Jacobsen, Matilde & Hansen, Eva Sahlholdt
Character count: 18300

Has social conformity increased during COVID-19?

Github: https://github.com/eveandmad/SocKult-Exam.git

Materials and methods

Meta-analysis

Literature search

The data was aggregated manually by 1) searching for relevant keywords and authors informed by advisors on Google Scholar, 2) screening citations in bibliographies of relevant papers, and 3) searching for papers citing relevant papers on Google Scholar (see search terms and platforms in SM table 1). The search was conducted between April 27 2020 and May 11 2020.

Study selection was based on certain inclusion criteria: 1) Use of the specific paradigm for estimating social conformity as described in the following section *Procedure (The conformity task)*; 2) A healthy participant sample.

Data extraction

Data extracted from studies included healthy sample size, study and sample origin, central demographic variables (age, gender, years of education), duration of distractor task, initial and second rating, group rating, and various estimates of change in rating between round 1 and round 2 (rating change) predicted by feedback. Those included hierarchical divisions of rating changes, i.e. it appeared to be common to divide the type of group rating compared to initial rating into three feedback groups (whether the group opinion were lower, i.e. negative feedback: –3 or –2; whether the group opinion were similar, i.e. equal feedback: –1, 0 or 1; whether the group opinion were higher, i.e. positive feedback: 2 or 3). One study reported "raw conformity change", i.e. a mean beta estimate for the difference between rating 1 and 2 as predicted by feedback (see SM table 2 for estimates used in the analysis). Where certain estimates were missing, contact was established to authors in order to access individual-level data.

Study selection and specifications

In total, 11 articles and one unpublished data collection fulfilling the inclusion criteria were assessed. Of those, seven articles provided the estimates necessary for analysis, i.e. estimates of rating change according to feedback. The authors of the remaining four articles and the unpublished data collection were contacted in order to obtain the missing estimates. Of those, the full datasets were provided for one article and the unpublished data collection. Two authors responded that due to the COVID-19 situation, the data could not be accessed. One author didn't respond. Thus, three articles were excluded, resulting in a total of nine papers included in the analysis with an aggregated sample size of n = 336.

The participants included had a mean age of 23.6 years (SD = 5.72), an average length of education of 13.43 years (SD = 0.89) and were approximately equally distributed across genders with 54.8 % females. Three studies were sampled in Denmark, two in the United States, one in Russia, one in China, one in the Faroe Islands and one in the Netherlands. All studies were published between 2011 and 2019. Despite all studies applying the same paradigm for testing social conformity, different tasks were used. Thus, four studies rated trustworthiness in faces, four studies rated attractiveness in faces, and one study rated preferences for food. All ratings were conducted on a scale from 1-8 except for one study that provided a scale from 1-7. All studies applied between 150 and 222 stimuli for rating with a median of 153. Individual study demographics are included in SM table 1.

Peri-COVID-19

Data collection

The data was collected between May 7 2020 and May 13 2020. Various platforms were used to obtain participants, including Facebook and word-to-word sharing amongst friends, family and colleagues. In order to obtain sufficient power, we set a lower boundary of 50 participants initiating the experiment as we expected 1) some having technical issues in data collection, and 2) a certain drop-out rate for completing the second part of the experiment. In total, we aimed to achieve a sample of 30 participants for analysis.

Participants

Before experiment initiation, participants reported age, gender, years of education, country of upbringing, country of residence and currently medically prescribed psychiatric diagnoses. Those variables were collected in order to assess possible bias due to sample variability and increase comparability with the sample included in the meta-analysis. Overall, 66 subjects completed round 1. Of those, 42 subjects completed round 2. Subjects only completing round 1 were excluded from analysis (n = 22). Additionally, two participants were excluded due to technical issues, and two subjects were excluded due to reporting medically prescribed psychiatric diagnoses. In total, 38 participants were included in the analysis. Demography of participants included follows in table 1.

Demographic variables	**Mean (SD)**	**n**	**Percentage**
Age	26.9 (7.46)	-	-
Female : Male	-	24 : 14	63 % : 37 %
Years of education	16.1 (2.2)	-	-
Country of upbringing: Denmark : Not Denmark	-	34 : 4	89 % : 11 %
Country of residence: Denmark : Not Denmark	-	32 : 6	84 % : 16 %

Table 1*: Demographics of participants included in the study as collected from Google Survey*

Procedure

General procedure

Before initiating the experiment, participants were primed to believe the experiment concerned interhuman trust relations during the COVID-19 pandemic to ensure that participants remained naïve of the actual purpose of the experiment. Further, they reported demographic variables as described and signed up to an automatic mail-service to get notification on the second part of the study. In order to match participants correctly, each participant assigned themselves an self-chosen identification term and inserted the URL from the experiment page into the survey. To partake in the second round of the experiment, each participant received an email one hour after initiation with a link to the second round. Here, each participant again reported their identification term and inserted the new URL to enable matching.

The conformity task

In the experiment, participants were instructed to rate 80 facial images individually according to how trustworthy they perceived them to be on a 1-8 scale (1: not trustworthy at all; 8: very trustworthy). After most images, participants were reminded of their own rating and viewed the average rating of previous participants (figure 1). All group ratings were –3, –2, 0, 2 or 3 points from the initial rating of the participant. The group opinion didn't show for some images in order to enable control of regression to the mean. To complete the second round of the experiment, participants received the experiment link after one hour and were instructed to initiate it within 24 hours. The waiting period of at least one hour ensured that subjects were distracted from the task long enough to not explicitly recollect their own ratings, and the upper limit of 24 hours was set to ensure that the period was short enough to expect an effect of social conformity (Huang et al., 2014) and to increase comparability with the meta-analytic estimates. Across participants included in the analysis, there was a mean waiting period of 11.54 hours (SD = 11.42).

