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

Nicholas A. Coles [1] ✉, David S. March [2], Fernando Marmolejo-Ramos [3], Jeff T. Larsen[4], Nwadiogo C. Arinze [5], Izuchukwu L. G. Ndukaihe [5], Megan L. Willis [6], Francesco Foroni [6], Niv Reggev [7,8], Aviv Mokady [7], Patrick S. Forscher [9], John F. Hunter [10], Gwenaël Kaminski [11], Elif Yüvrük[12], Aycan Kapucu[12], Tamás Nagy [13], Nandor Hajdu[13], Julian Tejada [14], Raquel M. K. Freitag [15], Danilo Zambrano [16], Bidisha Som [17], Balazs Aczel [13], Krystian Barzykowski [18], Sylwia Adamus [18], Katarzyna Filip [18], Yuki Yamada [19], Ayumi Ikeda [20], Daniel L. Eaves [21,22], Carmel A. Levitan [23], Sydney Leiweke[23], Michal Parzuchowski [24], Natalie Butcher [25], Gerit Pfuhl [26], Dana M. Basnight-Brown [27], José A. Hinojosa [28,29], Pedro R. Montoro [30], Lady G. Javela D [31], Kevin Vezirian[32], Hans IJzerman [32,33], Natalia Trujillo [34], Sarah D. Pressman[35], Pascal M. Gygax [36], Asil A. Özdoğru [37], Susana Ruiz-Fernandez [38,39], Phoebe C. Ellsworth [40], Lowell Gaertner[4], Fritz Strack[41], Marco Marozzi [42] and Marco Tullio Liuzza [43]

Following theories of emotional embodiment, the facial feedback hypothesis suggests that individuals' subjective experiences of emotion are influenced by their facial expressions. However, evidence for this hypothesis has been mixed. We thus formed a global adversarial collaboration and carried out a preregistered, multicentre study designed to specify and test the conditions that should most reliably produce facial feedback effects. Data from $n = 3{,}878$ participants spanning 19 countries indicated that a facial mimicry and voluntary facial action task could both amplify and initiate feelings of happiness. However, evidence of facial feedback effects was less conclusive when facial feedback was manipulated unobtrusively via a pen-in-mouth task.

The facial feedback hypothesis suggests that individuals' emotional experiences are influenced by their facial expressions. For example, smiling should typically make individuals feel happier, and frowning should make them feel sadder. Researchers suggest that these effects emerge because facial expressions provide sensorimotor feedback that contributes to the sensation of an emotion[1,2], serves as a cue that individuals use to make sense of ongoing emotional feelings[3,4], influences other emotion-related bodily responses[5,6] and/or influences the processing of emotional stimuli[7,8]. This facial feedback hypothesis is notable because it supports broader theories that contend emotional experience is influenced by feedback from the peripheral nervous system[9–11], as opposed to experience and bodily sensations being independent components of an emotion response[12–14]. Furthermore, this hypothesis supports claims that facial feedback interventions—for example, smiling more or frowning less—can help manage distress[15,16], improve well-being[17,18] and reduce depression[19–39].

Recently, a collaboration involving 17 independent teams consistently failed to replicate a seminal demonstration of facial feedback effects[40]. In the original study, the participants viewed humorous cartoons while holding a pen in their mouth in a manner that either elicited smiling (pen held in teeth) or prevented smiling (pen held by lips)[41]. Consistent with the facial feedback hypothesis, smiling participants

reported feeling more amused by the cartoons. This finding was influential because previous studies often explicitly instructed participants to pose a facial expression, raising concerns about demand characteristics[42–44]. Furthermore, theorists disagreed about whether these effects could occur outside of awareness[45–47]. Because the participants in this pen-in-mouth study were presumably unaware that they were smiling, the authors concluded that facial feedback effects were not driven by demand characteristics and could occur outside of awareness.

What implications does the failure to replicate have for the facial feedback hypothesis? One possibility is that the facial feedback hypothesis is false. However, this conclusion is unwarranted because this direct replication was limited to a specific test of the facial feedback hypothesis. Indeed, the replicators stated that their findings "do not invalidate the more general facial feedback hypothesis"[40]. Similarly, while arguing that the pen-in-mouth effect is unreliable, some researchers conceded that "other paradigms may produce replicable results"[48].

A second possibility is that both the facial feedback hypothesis and the original pen-in-mouth effect are true. If this is the case, researchers must determine why others were unable to replicate the pen-in-mouth effect. One suggestion is that the replicators did not perform a true direct replication because they deviated from the original study by overtly recording the participants (per the advice of an expert reviewer)[49]. According to this explanation, awareness of video recording may induce a self-focus that interferes with participants' internal experiences and emotional behaviour[49,50].

A third possibility is that the facial feedback hypothesis is true, but not in the context examined in the original pen-in-mouth study. Perhaps facial feedback effects occur only when participants are aware that they are posing a facial expression[45,46], a mechanism that the pen-in-mouth task was designed to eliminate. Alternatively, perhaps the pen-in-mouth task is not a reliable manipulation of facial feedback. Some theorists predict that facial feedback effects will emerge only when facial movement patterns resemble a prototypical emotional facial expression[5,51–55], and previous research indicates that the pen-in-mouth task does not reliably produce prototypical expressions of happiness[56]. Last, perhaps facial feedback influences only certain types of emotional experiences. Some researchers distinguish between self-focused and world-focused emotional experiences, and facial feedback theories have traditionally emphasized self-focused emotional experience[57,58]. However, in the original pen-in-mouth study, the participants were asked how amused a series of cartoons made them feel, which may have induced a world-focused emotional experience.

