**The Many Smiles Collaboration: A Multi-Lab Test of the Facial Feedback Hypothesis**

Pre-registration

**Study Information**

**Title**

A multi-lab test of the facial feedback hypothesis by The Many Smiles Collaboration

**Authors**

Nicholas Coles, David March, Fernando Marmolejo Ramos, Arinze Nwadiogo, Izuchukwu Ndukaihe, Asil Ali Özdoğru, Balazs Aczel, Nandor Hajdu, Tamas Nagy, Bidisha Som, Dana Basnight-Brown, Danilo Zambrano, Lady Grey Javela, Francesco Foroni, Megan Willis, Gerit Pfuhl, Gwenael Kaminski, Tracy Ehrengarth, Hans IJzerman, Kevin Vezirian, Hassan Banaruee, Omid Kathin-Zadeh, Isabel Suarez, Jeff T. Larsen, José Antonio Hinojosa Poveda, Pedro Montoro, Krystian Barzykowski, Katarzyna Filip, Sylwia Adamus, Michal Parzuchowski, Aneta Marczak, Michalina Tanska, Natalia Trujillo, Sandra Trujillo, Natalie Butcher, Daniel Eaves, Nikolay Dagaev, Elena Gorbunova, Niv Reggev, Pascal Gygax, Sarah Pressman, John Hunter, Miranda Pinks, Ian Waldrop, Susana Ruiz-Fernandez, Yuki Yamada, Ayumi Ikeda, Carmel Levitan, Fritz Strack, Lowell Gaertner, Phoebe Ellsworth, Marco Marozzi, Marco Tullio Liuzza

**Research Questions**

Q1. Does posing happy vs. neutral expression cause people to feel happier?

Q2. Do happy facial poses only influence feelings of happiness if they resemble a natural expression of happiness?

Q3. Can facial poses initiate emotional experience in otherwise neutral scenarios or can they only modulate ongoing emotional experiences?

Q4. Are facial feedback effects eliminated when controlling for awareness of the experimental hypothesis?

**Hypotheses**

Q1.H1: Participants will report experiencing more happiness when posing happy vs. neutral facial expressions (i.e., main effect for Pose).

Q2.H1: The difference in self-reported happiness when posing happy vs. neutral facial expressions will not be larger during tasks that produce more natural expressions of happiness.

Q3.H1: Participants will report experiencing more happiness when presented with positive stimuli (i.e., main effect for Stimuli Presence).

Q3.H2: The difference in self-reported happiness when posing happy vs. neutral facial expressions will be larger in the presence vs. absence of emotional stimuli (i.e., interaction between Pose and Stimuli Presence).

Q4.H1: Participants who are judged to be completely unaware of the purpose of the experiment will report experiencing more happiness when posing happy vs. neutral facial expressions (i.e., main effect for Pose when limiting analyses to completely unaware participants).

Q4.H2: Participants who engage in a Facial Movement Task that limits participant awareness (i.e., the pen-in-mouth task) will report experiencing more happiness when posing happy vs. neutral facial expressions.

**Sampling Plan**

**Existing Data**

Registration prior to creation of data.

**Explanation of existing data**

N/A

**Data collection procedures**

Each lab that participates in this study will be requested to collect data from 105 participants. At the time of pre-registration, 22 labs have joined the project. Therefore, we currently expect up to 2,310 participants.

**Sample size**

Up to 2,310 participants.

**Sample size rationale**

The project will use a 2 (Pose: happy or neutral) x 3 (Facial Movement Task: facial mimicry, voluntary facial action technique, or pen-in-mouth) x 2 (Stimuli Presence: present or absent) design, with type of pose manipulated within-participants and Facial Movement Task and stimuli presence manipulated between-participants.

Because of the clustered nature of our experimental design (individuals crossed with type of pose conditions, individuals nested within research groups), we plan to analyze the data using Linear Multilevel Modeling (LMM, Pinheiro and Bates 2000).

Primary Power Analysis

Our primary power analysis was performed via simulation (Gelman & Hill, 2007) in R (R Core Team, 2018, R version 3.4.4) using the lme4 package (Bates et al., 2015, version 1.1-17).