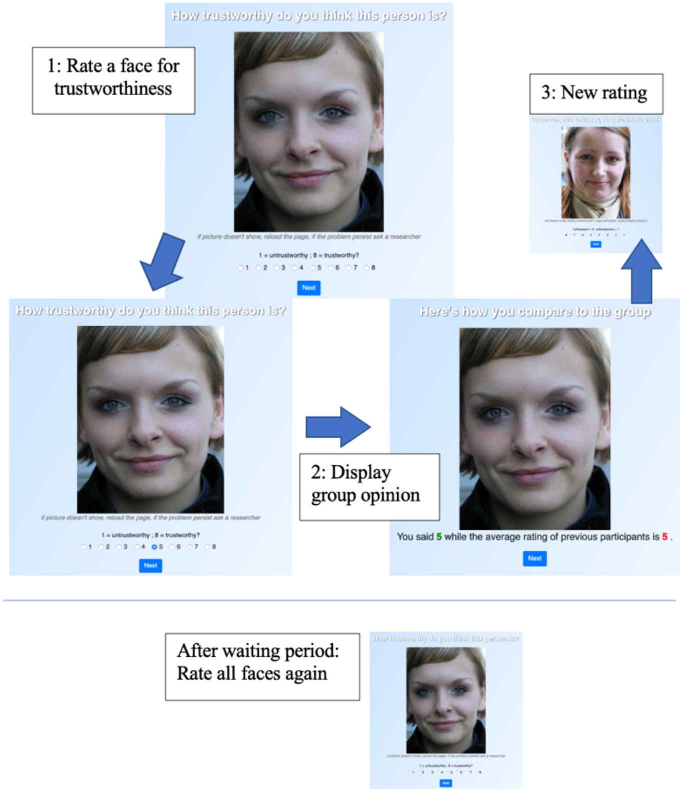

***Figure 1**: Social Conformity task paradigm as adopted from previous studies (Campbell-Meiklejohn et al., 2012; Klucharev et al., 2009; Simonsen et al., 2019). Participants rated 80 female faces for trustworthiness by choosing one number on a scale from 1-8. When proceeding to the next page the participant's rating was highlighted with green color while the group rating was highlighted with red color. Participants unexpectedly had to rate the faces again after at least 1 hour waiting, this time without getting the feedback of group rating.*

Materials and stimuli

The experiment consisted of two OTree scrips (Chen, Schonger & Wickens, 2016) obtained from a previous study (Vermillet, 2020). The scripts were modified to match an online at-home setting. The scripts for both rounds consisted of 80 images of female Caucasian faces from the shoulders up shown in a random order. Only females were used in order to reduce gender-related differences in ratings (Cloutier et al., 2008).
Group feedback was given based on a pseudo-randomization algorithm adopted from previous paradigms. The algorithm ensured that approximately one third of the group ratings matched participant rating (feedback = 0), one third of the ratings were above participant rating (feedback = +2 or +3) and one third of the ratings were below participant rating (feedback = –2 or –3).
Demographics and identification terms were collected using Google Forms. To send the automated emails with links to round 2, the online service Mailchimp was used.

Statistical methods

The entire analysis was conducted in R Studio (R Core Team, 2019).

Meta-analysis

A Bayesian Meta-Analysis was conducted using the package "brms" (Bürkner, 2017) with procedures described in the literature (Harrer et al., 2019). To estimate differences in rating change, mean estimates of change according to feedback were extracted. As rating scales were equal across studies and group feedback were, too, assigned in a similar manner, no standardization procedure was applied. For studies reporting estimates of change in relation to type of feedback (negative, equal or positive), estimates were assigned as rating change according to feedback type -2.5, 0 or 2.5. For the study reporting raw conformity score, an estimate of rating change for a feedback of 1 was modelled.
To assess the pooled effect size we used a random effects model relying on a Gaussian likelihood. The outcome variable consisted of estimated rating change and the standard error of these estimates as predicted

by feedback. A random intercept for feedback and a random slope for each article was included as we expected the effect to vary between studies (see formula 1).

Change | se(Standard Deviation of Change) ~ 1 + Feedback + (1 + Feedback | Article)
Formula 1: *Model formula for meta-analysis*

We used weakly informative priors for our model estimates. We expected rating change to lie between -1 and 1 and thus defined a normally distributed prior with a mean of 0 and variance of 0.5 ($\mu \sim N(0,0.5)$). For between study variance, we expected a positive deviation with high density close to zero while not constraining the probability of larger values. Thus, we defined it as a Half Cauchy distributed prior with a mean of 0 and variance of 0.3 ($\tau \sim HC(0,0.3)$) (Williams et al., 2018). We expected the slope for change according to feedback to lie between -0.5 and 0.5 (as observed in the literature included in the Meta-Analysis) and thus defined a normally distributed prior for the beta with a mean of 0 and variance of 0.25 ($\beta \sim N(0,0.25)$). Finally, we defined a prior for the correlation within varying effects to be an LKJ distribution with $\eta = 5$ in order to constrain the plausibility of a correlation of 1 (r = LKJ(5)).

To assess model quality we performed prior and posterior predictive checks (SM figure 1). Additionally, we ensured model convergence by looking at trace plots and ranked Markov chains (SM figure 2 and 3), and assessing that Rhat estimates were smaller than 1.05 and that effective sample sizes for both bulk and tail were larger than 200 (McElreath, 2020). Additionally, we performed a hypothesis test of the results as we expected a conformity effect larger than zero across studies. This test was based on the evidence ratio (the amount of evidence supporting the hypothesis compared to evidence against the hypothesis). An evidence ratio above 3 was interpreted as substantial evidence for the alternative hypothesis, whereas a ratio below ⅓ was interpreted as substantial evidence for the null-hypothesis.

Regression to the mean
To estimate the hidden impact of regression to the mean, the method of control in the respective analyses of the studies included in the meta-analysis was assessed. Most studies did not report proper control for this confound, suggesting inflation of the meta-analytic pooled effect size. To estimate the underlying true effect, further analysis was conducted. First, an estimate of approximate regression to the mean was computed with linear regression predicting second rating from first rating (formula 2) from one meta-analytic study where we had access to the required variables (Unpublished in-class experiment 2020). Where the model failed to converge, the varying structure was simplified.

SecondRating ~ 1 + FirstRating + (1 + FirstRating | ID)
Formula 2*: Model used to estimate regression to the mean*

The approximate estimate for regression to the mean was inserted as a fixed effect in a simulation based on the social conformity paradigm, built to reveal the true conformity effect when different levels of regression to the mean was inferred (Fusaroli, 2020). A simulated sample size of 300 participants was used to run 1000 simulations.

Comparative analysis: Peri-COVID-19

To test the hypothesis that private social conformity increased during the COVID-19 pandemic a Bayesian multilevel interaction model relying on a Gaussian likelihood was built using the package "brms" (Bürkner, 2017). To control for regression to the mean, a subset of data from the meta analysis with sufficient information provided was used for analysis (including Simonsen et al. 2019 and Unpublished in-class experiment 2020). In total, 121 subjects were included in this analysis.

Rating change as predicted by first rating, feedback and condition (whether the data was obtained pre-pandemic outbreak or peri-pandemic outbreak) was modelled, including first rating and feedback varying by subject and facial stimuli (formula 4).

Change ~ 0 + FirstRating:Condition + Feedback:Condition + (1 + FirstRating + Feedback | ID) + (1 + FirstRating + Feedback | FaceID))
Formula 3*: Model formula for comparative analysis interacting by condition*

We used weakly informative priors for our model estimates. We expected the effect of rating change according to feedback for both conditions to lie between -0.5 and 0.5 and thus defined a normally distributed prior for the beta of feedback with a mean of 0 and variance of 0.3 ($\beta \sim N(0,0.3)$). For the slope of change according to first rating we expected a slightly bigger variation, as regression to the mean likely has influence in this effect and thus defined a normally distributed prior with a mean of 0 and variance of 0.5. For the varying effects of the model we defined a similarly broad prior with a normally distributed mean of 0 and variance of 0.3 as we did not have much prior knowledge about the estimate ($\sigma \sim N(0,0.3)$). Finally, we defined a prior for the correlation within varying effects to be an LKJ distribution with $\eta = 5$ in order to constrain the plausibility of a correlation of 1 (r = LKJ(5)).