Amid the uncertainty created by the failure to replicate, a meta-analysis was performed on 286 effect sizes from 137 studies testing the effects of various facial feedback manipulations on emotional experience[59]. The results indicated that facial feedback has a small but highly varied effect on emotional experience. Notably, this effect could not be explained by publication bias. Published and unpublished studies yielded effects of similar magnitude, analyses failed to uncover significant evidence of publication bias and bias-corrected overall effect size estimates were significant. However, this meta-analysis did not explain why facial feedback effects were not observed in the pen-in-mouth replication study. Inconsistent with preliminary evidence that video-recording awareness interferes with facial feedback effects[50], the meta-analysis revealed significant facial feedback effects regardless of whether studies used overt video recording[59].

Although the meta-analysis suggests that the facial feedback hypothesis is valid, there are at least three limitations that could undermine this conclusion. First, since publication bias analyses often have low power[60–62], it is possible that seemingly robust facial feedback effects are driven by studies with undetected questionable research practices. Second, it is possible that the overall effect size estimates in this literature are driven by low-quality studies[63]. Third, even relatively similar subsets of facial feedback studies varied beyond what would be expected from sampling error alone, meaning that moderator analyses

had lower power and potentially contained unidentified confounds. Consequently, the meta-analysis could not reliably identify moderators that may help explain why some researchers fail to observe facial feedback effects.

Both the failure to replicate the pen-in-mouth study and the meta-analysis have a unique set of limitations that make it difficult to resolve the debate regarding whether the facial feedback hypothesis is valid. We therefore came together to form the Many Smiles Collaboration. We are an international group of researchers—some advocates of the facial feedback hypothesis, some critics and some without strong beliefs—who collaborated to (1) specify our beliefs regarding when facial feedback effects, if real, should most reliably emerge; (2) determine the best way(s) to test those beliefs; and (3) use this information to design and execute an international multi-lab experiment.

We agreed that one of the simplest necessary conditions for facial feedback effects to emerge is that participants pose an emotional facial expression and subsequently self-report the degree to which they are experiencing the associated emotional state. Therefore, our main research question was whether participants would report feeling happier when posing happy versus neutral expressions. On the basis of outstanding theoretical disagreements in the facial feedback literature, we also questioned (1) whether happy facial poses only influence feelings of happiness if they resemble a natural expression of happiness, (2) whether facial poses can initiate emotional experience in otherwise neutral scenarios or only amplify ongoing emotional experiences, and (3) whether facial feedback effects are eliminated when controlling for awareness of the experimental hypothesis. These disagreements ultimately informed the final experimental design: a 2 (Pose: happy or neutral) × 3 (Facial Movement Task: facial mimicry, voluntary facial action or pen-in-mouth) × 2 (Stimuli Presence: present or absent) design, with Pose manipulated within participants and Facial Movement Task and Stimuli Presence manipulated between participants (Supplementary Fig. 1).

To provide an easy-to-follow task that would produce more prototypical facial expressions, we used a facial mimicry paradigm, wherein the participants were asked to mimic images of actors displaying prototypical expressions of happiness[64]. To produce less prototypical facial expressions, some participants completed the voluntary facial action task[65], wherein they were asked to move some—but not all—facial muscles associated with prototypical expressions of happiness[56]. We also added the pen-in-mouth task after Stage 1 reviewer feedback, wherein the participants held a pen in their mouth in a manner that either elicited smiling (pen held in teeth) or prevented smiling (pen held by lips)[41]. While engaging in the facial feedback tasks, half of the participants viewed a series of positive images[57,58].

We hypothesized that participants would report experiencing more happiness when posing happy versus neutral facial expressions. Furthermore, we hypothesized that the magnitude of this effect would be similar across tasks that produce less (the voluntary facial action and pen-in-mouth tasks) versus more (the mimicry task) prototypical expressions of happiness. We also expected that facial feedback effects would be smaller in the absence than in the presence of positive stimuli. Last, we expected to observe facial feedback effects even when limiting our analyses to participants who were completely unaware of our hypothesis. Two pilot studies (n = 206; Supplementary Information) confirmed these predictions. A third pilot study conducted after initial Stage 1 acceptance (n = 119; Supplementary Information) provided preliminary evidence in favour of some—but not all—of our predictions. These pilot results led to minor refinements to the methodology but did not change our final set of predictions. Our research questions and hypotheses are summarized in Table 1.

Results

We conducted all analyses using R (v.4.1.2)[66]. For the frequentist analyses, we fit mixed-effect models using the lme4 package[67]. Some of these

Table 1 | Research questions and associated hypotheses

Research question	Hypothesis
Does posing happy versus neutral expressions cause people to feel happier?	Participants will report experiencing more happiness when posing happy versus neutral facial expressions.
Do happy facial poses only influence feelings of happiness if they resemble natural expressions of happiness?	The difference in self-reported happiness when posing happy versus neutral facial expressions will not be larger during tasks that produce more natural expressions of happiness.
Can facial poses initiate emotional experiences in otherwise neutral scenarios, or can they only modulate ongoing emotional experiences?	The difference in self-reported happiness when posing happy versus neutral facial expressions will be significant in both the presence and absence of emotional stimuli. This difference will be larger in the presence of emotional stimuli.
Are facial feedback effects eliminated when controlling for awareness of the experimental hypothesis?	Participants who are judged to be completely unaware of the purpose of the experiment will report experiencing more happiness when posing happy versus neutral facial expressions.

models contained random slopes and thus have smaller degrees of freedom. For tests of main effects, simple effects and interactions, we used the lmerTest package to derive analysis-of-variance-like F values with Satterthwaite degrees of freedom[68]. When we observed higher-order interactions, we used the emmeans package to decompose them using simple effect tests and pairwise contrasts[69]. We used model-derived mean difference estimates as our effect size of interest. However, we also report semi-standardized mean difference estimates, wherein the model-derived mean difference is divided by the total range of the measured dependent variable.