Power analysis specs are below:

* Number of research groups was set to 22
* Effect size (ES) estimates for the Pose (happy vs. neutral), Stimuli Presence (present vs. absent), and their interaction were set based on the ESs observed in the two pilot studies[[1]](#footnote-1). Using an effect coding scheme, the following estimates were derived:
  + Within-subjects effect of Pose= 0.39
  + Between-subjects effect of Stimuli Presence = 0.68
  + Within-between-subjects effect of the interaction = - 0.29
  + All the other effects were set to zero, and therefore serve the purpose of estimating the rate of false positives (Type I error)
* For participants (ID\_Sub), we included random intercepts with SD = 0.7, based on the estimate from the pilot study. We did not simulate random slopes for the single participants because the number of observations within each participant are too few, which would likely lead to convergence issues.
* For research groups (ID\_Lab), we included random intercepts with SD = 0.00 because, for simplicity, we simulated data analysis with scores standardized within each research group. This will not be our final choice but is an assumption that kept the simulation parsimonious. After collecting data, we will analyze the raw scores. Therefore, the research group intercepts will be different from zero.
* For research groups (ID\_Lab), we included random slopes. In modeling the random structure, we kept it maximal in order to have a better control of the Type I error (Barr et al. 2013). However, to improve the model convergence, we did not model the correlation between slope and intercept. In our final data analysis, we will include the correlation between slopes and intercepts unless we encounter convergence issues. In that case, we will first try to standardize our variables. If our models will still fail to converge, we will set the covariance between slopes and intercepts to zero.
  + For our non-null effects of interest, we set the random slope as big as the standard deviation for the meta-analytic effect from Wagenmakers et al. (2016). The authors reported a standard error = 0.068 from 17 labs, which leads to an approximate estimation of the standard deviation = 0.28. For expected null effects, we simulated a random slope as small as the random slope of the data from Wagenmakers et al. (2016), who found an estimate tau2 that approaches zero.
* Residual variance was set to 0.6, based on the estimate from the pilot study.

For the power analysis, we randomly generated normally distributed data for 1000 iterations with sample size set to 72, 96, and 120. These sample sizes were chosen because they represented the number of participants we believed each research group could feasibly collect. For each iteration, the following model was tested (in Wilkinson notation)[[2]](#footnote-2):

DV ~ Pose \* Facial Movement Task \* Stimuli Presence +

(1 | ID\_Sub) +

(0 + Pose | ID\_Lab) +

(0 + Stimuli Presence | ID\_Lab) +

(0 + Pose : Stimuli Presence | ID\_Lab)

In the model there are three factors: Pose (happy or neutral), Facial Movement Task (facial mimicry, voluntary facial action technique, or pen-in-mouth task), and Stimuli Presence (present vs. absent).

Power was computed as the proportion of absolute t values > 2, as suggested in footnote 1 of Baayen, Davidson, and Bates (2008). The same percentage provides information on the Type I error, when looking at the same proportion for effect that were set as 0 (e.g., the effect of task).

Secondary Power Analysis

To ensure that our power estimates were robust, we also re-ran the power simulation using simR package (Green & McLeod, 2016), wherein power was estimated from the model fitted on our simulated data. In order to keep the computational burden lower, we estimated the model with only the within-subjects effect of Pose, the between-subjects effect of Stimuli Presence = 0.68, and the mixed effect of the hypothesized interaction = - 0.29.

Power Analyses Results

Results from the primary power analysis indicated that, with a sample of n = 72, we can achieve (a) approximately 99% power to detect a main effect for Pose (happy, vs. neutral) and Stimuli Presence (present vs. absent) and (b) approximately 96% power to detect the hypothesized interaction between Pose and Stimuli Presence.

Results from the secondary power analysis indicated that, with a sample of n = 27, we can achieve (a) = approximately 93% power to detect the main effects of Pose and Stimuli Presence. However, in order to achieve over ≥ 95% for the hypothesized interaction between Pose and Stimuli Presence (per NHB power guidelines), a sample size of n ≥ 96 is necessary.