To assess model quality, prior and posterior predictive checks were performed (SM figure 5). Model convergence was ensured following the same procedures as described above (SM figure 6 and 7). A hypothesis test relying on evidence ratio was conducted to test whether the effect of feedback on social conformity was bigger in the peri-COVID-19 than pre-COVID-19 condition.

Additionally, the new data was included in the meta-analysis to assess heterogeneity from meta-analytic studies. Importantly, neither the meta-analytic nor the new data was corrected for regression to the mean. Estimates of mean rating change according to type of feedback of the new data collection is included in SM table 2 and a new forest plot (SM figure 8) was performed to assess the difference.

Results

Meta-analysis

The Bayesian meta-analysis model revealed a significant effect of feedback on rating change, (0.17, 95 % CIs: 0.11, 0.21). Additionally, evidence testing of the hypothesis that rating change as predicted by feedback is above zero proved highly evident, (ER = 5999, 95 % CIs: 0.13, 0.2), see figure 2. A forest plot (SM figure 4) was generated to assess the heterogeneity between studies. In general, low heterogeneity was found (τ = 0.03, 95% CIs: 0.00, 0.09), likely as a result of highly similar methods and population samples.

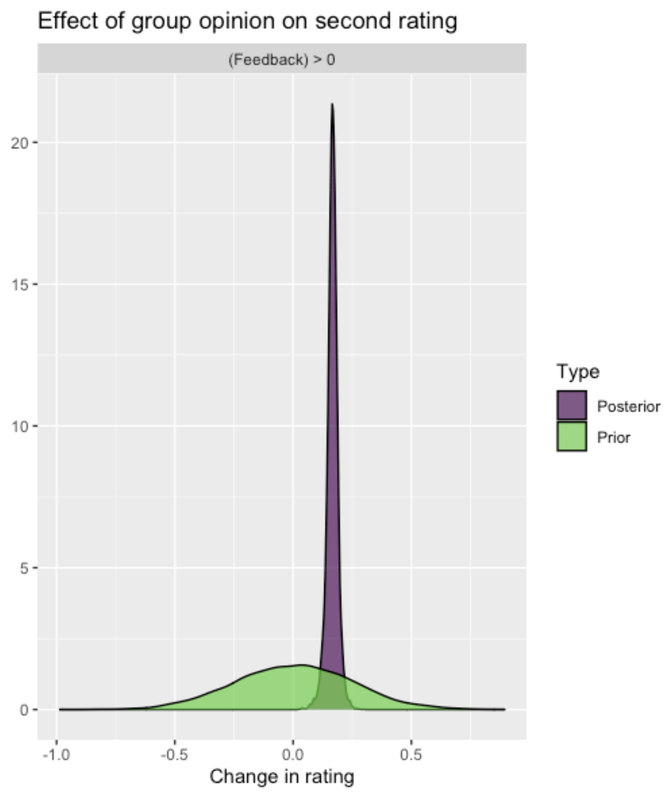

Figure 2*: Evidence ratio indicating a great deal of certainty that there is a rating change > 0*

Regression to the mean

The regression model (formula 2) revealed a significant effect of first rating as a predictor of second rating ($\beta = 0.666$, SE = 0.033, t = 19.95, $p < 0.001$). Thus, a fixed conformity estimate of 0.66 was included in the simulation. With this estimate, a fixed true conformity effect of -0.17 was necessary to obtain an estimated conformity effect of 0.17 as found in the meta-analysis (see SM table 4 for an overview of simulated values). Inversely, when controlling for regression to the mean (following SM formula 1) on the subset of data used for the regression analysis, a true conformity effect of 0.02 was obtained (ER > 3, 95% CI's: 0,0.04), identical with controlled results from other studies (Simonsen et al., 2019). Based on the simulation, this effect should show an estimated conformity of 0.52. As this is not the case in the meta-analysis as visualized in the forest plot (SM figure 4, Unpublished in-class experiment 2020), a decreased regression to the mean estimate, other confounds or masking effects might also play a role in the pooled meta-analytical effect size.

Comparative analysis: Peri-COVID-19

The Bayesian multilevel model revealed similarly small main effect of feedback for both conditions (pre-COVID-19: $\beta = 0.03$, 95% CI's = 0.01,0.05, peri-COVID-19: $\beta = 0.03$, 95% CI's = 0.00, 0.06) as well as similarly negative effects of initial ratings for both conditions (peri-COVID-19: $\beta = -0.19$, 95% CI's = -0.26, -0.13, pre-COVID-19: $\beta = -0.20$, 95% CI's = -0.26, -0.14) on change of trustworthiness ratings. When testing the hypothesis that a conformity effect above zero existed in both conditions, an evidence ratio above 3 provided substantial evidence of the alternative hypothesis. For the hypothesis that the effects of feedback has increased peri-COVID-19 compared to the pre-COVID-19 effect, no substantial evidence was found ($\beta = 0$, 95% CI's = -0.02, 0.03, ER = 1.63). See full details in table 6.

When not correcting for regression to the mean, means and variance of raw change according to feedback of the peri-COVID-data was highly compatible with those obtained for the pre-COVID-data included in the comparative model (figure 3).