For the Bayesian re-analysis of the hypotheses in Table 1, we used the BayesFactor package to fit models using medium Cauchy priors (r scale, 1/2) on the alternative hypotheses and the default Markov chain Monte Carlo settings[70]. We also performed sensitivity analyses with wide (r scale, √2/2) and ultrawide (r scale, 1) priors, and we thus report a range of Bayes factors (BFs). For tests of main effects, interactions and simple effects, we computed BFs by comparing models containing versus excluding the terms representing the tested effect.

Participants

We made two minor deviations from the preregistered sampling plan. First, due to constraints created by COVID-19, no research group collected data in person. We were thus unable to test whether our pattern of results differed by in-person versus online data collection. Second, we had 80 fewer participants than we initially planned for our primary analyses.

Depending on the research site, the participants completed the study on a completely volunteer basis, for partial course credit, for extra credit, for entrance into a lottery (for example, for a gift box), for a prize (for example, a pen) or for money (US$0.75–US$5). We stopped data collection when at least 22 research groups had each collected at least 105 participants, totalling 3,878 participants from 26 groups (Fig. 1; mean age (M_{age}), 26.6; s.d.$_{age}$, 10.6; 71% women, 28% men, 1% other). For the primary analyses, we excluded participants if they failed an attention check (17% fail rate), completed the study on a mobile device (3%), reported deviating from the pose instructions (1%), reported that their posed expression did not match an image of an actor completing the task correctly (3%), indicated that they were very distracted (3%) or exhibited any awareness of the study hypothesis (46%). (For the country-specific exclusion criteria rates, see the Supplementary Information.) An unexpectedly large number of participants were excluded for exhibiting awareness of the study hypothesis—but this may reflect an unusually strict classification scheme (that is, that two

coders must judge the participant as being completely unaware). This left 1,504 participants for the primary analyses.

Primary analyses

We hypothesized that participants would report higher levels of happiness (1) in the presence versus absence of emotional stimuli and (2) after posing happy versus neutral facial expressions. We also predicted that the effect of posed expressions on happiness would be larger in the presence than in the absence of positive stimuli. Following the study design (Supplementary Fig. 1), we modelled happiness reports with (1) Pose (happy or neutral), Facial Movement Task (facial mimicry, voluntary facial action or pen-in-mouth) and Stimuli Presence (present or absent) entered as effect-coded factors; (2) all higher-order interactions; (3) random intercepts for participants and research groups; and (4) random slopes for research groups.

Participants reported higher levels of happiness in the presence than in the absence of positive images ($M_{diff} = 0.30$; 95% confidence interval (CI), (0.12, 0.48); 5% scale range; $F(1, 22.65) = 10.67$; $P = 0.003$). However, the Bayesian analyses were inconclusive ($BF_{10} = 0.71$–1.25). Participants also reported more happiness after posing happy versus neutral expressions ($M_{diff} = 0.31$; 95% CI, (0.21, 0.40); 5.17% scale range; $F(1, 24.34) = 39.86$; $P < 0.001$; $BF_{10} = 61.06$–102.63). Contrary to our hypothesis, the Pose effect was not significantly larger in the presence than in the absence of positive stimuli ($F(1, 29.50) = 1.33$, $P = 0.26$, $BF_{10} = 0.06$–0.13).

Unexpectedly, there was an interaction between Pose and Facial Movement Task ($F(2, 32.95) = 17.11$, $P < 0.001$, $BF_{10} = 34.13$–100.14, Fig. 2). The effect of Pose on self-reported happiness was the largest in the facial mimicry task ($M_{diff} = 0.49$; 95% CI, (0.36, 0.61); 8.17% scale range; $F(1, 28.62) = 57.55$; $P < 0.001$; $BF_{10} > 100$) and the voluntary facial action task ($M_{diff} = 0.40$; 95% CI, (0.23, 0.56); 6.67% scale range; $F(1, 25.48) = 22.93$; $P < 0.001$; $BF_{10} = 25.20$–39.26). There was moderate support for the null hypothesis in the pen-in-mouth condition ($M_{diff} = 0.04$; 95% CI, (−0.07, 0.15); 0.67% scale range; $F(1, 24.74) = 0.57$; $P = 0.46$; $BF_{10} = 0.11$–0.17).

Secondary analyses

Our secondary analyses were designed to further probe the nature of facial feedback effects.

Potential aversion to the neutral expression posing task. The primary analyses suggest that posing happy versus natural expressions can increase feelings of happiness. However, an alternative explanation is that these effects are driven by hypothesis-irrelevant decreases in happiness after neutral poses (for example, as a result of boredom)[71]. To test this, we refit the primary analysis model with an effect-coded Pose factor that compared happy pose with filler trials that the participants completed. We focused on participants who were not exposed to positive images because these images were shown only during the facial posing trials (thus confounding their comparison with the filler trials). Nevertheless, similar results were observed in analyses that included participants who viewed positive images (Fig. 2).

Like the primary analyses, there was an interaction between Pose and Facial Movement Task ($F(2, 18.02) = 20.47$, $P < 0.001$). Participants reported higher levels of happiness after posing happy expressions versus completing filler tasks in both the facial mimicry task ($M_{diff} = 0.48$; 95% CI, (0.29, 0.67); 8% scale range; $t(22.4) = 5.23$; $P < 0.001$) and the voluntary facial action task ($M_{diff} = 0.20$; 95% CI, (0.05, 0.36); 3.33% scale range; $t(19.6) = 2.69$; $P = 0.01$). In the pen-in-mouth task, participants reported less happiness after completing the happy versus filler task ($M_{diff} = -0.15$; 95% CI, (−0.28, 0.02); 2.5% scale range; $t(31.5) = 2.39$; $P = 0.02$).