Conclusion

Results from the primary power simulation indicate that over 95% power for all our hypothesized effects could be obtained with at least 1,584 participants. However, based on a pilot study, we estimate that 44% of participants will not meet our strict inclusion criteria, leading to a desired sample of 2,281. Consequently, we will stop collecting data once one of the following conditions are met: (a) each lab has collected 105 participants, or (b) at least six months have elapsed since the start of data collection and we have a total of 2,281 participants.

**Stopping rule**

We will stop collecting data once one of the following conditions are met: (a) each lab has collected 105 participants, or (b) at least six months have elapsed since the start of data collection and we have a total of 2,281 participants.

**Variables**

**Manipulated variables**

The project will use a 2 (Pose: happy or neutral) x 3 (Facial Movement Task: facial mimicry, voluntary facial action technique, or pen-in-mouth) x 2 (Stimuli Presence: present or absent) design, with Pose manipulated within-participants and Facial Movement Task and Stimuli Presence manipulated between-participants.

**Measured variables**

\*Self-reported emotional experience\*: after each task, participants will report the extent to which they experienced happiness, anger, anxiety, tiredness, and confusion using a subset of the Discrete Emotions Questionnaire (Harmon-Jones, Bastian, & Harmon-Jones, 2016). We will include four items measuring happiness, two items measuring anxiety, and single items measuring anger, tiredness, and confusion. Following the instructions in the Discrete Emotions Questionnaire, to form a happiness score and an anxiety score, responses corresponding to those emotions will be averaged.

\*Self-reported difficulty and liking of the task and arithmetic problems\*: after each task, participants will report the difficulty (‘extremely easy’ to ‘extremely difficult’) and liking (‘not at all’ to ‘an extreme amount’) of the physical task and the accompanying math problems. These questions were included to reinforce the cover story, but these questions could be of interest for future researchers interested in the potential role of task difficulty and liking.

\*Attention check\*: In the non-filler tasks, an attention check item asking participants to choose a specific response option will be randomly inserted in the questions regarding the task and arithmetic problem difficulty.

\*Interoceptive awareness\*: Interoceptive awareness will be measured using the Body Awareness Questionnaire (Shields, Mallory & Simon, 1989) and scored per their instructions (i.e., reverse score item 10 and then take the average of all items).

\*Self-reports of quality of task completion\*: To have some indicator of how well participants completed the task, participants will be asked at the end of the experiment how well they believe the executed the task.

\*Awareness of the purpose of the experiment\*: At the end of the experiment, participants will complete several open-ended questions about they believed the purpose of the experiment was. Using two coders who are blind to the results of the experiment, each lab will code the recorded debriefing of the response prompts (1 = ‘not at all aware’ to 7 = ’completely aware’). These ratings will then be averaged to form an Awareness score. Note: if the reliability of coded responses is low (Cohen’s Kappa < .50), we will urge caution when interpreting results related to this variable.

\*Quality of responses questions\*: Towards the end of the study, participants will answer several questions related to the quality of their response. First, participants will be re-presented with their assigned happy pose instructions and asked to retrospectively rate how well they followed the instructions earlier in the study (1 = “not at all” to 7 = “exactly”). Second, participants will be asked to repeat the task and rate the degree to which it feels like they are expressing happiness (1 = “not at all” to 7 = “exactly”). Third, participants will be asked to watch themselves repeat the task (e.g., via a mirror or camera phone) and indicate the degree to which their expression matches an image of an individual completing the task correctly (1 = “not at all” to 7 = “exactly”). Fourth, participants will be asked to describe any issues that may compromise the quality of their data (e.g., distractions). Two coders from each research group who are blind to the results of the experiment will review the responses and rate the degree to which each participant was distracted (1 = “not at all distracted” to 7 “completely distracted”). Participants will be told that there will not be a penalty for indicating that they did not complete the task correctly or that there are issues with the quality of their data.

\*In-Person vs. online data collection\*: Research groups will have the option to collect data either in-person or online. The mode of data collection will be recorded.