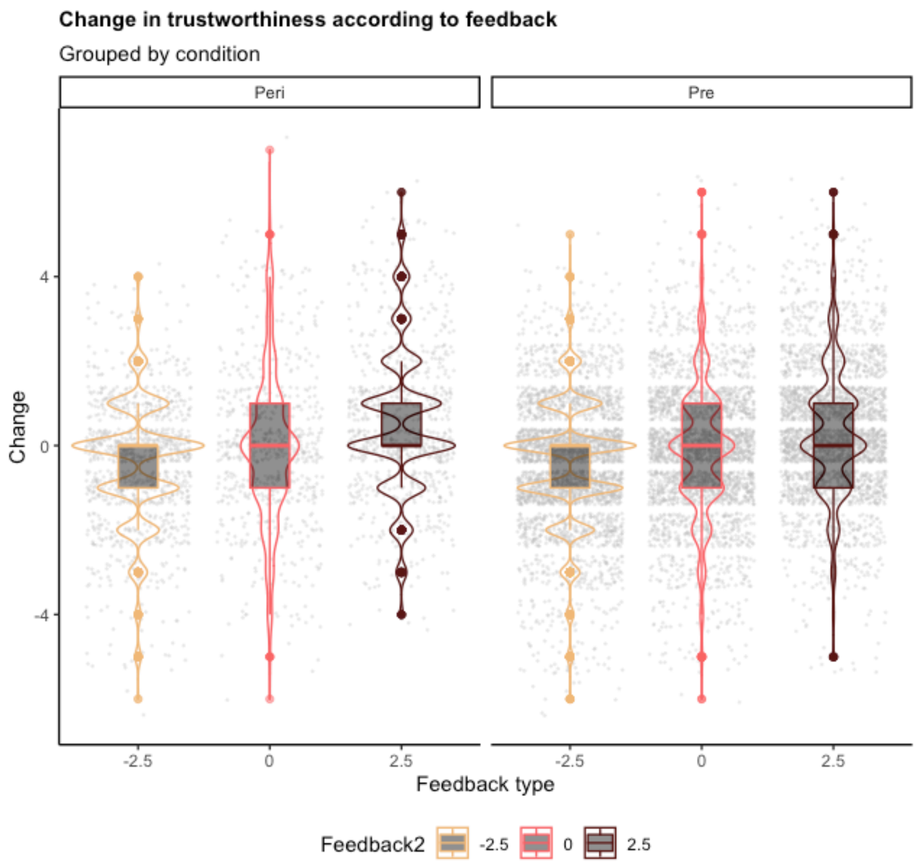

***Figure 3**: A violin distribution of the effects of change divided into different groups of feedback.*

References

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Campbell-Meiklejohn, D. K., Simonsen, A., Jensen, M., Wohlert, V., Gjerløff, T., Scheel-Kruger, J., Møller, A., Frith, C. D., & Roepstorff, A. (2012). Modulation of Social Influence by Methylphenidate. *Neuropsychopharmacology*, *37*(6), 1517–1525. https://doi.org/10.1038/npp.2011.337

Cloutier, J., Heatherton, T. F., Whalen, P. J., & Kelley, W. M. (2008). Are Attractive People Rewarding? Sex Differences in the Neural Substrates of Facial Attractiveness. *Journal of Cognitive Neuroscience*, *20*(6), 941–951. https://doi.org/10.1162/jocn.2008.20062

Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2019). *Doing Meta-Analysis in R: A Hands-on Guide*. PROTECT Lab. https://bookdown.org/MathiasHarrer/Doing_Meta_Analysis_in_R/

Huang, Y., Kendrick, K. M., & Yu, R. (2014). Conformity to the Opinions of Other People Lasts for No More Than 3 Days. *Psychological Science*, *25*(7), 1388–1393. https://doi.org/10.1177/0956797614532104

Klucharev, V., Hytönen, K., Rijpkema, M., Smidts, A., & Fernández, G. (2009). Reinforcement Learning Signal Predicts Social Conformity. *Neuron*, *61*(1), 140–151. https://doi.org/10.1016/j.neuron.2008.11.027

McElreath, R. (2020). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan* (2nd ed.). CRC Press.

Simonsen, A., Fusaroli, R., Skewes, J. C., Roepstorff, A., Mors, O., Bliksted, V., & Campbell-Meiklejohn, D. (2019). Socially Learned Attitude Change is not reduced in Medicated Patients with Schizophrenia. *Scientific Reports*, *9*(1), 992. https://doi.org/10.1038/s41598-018-37250-x

Williams, D. R., Rast, P., & Bürkner, P.-C. (2018). *Bayesian Meta-Analysis with Weakly Informative Prior Distributions* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/7tbrm

Paul-Christian Bürkner (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. Journal of Statistical Software, 80(1), 1-28. doi:10.18637/jss.v080.i01

Chen, D.L., Schonger, M., Wickens, C., 2016. oTree - An open-source platform for laboratory, online and field experiments. Journal of Behavioral and Experimental Finance, vol 9: 88-97

Vermillet, 2020: Did not respond in reference to assess citation. Contact: arnault@cc.au.dk

Fusaroli, 2020. Code for simulation adapted from script.

Supplementary Materials

Meta-analysis

Paper	Simonsen et al. (2014)	Campbell-Meiklejohn et al. (2012)	Zhao et al. (2016)	Klucharev et al. (2011)	Simonsen et al. (2019)	Unpublished in-class experiment (2020)	Shestakova et al (2012)	Nook and Zaki (2015)	Zaki et al. (2011)
Title	Serotoninergic effects on judgments and social learning of trustworthiness	Modulation of Social Influence by Methylphenidate	Investigating the Genetic Basis of Social Conformity: The Role of the Dopamine Receptor 3 (DRD3) Gene	Downregulation of the Posterior Medial Frontal Cortex Prevents Social Conformity	Socially Learned Attitude Change is not reduced in Medicated patients with schizophrenia	-	Electrophysiological precursors of social conformity	Social Norms Shift Behavioral and Neural Responses to Foods	Social Influence Modulates the Neural Computation of Value
Authors	A. Simonsen & J. Scheel-Krüger & M. Jensen & A. Roepstorff & A. Møller & C. D. Frith & D. Campbell-Meiklejohn	D. Campbell-Meiklejohn & A. Simonsen & M. Jensen & V. Wohlert & T. Gjerløff & J. Scheel-Kruger & A. Møller & C. Frith & A. Roepstorff	C. Zhao, J. Liuf, P. Gongh J. Hua, X. Zhoua	V. Klucharev, M. A. M. Munneke, A. Smidts, G. Fernández	A. Simonsen, R. Fusaroli, J.C. Skewes, A. Roepstorff, O. Mors, V. Bliksted, D. Campbell-Meiklejohn	-	A. Shestakova, J. Rieskamp, S. Tugin, A. Ossadtchi, J. Krutitskaya, V. Klucharev	E.C. Nook & J. Zaki	J. Zaki, J. Schirmer, and J.P. Mitchell
Search specification	Google Scholar	Followed citation from Simonsen et al 2019	Google Scholar	Google Scholar	Professor advice	Professor advice	Google Scholar	Followed citation from Wu et al. (2016)	Followed citation from Wu et al. (2016)
Search term	“social conformity task face trustworthiness”	-	“facial trustworthiness”	“group opinion social conformity”	-	-	“Klucharev”	-	-
Area	Denmark	Denmark	China	The Netherlands	Faroe Islands	Denmark	Russia	USA	USA
Sample size (healthy)	20	19	149	15	39	44	15	21	14
Female	20	19	60	15	12	25	15	18	0
Age: Mean (sd)	23.5 (2.5)	23 (2.7)	22.5 -	21.1 -	39.2 (10.6)	22 -	19.9 -	20.1 -	21.8 -
Years of education: Mean (sd)	13.8 (1.9)	13.8 (1.8)	-	-	14.2 (3.1)	-	-	-	-
Specification of education	-	-	University students	-	-	Cognitive Science students	Students	Undergraduates from Stanford	-

Task	Trustworthi-ness	Trustworthi-ness	Attractive-ness	Attractive-ness	Trustwor-thiness	Trustwor-thiness	Attractive-ness	Food rating	Attractive -ness
Rating scale	1-8	1-8	1-8	1-8	1-8	1-8	1-8	1-8	1-7
Number of stimuli	153	153	120	222	153	153	222	150	180
Distraction time (min)	30	30	30	30	60	15	30	5	30