Moderating role of pose quality. We next examined the moderating role of three indicators of the quality of posed expressions: the
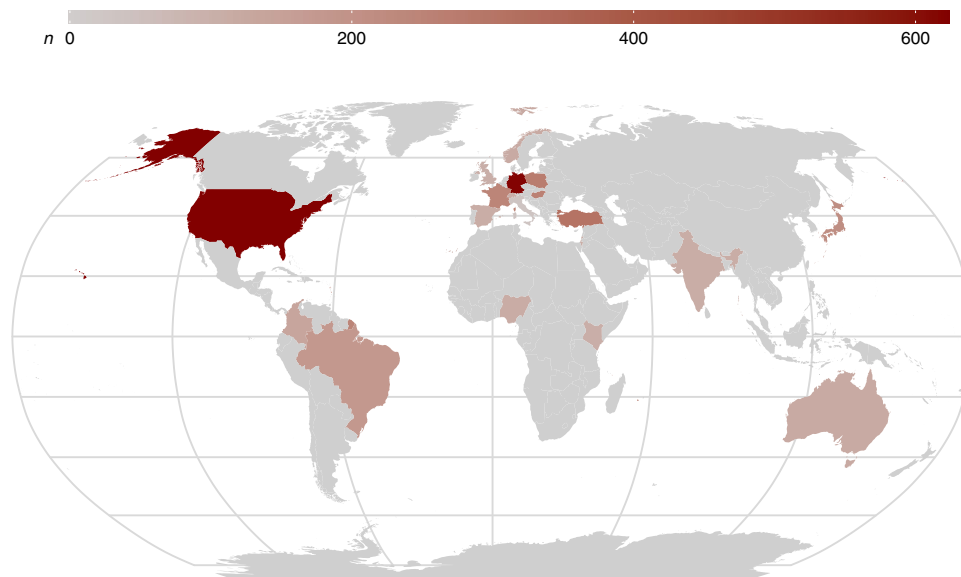
Fig. 1 | Country-specific sample sizes. Data were collected from 3,878 participants in 19 countries. Darker shades of red denote larger country-specific sample sizes.

participants' reports of the extent to which they followed pose instructions (compliance ratings), felt that their self-monitored expression matched an image of an actor successfully completing the task (similarity ratings) and felt that their posed expression resembled a genuine expression of happiness (genuineness ratings). For each quality indicator, we refit the primary analysis model with (1) the indicator entered mean-centred and (2) a term denoting its interaction with Pose. For each quality indicator, there was an interaction with Pose (Fig. 3). The effect of facial poses on happiness was larger among participants with higher compliance ($\beta = 0.08$; 95% CI, (0.05, 0.12); $t(1,482.63) = 4.33$; $P < 0.001$), similarity ($\beta = 0.03$; 95% CI, (0.01, 0.06); $t(1,358.62) = 3.37$; $P < 0.001$) and genuineness ratings ($\beta = 0.08$; 95% CI, (0.06, 0.09); $t(1,420.95) = 10.57$; $P < 0.001$).

Pose quality in different facial movement tasks. To examine whether pose quality varied between facial movement tasks, we used data from all 3,878 participants and modelled each quality indicator with (1) Facial Movement Task and Stimuli Presence entered as effect-coded factors, (2) random intercepts for research groups and (3) random slopes for research groups.

Compliance ratings varied by Facial Movement Task ($F(2, 18.18) = 10.50$, $P < 0.001$), but not Stimuli Presence ($M_{diff} = 0.03$; 95% CI, (−0.05, 0.11); 0.5% scale range; $F(1, 37.63) = 0.60$; $P = 0.44$). Compliance ratings were high across all tasks, but slightly lower in the facial mimicry task ($M = 6.45$, s.d. $= 1.07$) than in the voluntary facial action ($M = 6.57$; s.d. $= 0.93$; $M_{diff} = −0.15$; 95% CI, (−0.28, −0.02); 2.5% scale range; $t(23.5) = −2.47$; $P = 0.02$) and pen-in-mouth tasks ($M = 6.68$; s.d. $= 1.01$; $M_{diff} = −0.25$; 95% CI, (−0.37, −0.14); 4.17% scale range; $t(22.8) = −4.49$; $P < 0.001$). Compliance ratings were also slightly higher in the pen-in-mouth task than in the voluntary facial action task ($M_{diff} = 0.10$; 95% CI, (−0.01, 0.21); 1.67% scale range; $t(21.9) = 1.96$; $P = 0.06$).

Likewise, similarity ratings varied by Facial Movement Task ($F(2, 40.12) = 7.35$, $P = 0.002$), but not Stimuli Presence ($M_{diff} = −0.12$; 95% CI, (−0.25, 0.02); 2% scale range; $F(1, 19.18) = 3.15$; $P = 0.09$). Similarity ratings were high across all tasks but higher in the facial mimicry task ($M = 5.30$, s.d. $= 1.36$) than in the voluntary facial action ($M = 5.09$; s.d. $= 1.73$; $M_{diff} = 0.23$; 95% CI, (0.03, 0.43); 3.83% scale range; $t(22.7) = 2.43$; $P = 0.02$) and pen-in-mouth tasks ($M = 5.07$; s.d. $= 1.61$; $M_{diff} = 0.24$; 95% CI, (0.11, 0.36); 4% scale range; $t(194) = 3.63$; $P < 0.001$).