**Indices**

Described above.

**Design Plan**

**Study type**

Experiment - A researcher randomly assigns treatments to study subjects, this includes field or lab experiments. This is also known as an intervention experiment and includes randomized controlled trials.

**Blinding**

For studies that involve human subjects, they will not know the treatment group to which they have been assigned

**Study design**

The project will use a 2 (Pose: happy or neutral) x 3 (Facial Movement Task: facial mimicry, voluntary facial action technique, or pen-in-mouth) x 2 (Stimuli Presence: present or absent) design, with Pose manipulated within-participants and Facial Movement Task and Stimuli Presence manipulated between-participants.

**Randomization**

Participants will be randomly assigned to either perform the facial mimicry, voluntary facial action technique, or pen-in-mouth task using a built-in randomization function in Qualtrics. In each of these conditions, participants will pose a happy and neutral expression in a randomized order (using the randomization function in Qualtrics). Participants will also be randomly assigned to either view a blank screen or emotional stimuli during these poses.

**Analysis Plan**

**Statistical Models**

Because of the clustered nature of our experimental design (individuals crossed with pose, individuals nested within research groups), we plan to analyze happiness self-reports through a linear multilevel modeling approach (LMM, Pinheiro & Bates, 2000). The analyses will be conducted using the lme4 package (Bates et al., 2015, version 1.1-17) using the following formula (in Wilkinson notation):

Happiness.Reports ~ Pose \* Facial Movement Task \* Stimuli Presence +

(1 | ID\_Lab) + (1 | ID\_Sub) +

(0 + Pose | ID\_Lab) +

(0 + Facial Movement Task | ID\_Lab) +

(0 + Stimuli Presence | ID\_Lab) +

(0 + Pose : Stimuli Presence | ID\_Lab) +

(0 + Pose : Facial Movement Task | ID\_Lab) +

(0 + Facial Movement Task : Stimuli Presence | ID\_Lab) +

(0 + Pose : Facial Movement : Stimuli Presence | ID\_Lab).

This analytical approach will model the research groups and participants as sources of randomness. Importantly, by taking into account random slopes for each research group, we will attain a better control of the Type I error (Barr et al., 2013, Barr 2013) while maintaining good power (see the Sample size rationale paragraph).

In order to be able to quantify the relative evidence between the null and alternative hypothesis, we will also conduct the same analysis within a Bayesian Hypothesis testing framework (i.e., Bayes Factor, Wagenmakers, Morey & Lee, 2016) using the Bayes Factor package in R (Morey & Rouder, 2018). To do so, we will fit our full model using the *lmBF* function and will compare it to the simpler nested models, while keeping the random effects constant.

The *lmBF* function uses Cauchy priors on the Fixed Effects. We will conduct a sensitivity analysis using different prior *r scales*. An r scaling parameter for the Cauchy prior represents the ± scaled values where you expect to observe your effect 50% of the times. We will set our r scale prior to 0.5, √2/ 2, and 1 in order to provide sensitivity analyses.

Importantly, we will run this analysis both with and without participants who exhibited awareness of the experimental hypothesis. However, but will consider the test of the Pose factor without aware participants to be the strongest and primary test of the facial feedback hypothesis.

**Transformations**

N/A

**Follow-up analyses**

Are Facial Feedback Effects Moderated by Self-Reported Quality of Task Execution?

Although secondary to the purposes of this research, we expect that participants who report executing the Facial Movement Task better should exhibit larger facial feedback effects. This hypothesis will be tested through the following model:

Happiness.Reports ~ Pose + Pose.Quality+ Pose : Pose.Quality + Facial Movement Task+ Stimuli Presence + (1 | ID\_Lab) + (1 | ID\_Sub) + (0 + Pose | ID\_Lab) + (0 + Stimuli Presence | ID\_Lab) + (0 + Facial Movement Task | ID\_Lab) +  (0 + Pose : Pose.Quality | ID\_Lab).

The Pose : Pose.Quality interaction will serve as our test of whether people with higher reported levels of pose quality exhibit larger facial feedback effects.