***Table 1**: Individual demographic variables of studies included in Meta-Analysis*

Paper	**Change Low: Mean (SD)**	**Change Same: Mean (SD)**	**Change High: Mean (SD)**	**Raw Change: Mean (SD)**	**First Rating: Mean (SD)**	**Second Rating: Mean (SD)**
Simonsen et al. (2014)	-0.39 (0.05)	-	0.46 (0.05)	-	4.87 (0.10)	4.94 (0.14)
Campbell-Meiklejohn et al. (2012)	-0.31 (0.11)	-	0.50 (0.10)	-	4.88 (0.11)	4.93 (0.15)
Klucharev et al. (2011)	-0.40 (0.25)	-0.05 (0.28)	0.28 (0.24)	-	-	-
Simonsen et al. (2019)	-0.41 (1.53)	-0.08 (1.59)	0.46 (1.59)	-	5.0 (1.6)	5.0 (1.6)
Unpublished in-class experiment (2020)	-0.54 (1.46)	-0.22 (1.49)	0.14 (1.41)	-	4.63 (1.76)	4.40 (1.81)
Shestakova et al (2012)	-0.74 (0.42)	-0.18 (0.27)	0.43 (0.32)	-	-	-
Nook and Zaki (2015)	-0.11 -	0.03 -	0.07 -	-	-	-
Zaki et al. (2011)	-0.33 -	-	0.10 -	-	-	-
Zhao et al. (2016)	-	-	-	0.26 (0.095)	-	-

***Table 2**: Individual estimates from studies included in Meta-Analysis*

Estimate type	**Estimate**	**Est.Error**	**Q2.5**	**Q97.5**
b_Intercept	0.031	0.043	-0.062	0.115
b_Feedback	0.166	0.024	0.113	0.210
sd_Article__Intercept	0.046	0.043	0.002	0.157
sd_Article__Feedback	0.027	0.025	0.001	0.094

***Table 3**: Summary of estimates from meta-analysis*

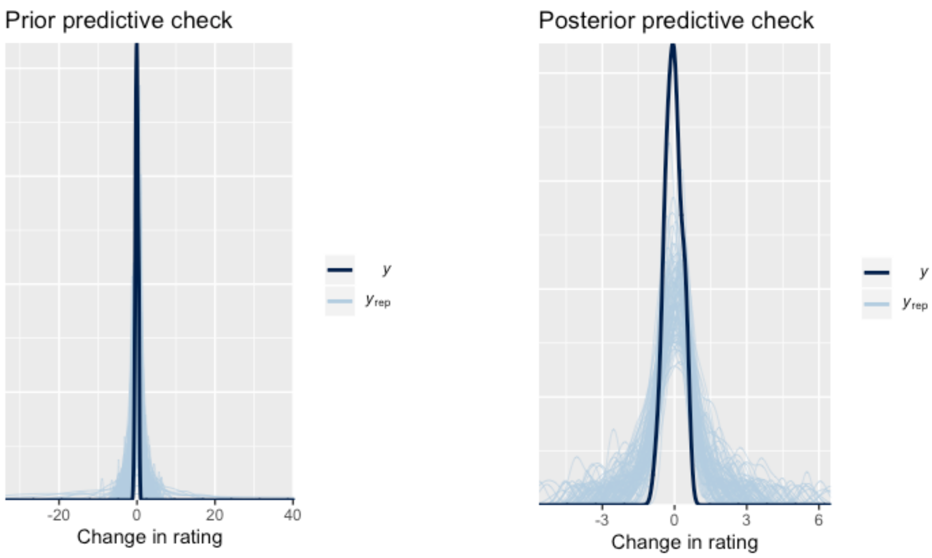

***Figure 1**: Prior and posterior predictive checks of meta-analytical model*

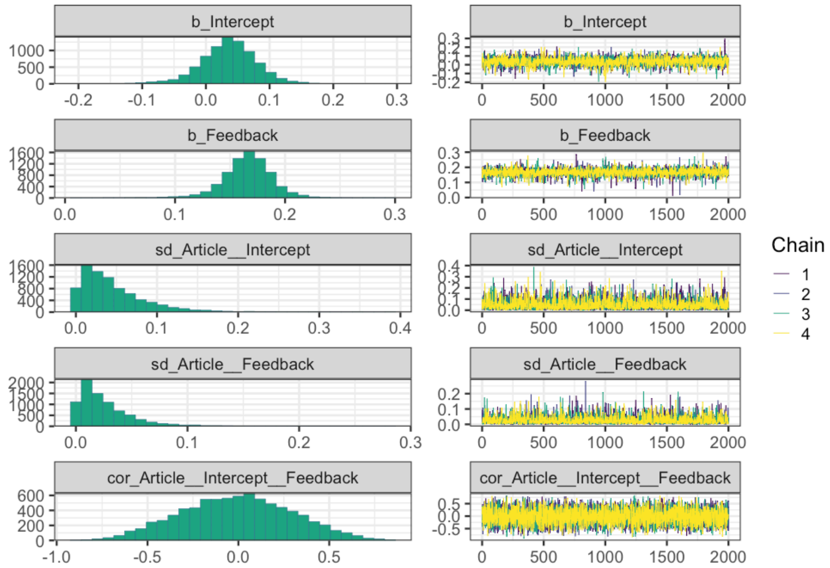

Figure 2: *Model assessment I: Histogram and trace plots of meta-analytical model estimates*

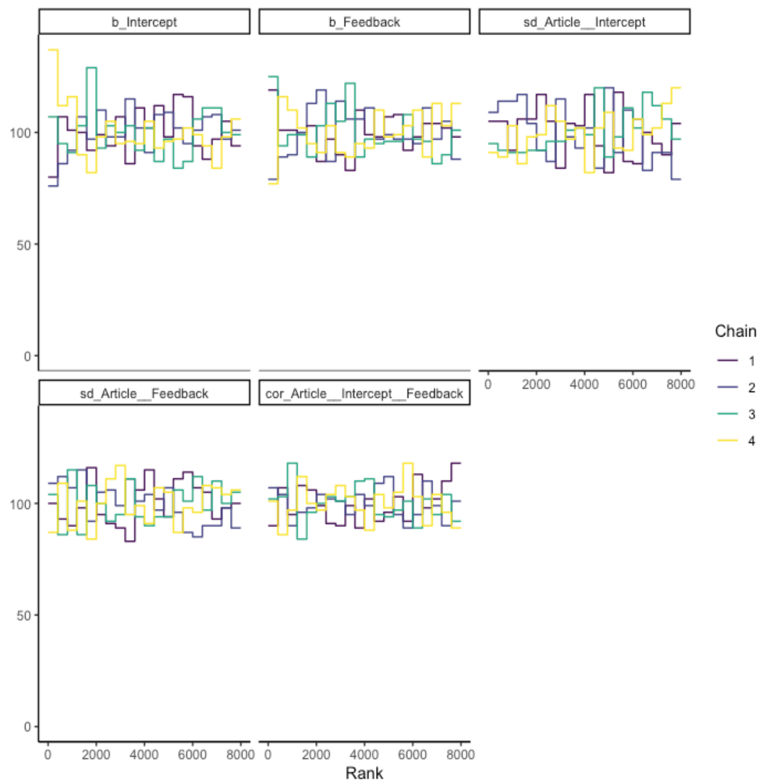

Figure 3: *Model assessment II: Ranked Markov chains of meta-analytical model convergence*