Genuineness ratings strongly varied by Facial Movement Task ($F(2, 13.69) = 82.56$, $P < 0.001$). Genuineness ratings were substantially lower in the pen-in-mouth task ($M = 2.98$; s.d. $= 1.89$) than in the facial mimicry ($M = 4.15$; s.d. $= 1.92$; $M_{diff} = −1.15$; 95% CI, (−1.34, −0.97); 19.17% scale range; $t(23.85) = 12.85$; $P < 0.001$) and voluntary facial action tasks ($M = 3.91$; s.d. $= 2.00$; $M_{diff} = −0.89$; 95% CI, (−1.12, −0.66); 14.83% scale range; $t(24.92) = 8.00$; $P < 0.001$). Genuineness ratings were also lower in the voluntary facial action task than in the facial mimicry task ($M_{diff} = −0.26$; 95% CI, (−0.48, −0.05); 4.33% scale range; $t(6.67) = −2.90$; $P = 0.02$). Participants also reported higher genuineness ratings in the presence ($M = 3.78$, s.d. $= 2.00$) than in the absence ($M = 3.57$, s.d. $= 2.00$) of positive images ($M_{diff} = 0.23$; 95% CI, (0.11, 0.34); 3.83% scale range; $F(1, 1,538.52) = 13.66$; $P < 0.001$).

Awareness of the study purpose. To examine whether some facial feedback tasks lead participants to be more aware of the study purpose, we used data from all 3,878 participants and modelled coder ratings of the extent to which they were aware with (1) Facial Movement Task and Stimuli Presence entered as effect-coded factors, (2) random intercepts for research groups and (3) random slopes for research groups. Awareness scores varied by Facial Movement Task ($F(2, 19.70) = 13.54$, $P < 0.001$), with participants being less aware in the pen-in-mouth task ($M = 1.75$, s.d. $= 1.41$) than in the voluntary facial action task ($M = 2.28$; s.d. $= 1.78$; $M_{diff} = −0.48$; 95% CI, (−0.67, −0.29); 8.02% scale range; $t(24) = −5.19$; $P < 0.001$) and the facial mimicry task ($M = 2.05$; s.d. $= 1.52$; $M_{diff} = −0.27$; 95% CI, (−0.43, −0.11); 4.48% scale range; $t(15.4) = −3.66$; $P < 0.05$). Participants were also less aware in the facial mimicry task than in the voluntary facial action task ($M_{diff} = −0.21$; 95% CI, (−0.36, −0.07); 3.53% scale range; $t(39.4) = −2.97$; $P = 0.005$).

To test whether facial feedback effects are amplified by awareness of the study purpose, we modelled happiness reports with (1) Pose, Facial Movement Task and Stimuli Presence entered as effect-coded factors; (2) awareness scores entered mean-centred; (3) a higher-order interaction term for Pose and awareness scores; (4) random intercepts for participants and research groups; and (5) research group random slopes for all terms other than awareness scores. The results indicated that the Pose effect was larger among participants who were more aware of the study hypothesis ($\beta = 0.08$; 95% CI, (0.06, 0.10); $t(22.74) = 7.55$; $P < 0.001$) (Fig. 3).
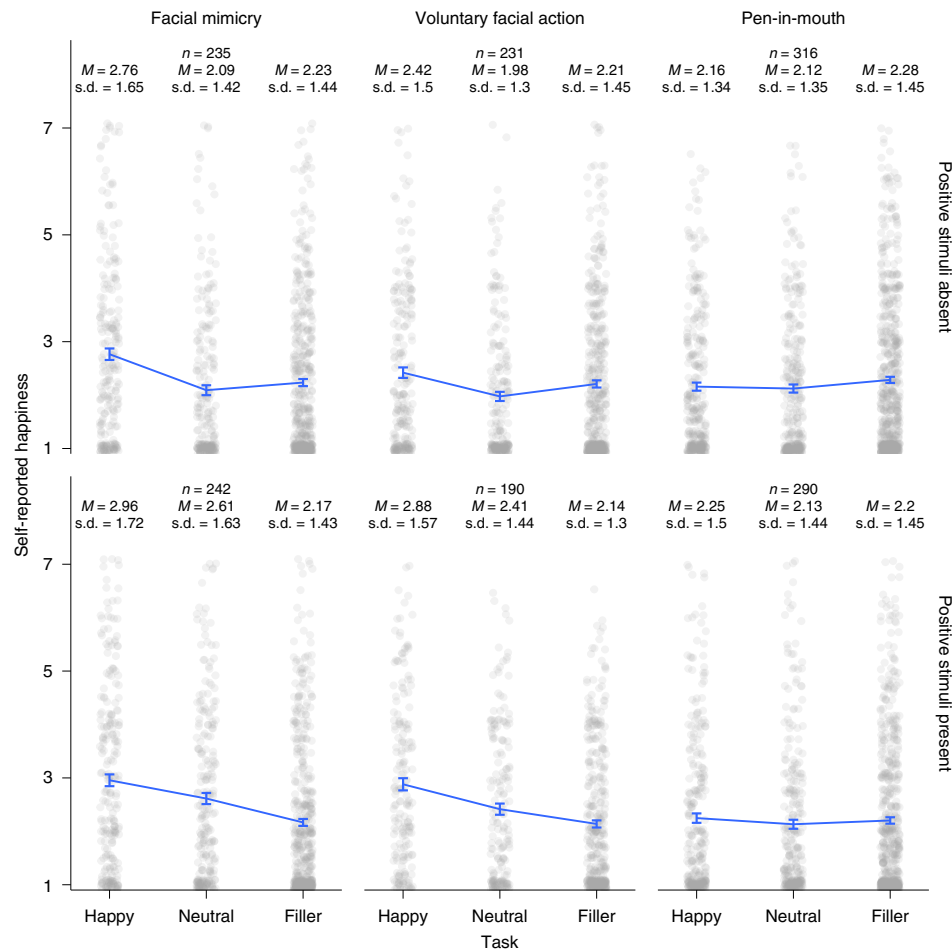
Fig. 2 | Effects of facial expression poses and filler tasks on self-reported happiness in each study condition. Self-reported happiness (1 = 'not at all' to 7 = 'an extreme amount') after the participants posed happy facial expressions, posed neutral facial expressions or completed filler tasks. The panel columns indicate whether the participants completed the facial mimicry, voluntary facial action or pen-in-mouth task. The panel rows indicate whether positive images were absent or present during the facial pose tasks. The grey points represent jittered participant observations. The blue error bars represent mean ± 1 standard error. Condition-specific sample sizes, means and standard deviations are reported.

