, this was a necessary deviation since excluding participants at this stage could affect our assessment of influential cases.

To see if there were any outliers, we converted AmusedClip1-9 to z-scores. We made a descriptive statistics table and checked the minimum and maximum values for the z-scores. This showed that zAmusedclip7 has data points outside the range (-3, 3). To see which participants these data points belong to, we filtered cases where zAmusedclip7 was outside the range. This indicated that participants with the ID 100 and 148 had outliers. To check for influential cases, we calculated Cook's distance. For this, we computed the mean of AmusedClip1-9. The maximum value for Cook's distance was 0.044, higher than the cut-off 0.028, as determined by the 4/(N-P) (in this case, 4/(146-3)). Accordingly, we filtered cases where Cook's distance was higher than 0.028, which belonged to participants with ID 32, 46, 122. 147, 151. We noted outliers and influential cases to conduct ANOVA before and after excluding these participants.

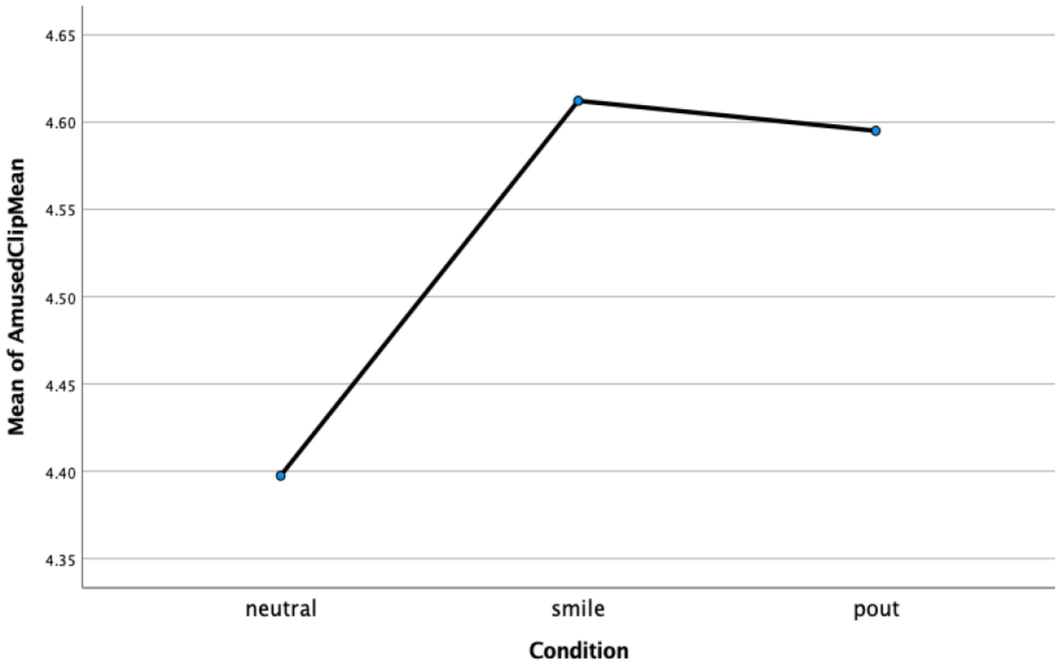
Then, we proceeded to check the assumptions of ANOVA. This was deemed an insignificant difference. To see if AmusedClipMean was normally distributed, the Shapiro-Wilk test was performed and did not show evidence of non-normality (W = 0.984, p = 0.084). Levene's test was conducted to verify the assumption of homogeneity of variances. Levene's test was not significant, indicating equal variances (F = 0.173, p = .842). Hence, the assumptions of ANOVA were confirmed.

One-way ANOVA was conducted to compare the effect of facial expression on amusement ratings across three conditions: pout, neutral, and smiling. The ANOVA showed that there was not a statistically significant difference between conditions within a 95% Confidence

Interval: F(2, 143) = 0.828, p = 0.439. η^2= 0.011. ANOVA was conducted again after removing outliers and influential cases (see Figure 1). Assumptions of normality and homogeneity of variances were verified once again via the Shapiro-Wilk test (W = 0.986, p = 0.168) and Levene's test (F = 0.538, p = .585), respectively. The ANOVA showed that there was not a statistically significant difference between conditions within a 95% Confidence Interval: F(2, 136) = 1.683, p = 0.190, η^2= 0.024. As a result, post-hoc analyses such as Tukey's Honest Significant Difference Test – and hence checking for unequal group sizes – were deemed unnecessary. The results suggest that facial expression does not significantly affect amusement.

Figure 1.