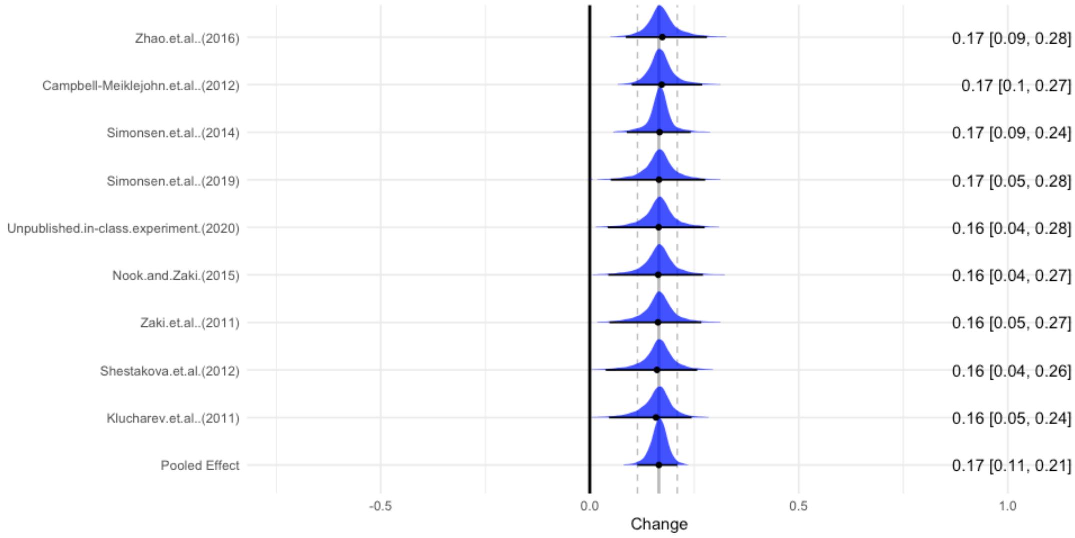

Figure 4: *Forest plot of meta-analytical effect sizes*

Regression to the mean

True regression to the mean	True conformity	Estimated conformity
0.66	-0.20	0.09
0.66	-0.17	0.17
0.66	-0.10	0.28
0.66	-0.05	0.47
0.66	0.02	0.52
0.52	-0.20	0.26
0.52	-0.17	0.35
0.52	-0.10	0.40
0.52	-0.05	0.44
0.52	0.02	0.55

Table 4: *Estimates as obtained from a simulation assessing causal mechanisms of the social conformity paradigm. "True regression to the mean" estimates were obtained through a regression analysis of real data, whereas the "True conformity" scores were changed manually in the simulation to get the estimated conformity effect as explained by the simulation (see provided code for more details on the simulation).*

Comparative analysis: Peri-COVID-19

Across participants (n = 38)	Mean (SD)	Number of ratings
First rating	4.90 (1.83)	2971
Second rating	4.89 (1.79)	2971
Change in rating with low group rating (feedback = -2, -3)	-0.412 (1.49)	1090
Change in rating with same group rating (feedback = 0)	-0.076 (1.78)	668
Change in rating with high group rating (feedback = +2, +3)	0.52 (1.74)	840

Table 5: *Means of ratings and means of change according to feedback with count estimates*

Estimate type	Estimate	Est.Error	Q2.5	Q97.5
b_FirstRating:ConditionPeri	-0.194	0.033	-0.262	-0.130
b_FirstRating:ConditionPre	-0.199	0.030	-0.260	-0.141
b_ConditionPeri:Feedback	0.031	0.015	0.003	0.061
b_ConditionPre:Feedback	0.027	0.010	0.009	0.048
sd_FaceID Intercept	0.540	0.059	0.431	0.661
sd_FaceID FirstRating	0.072	0.012	0.049	0.096
sd_FaceID Feedback	0.016	0.010	0.001	0.037
sd_ID Intercept	2.162	0.109	1.953	2.377
sd_ID FirstRating	0.320	0.027	0.268	0.374
sd_ID Feedback	0.013	0.009	0.001	0.034