Body awareness. To examine the moderating role of body awareness, we re-ran our primary analysis model with (1) participants' responses on a body awareness measure entered mean-centred and (2) a higher-order interaction term for Pose and awareness. No moderating role of body awareness was detected ($\beta = 0.00$; 95% CI, (-0.03, 0.03); $t(9.87) = 0.02$; $P = 0.99$) (Fig. 3).

Between-condition differences in other inclusion criteria. Next, we examined whether there were between-condition differences in the extent to which participants used an incorrect device to complete the study (for example, a phone) or failed attention checks. We separately modelled the probability that participants failed to meet each inclusion criterion using logistic mixed-effect regression with (1) Facial Movement Task and Stimuli Presence entered as effect-coded factors, (2) random intercepts for research groups and (3) random slopes for research groups.

The probability that participants used the incorrect device did not vary by Facial Movement Task (96%, 97% and 97% pass rates in the facial mimicry, voluntary facial action and pen-in-mouth tasks; $\chi^2(2) = 3.06$; $P = 0.22$) or Stimuli Presence (97% pass rate in the absence and presence of positive stimuli; $\chi^2(1) = 0.11$; $P = 0.74$). Likewise, the probability that participants failed attention checks did not vary by Facial Movement Task (84%, 82% and 83% pass rates in the facial mimicry, voluntary facial action and pen-in-mouth tasks; $\chi^2(2) = 1.28$; $P = 0.53$) or Stimuli Presence

(84% and 82% pass rates in the absence and presence of positive stimuli; $\chi^2(1) = 2.54$; $P = 0.11$).

We also tested for between-condition differences in coder ratings of the extent to which participants were distracted using linear mixed-effect regression with (1) Facial Movement Task and Stimuli Presence entered as effect-coded factors, (2) random intercepts for research groups and (3) random slopes for research groups. Distraction scores did not significantly vary between the facial mimicry ($M = 2.01$, s.d. $= 1.17$), voluntary facial action ($M = 1.92$, s.d. $= 1.14$) and pen-in-mouth ($M = 1.92$, s.d. $= 1.14$) tasks ($F(2, 18.57) = 2.45$, $P = 0.11$). Distraction scores also did not vary in the absence ($M = 1.94$, s.d. $= 1.15$) versus presence ($M = 1.96$, s.d. $= 1.16$) of positive stimuli ($F(1, 900.52) = 0.02$, $P = 0.90$).

Anger and anxiety. We next examined whether posed happy expressions decreased self-reported negative emotions and whether some facial movement tasks were more frustrating and anxiety-provoking than others. To do so, we separately re-ran our primary analyses with anxiety and anger reports as the dependent variables.

Happy versus neutral facial expression poses did not significantly decrease feelings of anger ($M_{\text{diff}} = -0.02$; 95% CI, (-0.07, 0.03); 0.33% scale range; $F(1, 20.71) = 0.85$; $P = 0.37$) or anxiety ($M_{\text{diff}} = -0.01$; 95% CI, (-0.06, 0.04); 0.17% scale range; $F(1, 25.36) = 0.32$; $P = 0.57$). However, feelings of anger ($F(2, 27.46) = 4.30$, $P = 0.02$) and anxiety
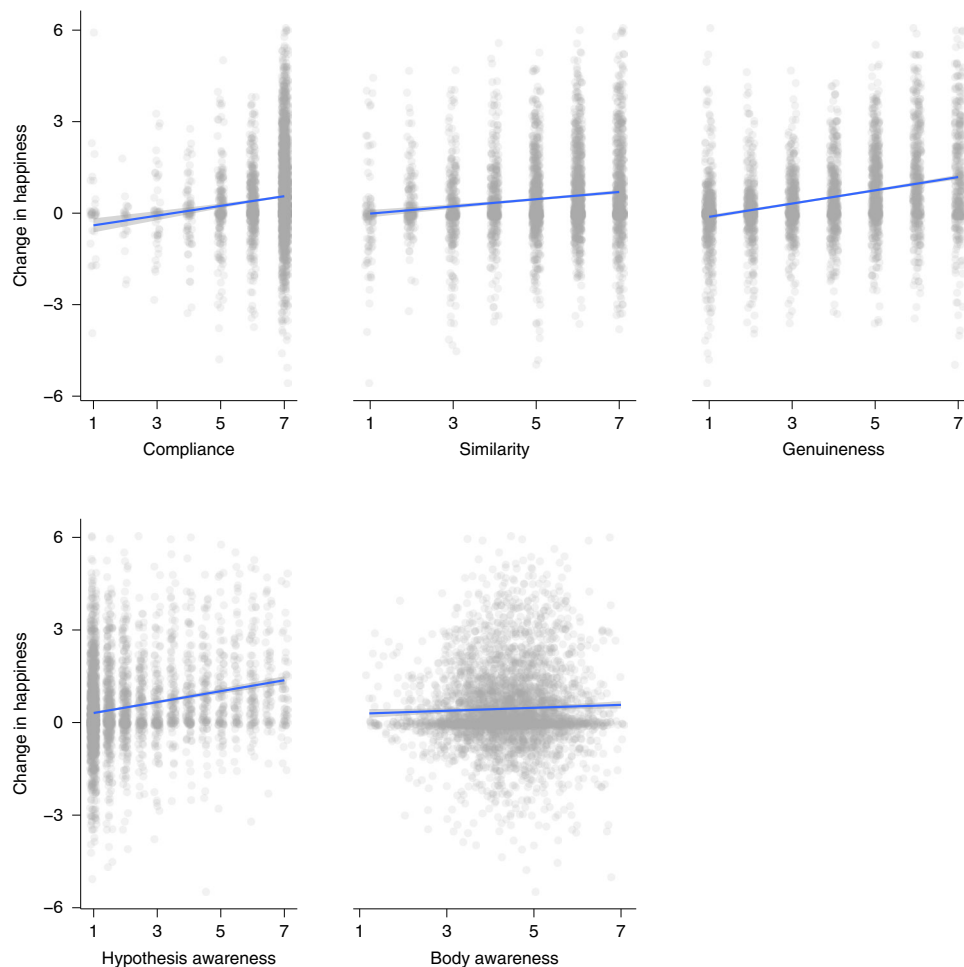
Fig. 3 | Potential moderators of facial feedback effects. The change in happiness (*y* axis) when the participants posed happy versus neutral expressions was moderated by compliance, similarity, genuineness and hypothesis awareness ratings, but not body awareness ratings (*x* axes). The grey points represent jittered participant observations. The blue lines represent the estimated linear relationships.