Are Participants Less Aware of the Experimental Hypothesis in the Pen-in-mouth Condition?

At the end of the experiment, participants will be asked several questions about what they believed the purpose of the experiment was in either an in-person interview or open-ended survey response prompt. Using two coders who are blind to the results, each lab will code these responses (1 = ‘not at all aware’ to 7 = ’completely aware’). These ratings will then be averaged to form an Awareness score.

The pen-in-mouth task was created to reduce the degree to which participants are aware of the experimental hypothesis (Strack, Martin, & Stepper, 1988). To examine whether awareness of the experimental hypothesis differs as a function of the Facial Movement Task, we will fit the following model:

Awareness.Ratings ~ Facial Movement Task + Stimuli Presence + (1 | ID\_Lab) + (0 + Facial Movement Task | ID\_Lab) + (0 + Stimuli Presence | ID\_Lab)

If the Facial Movement Task main effect is significant, we will perform follow-up pairwise contrasts.

Does Awareness of the Experimental Hypothesis Moderate Facial Feedback Effects?

First, we will examine the hypothesis that people with higher levels of awareness will exhibit larger facial feedback effects. This hypothesis will be tested through the following model:

Happiness.Reports ~ Pose + Awareness + Pose : Awareness + Facial Movement Task + Stimuli Presence + (1 | ID\_Lab) + (1 | ID\_Sub) + (0 + Pose | ID\_Lab) + (0 + Stimuli Presence | ID\_Lab) + (0 + Facial Movement Task | ID\_Lab) +  (0 + Pose : Awareness | ID\_Lab).

The Pose : Awareness interaction will serve as our test of whether awareness moderates facial feedback effects.

Do Participants Report Experiencing More Happiness During the Smiling vs. Filler Trials?

We will compare participants happiness ratings from the smiling trial to filler trials that participants complete to obscure the purpose of the experiment. To do so, we will create a dummy variable that indicates whether the trail was a filler or smiling task (0 = filler; 1 = happy). Then, we will fit the following model:

Happiness.Reports ~ Movement Type + Facial Movement Task + Stimuli Presence + Movement Type : Stimuli Presence + (1 | ID\_Lab) + (1 | ID\_Sub) + (0 + Movement Type | ID\_Lab) + (0 + Facial Movement Task | ID\_Lab) + (0 + Stimuli Presence | ID\_Lab) + (0 + Movement Type : Stimuli Presence).

Since the condition where no stimuli are present provides the cleanest comparison (because it is not confounded by the effects of the positive images on the happiness reports), we will pay particular attention to Movement Type : Stimuli Presence interaction. If significant, we will consider the effect of Movement Type in the no-stimuli-present condition to be a stronger test of this whether participants experience more happiness during the smiling vs. filler trials.

Do Participants Report Experiencing Less Anger after Posing Happy vs. Neutral Expressions?

We will re-run our primary analyses with self-reported anger (as opposed to happiness) as the DV. If participants report experiencing less anger after posing happy expressions, this would provide some evidence that facial feedback influences felt positivity/negativity as opposed to having only emotion-specific effects (see Coles et al. 2019 for more details on this debate).

Are Facial Feedback Effects Moderated by Individual Differences in Bodily Awareness?

To examine this, we will add both (a) the body awareness scores, and (b) a Pose \* Body Awareness interaction term into the primary statistical model.

Are Facial Feedback Effects Moderated by Indicators of Pose Quality?

To examine this, we will test whether the following variables interact with Pose in our primary statistical model: (1) the degree to which participants followed the happy pose instructions earlier in the study; (2) the degree to which it feels like they were expressing happiness during the happy; (3) the degree to which their participants felt that their happy pose expression matches an image of an individual completing the task correctly; (4) the degree to which participants are rated as distracted. Participants will be told that there will not be a penalty for indicating that they did not complete the task correctly or that there are issues with the quality of their data.