***Table 6**: Summary of estimates for interaction model*

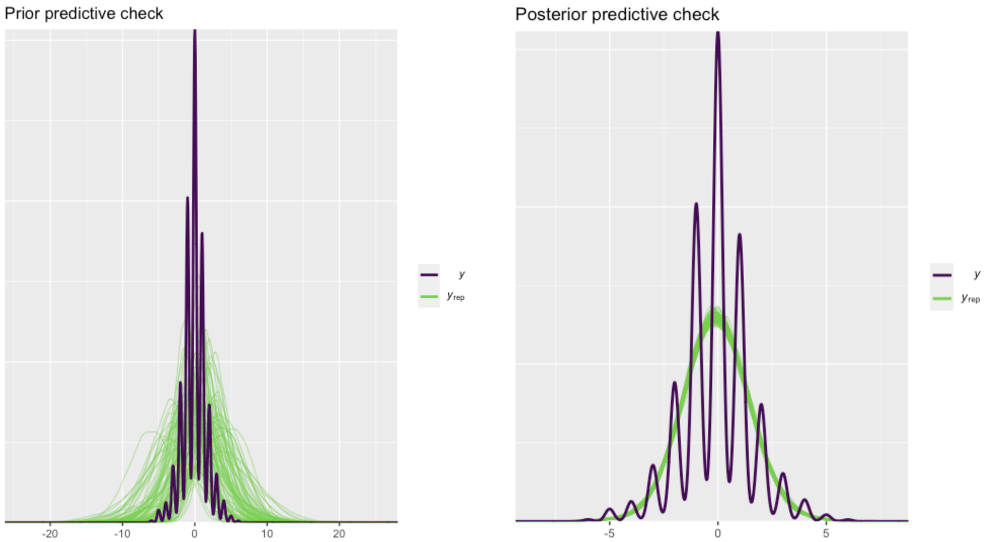

***Figure 5**: Prior and posterior predictive check of model interacting with condition*

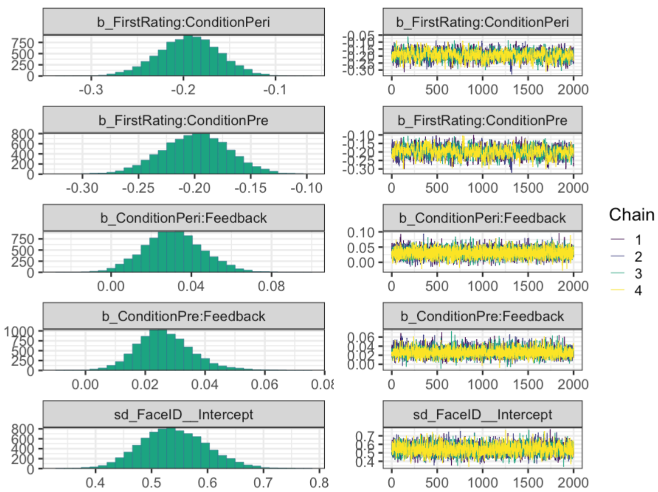

***Figure 6**: Model assessment 1: Histogram and trace plots for estimates of interaction model*

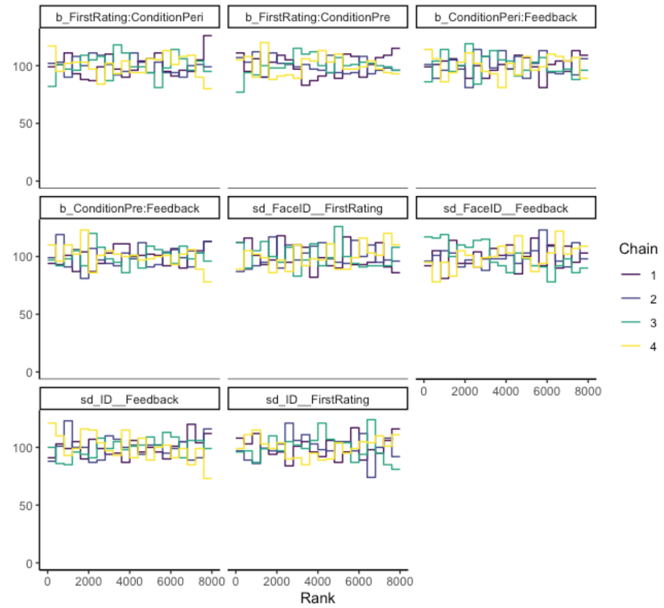

***Figure** 7: Model assessment II: Ranked Markov chains of model convergence of interaction model*

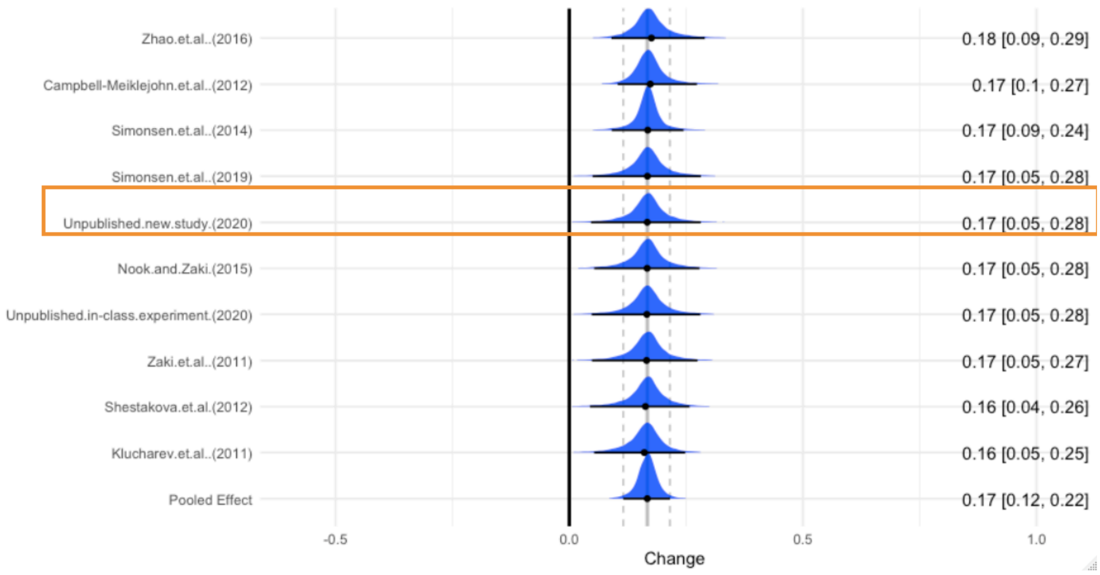

***Figure 8**: Forest plot including own study*

Individual analysis: Peri-COVID-19

A Bayesian multilevel model was built using the package "brms" (Bürkner, 2017) based on procedures described in the literature (McElreath, 2020). To estimate differences in change from rating 1 to 2, a subset of the data-set collected in the new study was assessed according to estimates of change where participants received a low feedback (-2 or -3) or high feedback (2 or 3) reducing the amount of rated images to 1930. This was done to avoid that the effect potentially would be drawn closer to zero due to an invisible conformity effect in the feedback type where participants received the same group rating as their own (e.g. if first rating and belonging group rating was low for a particular case, a similar second rating that would have been high if not for conformity was to be observed as regression to the mean rather than conformity and thus shrinking the effect rather than increasing it).

The following random effects model with a random slopes for each participant and stimuli was used:

Change ~ 1 + FirstRating + Feedback + (1 + FirstRating + Feedback | ID) + (1 + FirstRating + Feedback | FaceID))

Formula 1*: model formula for own study of peri-COVID-19 effect of private social conformity*

We used weakly informative priors for our model estimates. We expected rating change to lie between -2 and 2 and thus defined a normally distributed prior with a mean of 0 and variance of 1 ($\mu \sim N(0,1)$). We expected the slope for change according to feedback to lie between -0.5 and 0.5 (as observed in the literature included in the Meta-Analysis) and thus defined a normally distributed prior for the betas with a mean of 0 and variance of 0.25 ($\beta \sim N(0,0.25)$). For the varying effects of the model we defined a similarly broad prior with a normally distributed mean of 0 and variance of 0.3 as we did not have much prior knowledge about the estimate ($\sigma \sim N(0,0.3)$). Finally, we defined a prior for the correlation within varying effects to be an LKJ distribution with $\eta = 5$ in order to constrain the plausibility of a correlation of 1 (r = LKJ(5)).

To assess model quality we performed prior and posterior predictive checks (see SM figure 9). Additionally, we ensured model convergence by looking at trace plots, unranked and ranked (SM figure 10 and 11) and assessing that Rhat estimates were smaller than 1.01 and effective sample sizes for both bulk and tail were smaller than 200 (McElreath, 2020). Individual level estimates of participants were explored to ensure homogeneity within the data (SM figure 12). Additionally, we tested whether the effect was larger than chance (Feedback > 0) based on the evidence ratio.
To assess that the experiment worked as expected means and standard deviations of ratings were summarized across participants revealing a highly similar mean from first to second rating. Means and standard deviations of the raw change in rating were divided into three categories according to type of feedback (low, same, high) and summarized, revealing a general tendency to go down in rating when feedback was low and go up in rating when feedback was high. This tendency is most likely to be an effect of regression to the mean, though. The numbers of stimuli with certain feedback type as generated by the algorithm revealed a slight overweight in negative feedback, but still balanced enough to deem the automated feedback generation successful (see table 4).
As expected a small effect of private social conformity was observed based on the main effect of feedback after controlling for regression to the mean (β = 0.02, 95% CI's = 0 - 0.05, ER = 15.23). See full details of the model in table 7.