(F(2, 58.20) = 5.18, P = 0.008) did differ by Facial Movement Task. Participants reported higher levels of anger in the pen-in-mouth task than in the facial mimicry task (M_{diff} = 0.14; 95% CI, (0.03, 0.24); 2.33% scale range; t(24.2) = 2.64; P = 0.01) and the voluntary facial action task (M_{diff} = 0.12; 95% CI, (0.02, 0.21); 2% scale range; t(31.6) = 2.40; P = 0.02). Similarly, participants reported more anxiety in the pen-in-mouth task than in the facial mimicry task (M_{diff} = 0.13; 95% CI, (0.02, 0.24); 2.17% scale range; t(51.6) = 2.35; P = 0.02) and the voluntary facial action task (M_{diff} = 0.17; 95% CI, (0.06, 0.28); 2.83% scale range; t(79) = 3.00; P = 0.004). Nonetheless, follow-up exploratory analyses did not indicate that these increases in anxiety obfuscated facial feedback effects (Supplementary Information).

Exploratory analyses

For all analyses, we preregistered plans to model random slopes for research groups. However, random slopes often led to singular fit and convergence warnings, which is indicative of overfit models with potentially unreliable estimates[72]. Sensitivity analyses without (versus with) random slopes generally yielded identical inferences, except for the simple effect of Pose in the pen-in-mouth task. After we removed random slopes, the two-sided test of the effect of Pose was not significant (M_{diff} = 0.08; 95% CI, (−0.01, 0.16); 1.33% scale range; F(1, 1,498) = 2.78; P = 0.095), but an exploratory one-sided test was (one-sided P < 0.05). However, the Bayesian analyses were inconclusive (BF$_{10}$ = 0.46–0.96). Nonetheless, when we relaxed our inclusion criteria in a subsequent

sensitivity analysis, we found extremely strong evidence of a Pose effect in the pen-in-mouth task (M_{diff} = 0.14; 95% CI, (0.07, 0.21); 2.33% scale range; F(1, 3,872) = 16.37; P < 0.001; BF$_{10}$ > 100).

Discussion

Our project brought together a large adversarial team to design and conduct an experiment that best tested and clarified our disagreements about the facial feedback hypothesis. We designed our experiment not to provide close replications of any existing study but rather to provide informative tests of the facial feedback hypothesis. For example, our pen-in-mouth task was inspired by the original pen-in-mouth study that some, but not all[49], researchers have had difficulty replicating[40]. Nevertheless, our methodology differed in many ways from the original pen-in-mouth study. For example, we ran our study online (versus in person), focused on feelings of happiness (versus amusement), used a different cover story, had the participants pose expressions for a relatively short duration (five seconds) and did not instruct the participants to maintain the poses while they completed emotion ratings.

Our primary analyses replicated the pilot studies that informed the design of this study, albeit with more stringent inclusion criteria and a much larger and more culturally diverse sample (see Supplementary Fig. 2 for the country-specific effect size estimates). Contrary to theories that characterize peripheral nervous system activity and emotional experience as independent components of an emotion response[12–14], our results suggest that facial feedback can impact

feelings of happiness when using the facial mimicry and voluntary facial action tasks. Furthermore, these effects emerge in both the presence and absence of emotional stimuli—although, contrary to our prediction, the effect was not larger in the presence of emotional stimuli. Consistent with a previous meta-analysis, these results suggest that facial feedback can not only amplify ongoing feelings of happiness but also initiate feelings of happiness in otherwise neutral contexts[59].

Secondary analyses revealed that the observed facial feedback effects could not be explained by participants' aversion to the relatively inactive neutral pose task or demand characteristics. Even compared with relatively active filler trials, participants reported the most happiness after posing happy expressions. Furthermore, although facial feedback effects were larger among participants who were rated as more aware of the purpose of the study, we observed facial feedback effects among participants who did not exhibit such awareness. These results are consistent with recent experimental work demonstrating that demand characteristics can moderate, but do not fully account for, facial feedback effects[73].