In-Person vs. Online Data Collection

To examine if facial feedback effects are moderated by whether participants completed the study online vs. in a lab, we will re-run our primary analysis with data collection setting included as a moderator. However, we will note that this analysis may be confounded by (a) whether the research group is a proponent vs. critic of the facial feedback hypothesis (i.e., proponents may be more likely to collect data in the laboratory) and (b) the region of data collection (i.e., research groups in regions with fewer COVID-19 cases may be more likely to collect data in the laboratory).

Do Facial Movement Tasks Lead to Different Levels of Anxiety?

Although we do not anticipate a Pose by Facial Movement Task interaction, we note that it is possible that the pen-in-mouth condition leads to heightened levels of anxiety in the midst and/or aftermath of COVID-19. Although speculative, heightened levels of anxiety may interfere with facial feedback effects. Consequently, as an exploratory analysis, we will examine if anxiety ratings differ as a function of Facial Movement Tasks.

General Comment

Considering that our study is optimized, in terms of statistical power, to test our primary hypothesis, we will be cautious in the conclusions we will draw from these secondary analyses.

**Inference criteria**

For the frequentist Null Hypothesis Significance Testing (NHST), we set an alpha level = .05.

Bayes Factors in favor of the alternative (BF10) or the null (BF01) will be interpreted according to the recommendations from Lee and Wagenmakers (2013). Therefore, a 3 ≥ BF10 (BF01) < 10 will be regarded as moderate evidence for the alternative (the null), a ≥ 10 BF10 (BF01) < 30 will be regarded as strong evidence for the alternative (the null), etc.

**Data Exclusion**

Participants will be excluded from the primary analyses if they (1) exhibited any awareness of the facial feedback hypothesis (i.e., received an awareness score over 1 from two independent coders), (2) disclosed that they were very distracted during the study (i.e., received an average distraction score above 5 from two independent coders), (3) did not complete the study on a desktop computer or laptop, (4) indicated they did not follow the pose instructions, (5) indicated their expression during the happy pose task did not at all match the image of an actor completing the task correctly, or (6) failed attention checks.

**Missing data**

Participants with missing data on the self-reports of emotional experience will be omitted from all analyses.

**Exploratory analysis**

To examine whether any higher-order interactions exist in our primary analysis, we will fit a three-way interaction model and we will keep the random structure maximal as advised by Barr and colleagues (2013) (i.e., random intercepts for labs and subjects, random slopes for labs).

Since these analyses are exploratory, we will report the values of these interactions, but no *p*-values. However, we will report their Bayes Factors, as the interpretation of BFs does not depend upon the prior intentions of the experimenter and can be used as a way to assess the amount of evidence in favor of the alternative given the data.

**Other Comments**

The raw dataset (exported from Qualtrics), cleaned data set, data analysis scripts, and exported Qualtrics survey will be made available via the Open Science Framework after the final paper has been accepted.

Note: Given that we required a large sample to detect effects using the alpha = .05 significance threshold, we will not be able to adjust for family-wise error in this project. This is something we plan to discuss in the final manuscript.

1. We used ES estimates from the pilot as opposed to the Coles, Larsen, and Lench (2019) meta-analysis because their overall meta-analytic effect size was characterized by a large degree of heterogeneity, suggesting that it represents the weighted-average of effects from highly dissimilar facial feedback studies. Although moderator and subgroup analyses could hypothetically be used to identify relatively similar subsets of studies—e.g., ones that more closely meet our methodological specifications—these subsets of studies were also characterized by large degrees of heterogeneity (Coles, Larsen, & Lench, 2019, p. 34). Consequently, the pilot study provides more precise estimates of the effects underlying the exact methodological specifications used in the main study. Furthermore, using the observed effect sizes from the pilot study allows us to avoid making assumptions about the extent and impact of undetected questionable research practices on the meta-analytic effect size estimates. [↑](#footnote-ref-1)
2. Note: compared with the planned analysis (see below), here we removed the random slope effect for the Facial Movement Task condition and the random intercept for the ID\_Lab because these were set to zero or nearly to zero in our simulation. Therefore, modeling them in the simulated analyses raised convergence issues. [↑](#footnote-ref-2)