Estimate type	Estimate	Est.Error	Q2.5	Q97.5
b_Intercept	2.985	0.197	2.605	3.378
b_FirstRating	-0.615	0.032	-0.678	0.552
b_Feedback	0.019	0.014	-0.008	0.046
sd_FaceID__Intercept	0.440	0.059	0.325	0.561
sd_FaceID__FirstRating	0.027	0.018	0.001	0.066
sd_FaceID__Feedback	0.021	0.015	0.001	0.055
sd_ID__Intercept	0.955	0.105	0.764	1.174
sd_ID__FirstRating	0.150	0.023	0.109	0.198
sd ID Feedback	0.021	0.015	0.001	0.055

***Table** 7: Summary of estimates from individual analysis of new study*

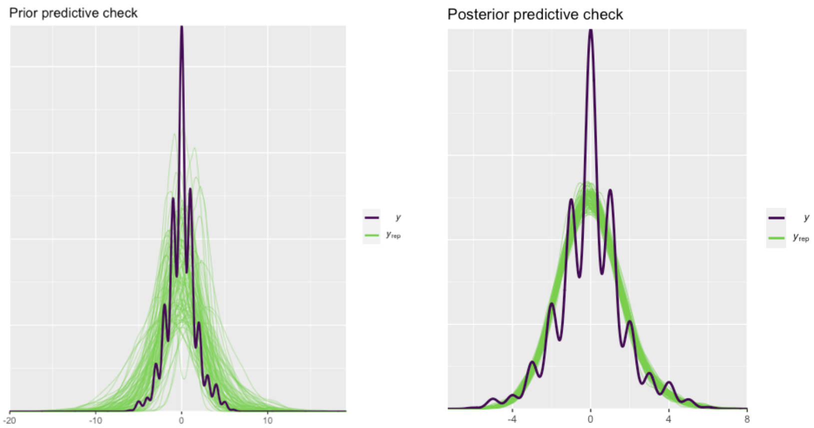

Figure 9: *Prior and posterior predictive check of peri-COVID19 study on high and low group feedback*

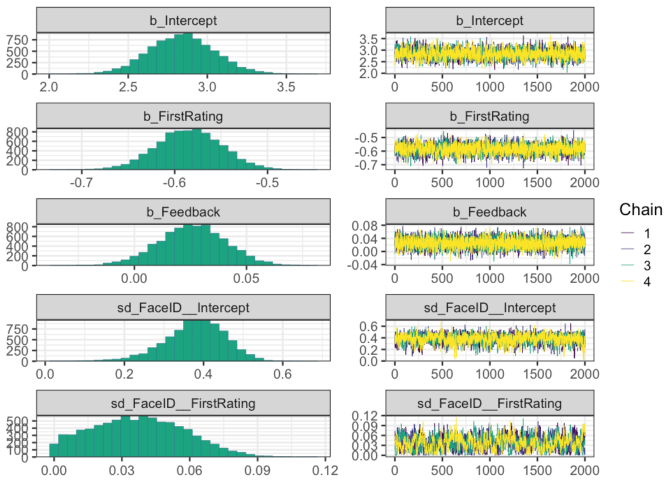

Figure 10: *Model assessment 1: Histogram and trace plots for estimates of peri-COVID19 study on high and low group feedback*

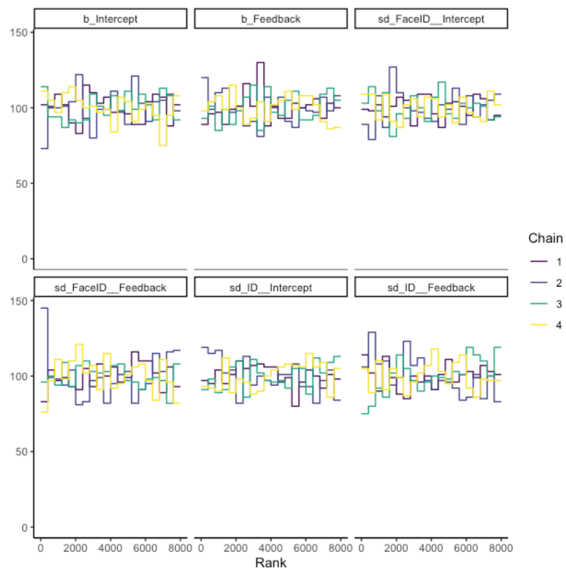

Figure 11: *Model assessment II: Ranked plots of model convergence of peri-COVID19 study on high and low group feedback*

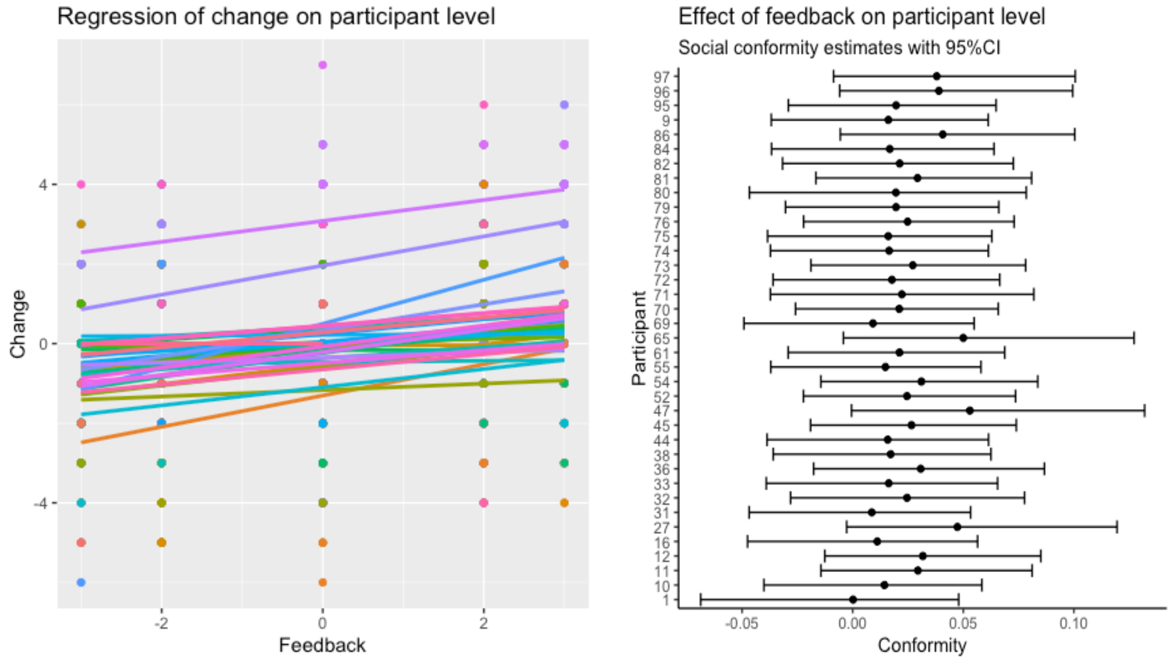

Figure 12*: Assessment of individual level estimates from peri-COVID-19 study*