Means Plot of AmusedClipMean

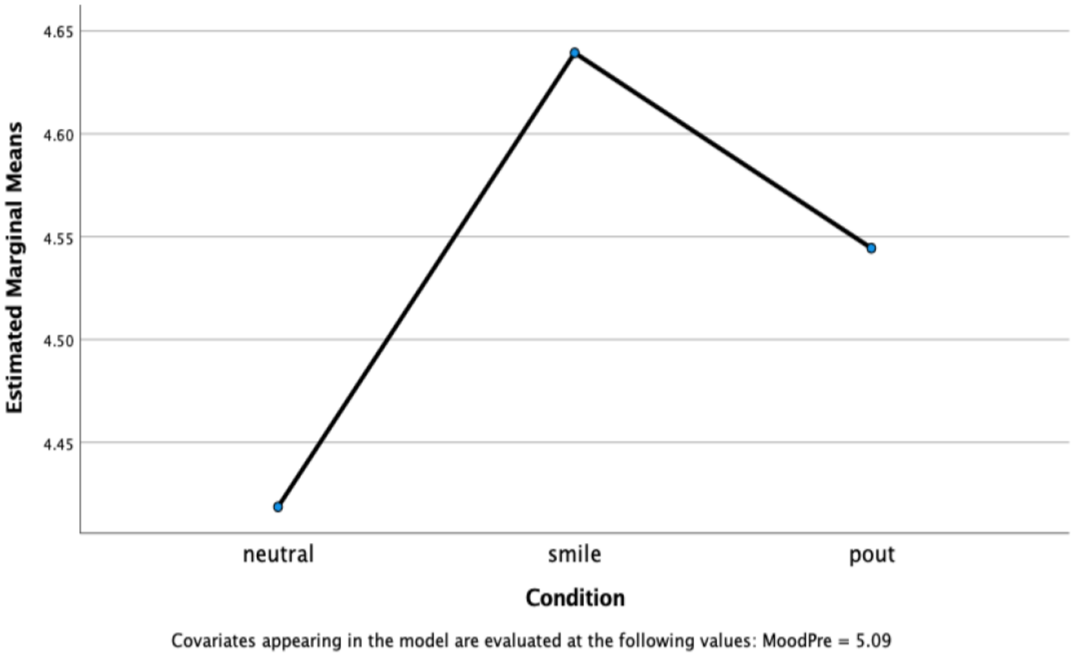


Exploratory Results

An Analysis of Covariance (ANCOVA) was conducted to examine the effect of condition on amusement ratings while controlling for mood (see Figure 2). The covariate, MoodPre, was significantly related at a 95% Confidence Interval level to the amusement ratings, F(1, 139) = 12.107, p < 0.001, n_p^2 = 0.082. After adjusting for MoodPre, there was no significant effect at a 95% Confidence Interval level of condition on amusement ratings, F(2, 139) = 1.600, p = 0.206, n_p^2 = 0.023.

Figure 2.

Estimated Marginal Means Plot of AmusedClipMean



Discussion

Our one-way ANOVA analysis aimed to investigate differences in amusement ratings across three facial expression conditions: neutral, pout, and smiling. The results did not yield statistically significant differences, and this finding remained consistent after controlling for outliers. These results suggest that, within the context of our study, the facial feedback hypothesis, as operationalised by the pen-in-mouth paradigm, may not have had the predicted effect on the subjective experience of amusement. The lack of significant findings contrasts Strack et al.'s original study (1988). This discrepancy could be attributed to several factors, including the possibility that the pen-in-mouth paradigm may not effectively elicit genuine emotional expressions or that the influence of induced facial expressions on emotional experiences may be more nuanced than previously understood.

We also explored the potential influence of participants' baseline mood on amusement ratings using an ANCOVA. While MoodPre, the covariate, was significantly related to amusement ratings with a large effect size (n_p^2 = 0.082), the ANCOVA revealed no significant effect of facial expression condition on amusement ratings after adjusting for MoodPre. This indicates that while mood significantly and largely practically contributed to the variation in amusement ratings, the facial expression conditions did not have an additional effect beyond what could be attributed to the participants' initial mood states.

The significant relationship between mood and amusement ratings underscores the importance of considering individual emotional states as a potential confounding variable in such studies. It suggests that baseline affective states can influence cognitive evaluations of stimuli, which aligns with previous research highlighting mood-congruent biases in information processing. Regardless, the absence of significant differences in our study raises questions about the robustness and generalisability of the facial feedback hypothesis. Our findings, along with failed replications of the original study (Wagenmakers et al., 2016), suggest that if there is an effect, it may be context-dependent or influenced by methodological variations.

Future research could benefit from exploring alternative paradigms to elicit and measure facial expressions and their emotional correlates. Moreover, considering the observed impact of mood, future studies should control for this variable or examine its interaction with other factors more closely. Longitudinal designs could clarify whether facial expressions have a cumulative effect on mood and emotional experiences over time.

Reflection

Our pre-analysis plan was mostly unambiguous and accounted for everything except that we did not have any guidelines for what counts as unequal group sizes and its possible implications. Ultimately, the (un)equality of the group sizes was irrelevant to our results as we only needed it to decide which post-hoc test to use in case the ANOVA was significant, which wasn't. Still, we would have been unprepared concerning its impact on our results and potential conclusion. This meant that we had to research this after conducting our analysis.

In retrospect, we realised that our first step should have been inspecting the data properly before applying our guidelines for missing data, influential cases and outliers. Instead, we started by applying our guidelines to the data, which hindered our progress because we realised afterwards that we had no missing data. This means that we did not have to apply our proposed guidelines to the dataset, and we wasted time when we could have worked on our analysis.

When creating our pre-analysis plan, we should have considered the practicality of implementing our guidelines. An aspect that we did not consider appropriately in our pre-analysis plan is whether we were able to execute the guidelines that we had set for ourselves using the software that we wanted to use. We experienced difficulties executing the guidelines for our missing data in SPSS because we had not considered if we could actually execute it the way we wanted to. We realised afterwards that our guidelines would have worked better if we had used R instead. This impacted the practicality of our analysis as we struggled for a while trying to execute what we wanted to do. If we were to write our pre-analysis again, we should consider in more depth which software is the most suitable by understanding how our guidelines would be applied to the software we want to use. Therefore, we should have spent more time researching the usage of SPSS and testing out different syntaxes when considering our guidelines.

Many of the difficulties we experienced could have been avoided if we had tried out our pre-analysis plan on a mock dataset because it would have allowed us to work out any potential flaws that could occur. Therefore, this would have enabled us to have a better pre-analysis plan.

In conclusion, next time, we will produce a more realistic and detailed pre-analysis plan by focusing on the order of steps and running the plan in a mock dataset to conduct the analysis more efficiently.

Syntax

```
* Encoding: UTF-8.

*Check for missing values and ranges by making a frequency table including all variables.
No missing values exist, and all variables are in the predetermined range.
FREQUENCIES VARIABLES=Condition Sexe Age AmusedClip1 AmusedClip2
AmusedClip3 AmusedClip4 AmusedClip5 AmusedClip6 AmusedClip7 AmusedClip8
AmusedClip9 AmusedClip10 MoodPre AttentionCheck1 AttentionCheck2
  /STATISTICS=STDDEV RANGE MINIMUM MAXIMUM MEAN MEDIAN MODE
  /ORDER=ANALYSIS.

*Compute the mean of AttentionCheck1 and AttentionCheck2.
DATASET ACTIVATE DataSet1.
COMPUTE AttentionCheckMean=MEAN(AttentionCheck1,AttentionCheck2).
EXECUTE.

*Select cases where AttentionCheckMean > 0. Filter unselected cases. Participants with ID
11, 21, 27, 31, 89, 105, 113 are 0 scorers on both AttentionCheck1 and AttentionCheck2.
USE ALL.
COMPUTE filter_$=(AttentionCheckMean > 0  ).
VARIABLE LABELS filter_$ 'AttentionCheckMean > 0   (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.

*Delete 0 scorers on AttentionCheck1 and AttentionCheck2.
FILTER OFF.
USE ALL.
SELECT IF (AttentionCheckMean > 0).
EXECUTE.

*Compute z-scores for AmusedClip1-7.
```

```
DESCRIPTIVES VARIABLES=AmusedClip1 AmusedClip2 AmusedClip3 AmusedClip4
AmusedClip5 AmusedClip6 AmusedClip7 AmusedClip8 AmusedClip9
  /SAVE
  /STATISTICS=MEAN STDDEV MIN MAX.

*To check for outliers, make a frequency table of zAmusedclip1-7. Check the maximum and
minimum values. zAmusedclip7 has data points outside the range (-3, 3).
FREQUENCIES VARIABLES=ZAmusedClip1 ZAmusedClip2 ZAmusedClip3
ZAmusedClip4 ZAmusedClip5 ZAmusedClip6  ZAmusedClip7 ZAmusedClip8
ZAmusedClip9
  /STATISTICS=STDDEV RANGE MINIMUM MAXIMUM MEAN MEDIAN MODE
  /ORDER=ANALYSIS.

*Filter cases where zAmusedclip7 is outside the range (-3, 3). Participants with ID 100 and
148 are outside the range.
USE ALL.
COMPUTE filter_$=(ZAmusedClip7 < 3  &  ZAmusedClip7 > -3).
VARIABLE LABELS filter_$ 'ZAmusedClip7 < 3  &  ZAmusedClip7 > -3 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.

*Turn off the filter for now to include outliers.
FILTER OFF.
USE ALL.
EXECUTE.

*Compute the mean of AmusedClip1-9.
DATASET ACTIVATE DataSet1.
COMPUTE
AmusedClipMean=MEAN(AmusedClip1,AmusedClip2,AmusedClip3,AmusedClip4,Amuse
dClip5,AmusedClip6,AmusedClip7,AmusedClip8,AmusedClip9).
EXECUTE.
```

```
*Calculate Cook's d to check for influential cases. Check maximum Cook’s d.
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT AmusedClipMean
  /METHOD=ENTER Condition
  /SAVE COOK.


*Filter influential cases, as indicated by Cook's d > 0.028. These belong to participants with ID 32, 46, 122. 147, 151.
DATASET ACTIVATE DataSet1.
USE ALL.
COMPUTE filter_$=(COO_1 < 0.028).
VARIABLE LABELS filter_$ 'COO_1 < 0.028 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.

*Turn off the filter for now to include influential cases.
FILTER OFF.
USE ALL.
EXECUTE.

*Check for the assumption of unequal variances via the Shapiro-Wilk test.
EXAMINE VARIABLES=AmusedClipMean
  /PLOT BOXPLOT STEMLEAF NPPLOT
  /COMPARE GROUPS
  /STATISTICS DESCRIPTIVES
  /CINTERVAL 95
```

```
  /MISSING LISTWISE
  /NOTOTAL.

* Conduct ANOVA [dependent list: AmusedClipMean; factor: Condition]. Check for the assumption of homogeneity via Levene's test.
ONEWAY AmusedClipMean BY Condition
  /STATISTICS HOMOGENEITY
  /MISSING ANALYSIS
  /CRITERIA=CILEVEL(0.95).

*Delete outliers listwise.
FILTER OFF.
USE ALL.
SELECT IF (ZAmusedClip7 < 3  &  ZAmusedClip7 > -3).
EXECUTE.

*Delete influential cases listwise.
FILTER OFF.
USE ALL.
SELECT IF (COO_1 < 0.028).
EXECUTE.
*Check for the assumption of unequal variances via the Shapiro-Wilk test again after removing outliers and influential cases.
EXAMINE VARIABLES=AmusedClipMean
  /PLOT BOXPLOT STEMLEAF NPPLOT
  /COMPARE GROUPS
  /STATISTICS DESCRIPTIVES
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.

*Conduct ANOVA again after removing outliers and influential cases.
ONEWAY AmusedClipMean BY Condition
  /STATISTICS HOMOGENEITY
```

```
  /MISSING ANALYSIS
  /CRITERIA=CILEVEL(0.95).

*Conduct ANCOVA [between-subjects factor: Condition; covariate: MoodPre].
UNIANOVA AmusedClipMean BY Condition WITH MoodPre
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /PRINT ETASQ DESCRIPTIVE
  /CRITERIA=ALPHA(.05)
  /DESIGN=MoodPre Condition.
```

References

Soderkvist, S., Ohlen, K., Dimberg, U. (2018). How the Experience of Emotion is Modulated by Facial Feedback. Journal of Nonverbal Behavior, 42, 129–151.

Strack, F., Martin, L. L., & Stepper, S. (1988). Inhibiting and Facilitating Conditions of the Human Smile: A Nonobtrusive Test of the Facial Feedback Hypothesis. *Journal of Personality and Social Psychology, 54,* 768-777.

Wagenmakers, E. J., Beek, T., Dijkhoff, L., Gronau, Q. F., Acosta, A., Adams, R. B., Albohn, D. N., Allard, E. S., Benning, S. D., Blouin-Hudon, E.-M., Bulnes, L. C., Caldwell, T. L., Calin-Jageman, R. J., Capaldi, C. A., Carfagno, N. S., Chasten, K. T., Cleeremans, A., Connell, L. . . . Zwaan, R. A. (2016). Registered Replication Report: Strack, Martin, & Stepper (1988). *Perspectives on Psychological Science, 11*, 917–928.

AI Disclosure

Throughout the writing of this report, several AI tools have been used: Grammarly for grammar and spelling checks, Scribbr for generating and organizing references, and OpenAI's GPT-4 for getting additional explanations on statistical models and comparing various statistical tests. These AI tools were instrumental in enhancing the quality of the report, ensuring correct grammar structure and better clarity, citation consistency in the correct format (APA 7th edition), and the increased reliability of the statistical analyses chosen. The use of these tools was strict as assistive technology; the final content, choices, interpretations, and conclusions remain the authors' sole responsibility.