Consistent with our predictions and a previous meta-analysis[59], facial feedback effects, when present, were small (see Supplementary Fig. 3 for the distribution of mean difference scores). Nonetheless, these effects were similar in size to the effect of mildly positive photos on happiness—that is, facial feedback was just as impactful as the external emotional context. Observing small effects is inconsistent with extreme claims that facial feedback is the primary determinant of emotional experience[2,74]. However, they support less extreme theories that characterize facial feedback as one of many components of the peripheral nervous system that contribute to emotional experience[47,75,76].

These results have implications for discussions about whether facial feedback interventions—such as those that might ask people to simply smile in the mirror for five seconds every morning—can be leveraged to manage distress[15,16], improve well-being[17,18] and reduce depression[19–39]. It is possible that relatively small facial feedback effects could accumulate into meaningful changes in well-being over time[77]. However, given that the similar-sized effect of positive images on happiness has not emerged as a serious well-being intervention, many (but not all) authors of this paper find it unlikely that facial feedback interventions will either.

Contrary to our predictions, the effect of posed facial expressions on happiness varied depending on the facial movement task. There was strong evidence of facial feedback effects in the facial mimicry and voluntary facial action tasks, but the evidence was less clear in the pen-in-mouth task. (This was despite avoiding video recording participants, which some[50]—but not all[59]—researchers argue interferes with facial feedback effects.) Our preregistered model with random slopes did not provide significant evidence of a simple effect of Pose in the pen-in-mouth condition, and Bayesian analyses provided moderate support for the null hypothesis. An exploratory one-sided test of this effect was significant when we removed random slopes from the model, but Bayesian analyses characterized the evidence as inconclusive. However, when we relaxed our inclusion criteria, both frequentist and Bayesian analyses provided strong evidence of a facial feedback effect in the pen-in-mouth task. Nonetheless, we preregistered that this would be considered a less stringent test of the facial feedback hypothesis.

Although it is less clear whether the pen-in-mouth task had a non-zero effect on feelings of happiness, the effect is clearly smaller than that produced by the facial mimicry and voluntary facial action tasks. This may suggest that different mechanisms underlie the effects produced by each task. Researchers do not agree on which mechanisms underlie facial feedback effects[73], but they may involve both inferential processes (for example, people inferring they are happy because they are smiling)[45,46] and non-inferential processes (for example, smiling automatically activating other physiological components of emotion)[5,54]. Unlike other facial feedback tasks, the pen-in-mouth task was designed to limit the role of inferential process by manipulating

facial expressions covertly[41]. Consistent with this goal, participants in the pen-in-mouth condition were less likely to report that the posed happy expression felt genuine. This may mean that inferential processes were minimized in this task, thus reducing the size of the facial feedback effect. Contrary to this explanation, though, we did not find that facial feedback effects were moderated by self-report measures of general attentiveness to non-emotional bodily process. (See the Supplementary Information for similar results from pilot studies using a multifaceted self-report of body awareness.)

Alternatively, the pen-in-mouth task may have created a less prototypical expression of happiness—which, regardless of the role of inferential processes, may attenuate facial feedback effects[51–53]. Specifically, facial feedback effects may be amplified when the task activates muscles typically associated with an emotional state and attenuated when the task activates muscles not typically associated with an emotional state. In retrospect, the pen-in-mouth task we used may simultaneously activate muscles associated with biting, which may attenuate its effect on happiness reports. Furthermore, a robust pen-in-mouth effect may emerge if one uses a variant of the task that better activates the orbicularis oculi muscles, which is associated with genuine expressions of happiness[56]. However, our results provide mixed support for these predictions. On one hand, facial feedback effects did not differ between the other two tasks, which were designed to produce less prototypical (voluntary facial action task) and more prototypical (facial mimicry task) expressions of happiness. On the other hand, facial feedback effects were larger when participants reported posing higher-quality expressions. Future research can further investigate this issue by more directly measuring muscle activity using facial action coding[78], electromyography[79], sonography[80] or thermography[81].

To conclude, our adversarial collaboration was partly inspired by conflicting narratives about the validity of the facial feedback hypothesis. We began the collaboration after a large team of researchers failed to replicate a seminal demonstration of facial feedback effects using a pen-in-mouth task[40], but a meta-analysis indicated that facial feedback has a small but significant effect on emotional experience[59]. Our results do not provide unequivocal evidence of a pen-in-mouth effect. Nonetheless, they do provide strong evidence that other tasks designed to produce partial or full recreations of happy expressions can both modulate and initiate feelings of happiness. It has been nearly 100 years since researchers began famously debating whether peripheral nervous system activity is merely a by-product of emotion processes. Consistent with theories positing that peripheral nervous system activity impacts emotional experience, our results a century later provide strong evidence of facial feedback effects. With this foundation strengthened, future researchers can turn their attention to answering new questions about when and why these effects occur.

Methods

Ethics
Each research group received approval from their local Ethics Committee or Institutional Review Board to conduct the study (for example, University of Tennessee IRB-19-05313-XM), indicated that their institution does not require approval for the researchers to conduct this type of research or indicated that the current study is covered by a pre-existing approval. At the time of Stage 1 submission, 22 research groups had ethics approval to collect data, but additional sites with pending ethics approval joined the project later. All participants provided informed consent.

Procedure
The experiment was presented via Qualtrics. Due to constraints created by COVID-19, we planned for data collection to primarily occur online. However, research groups were allowed to collect data in the laboratory if they indicated they could do so safely. Before beginning the study, the participants were asked to confirm that they had a clean

pen or pencil nearby that they were willing to place in their mouths, were completing the study on a desktop computer or laptop (details regarding the participants' operating systems were automatically recorded to confirm) and were in a setting with minimal distractions.

The participants were told that the study was investigating how physical movements and cognitive distractors influence mathematical speed and accuracy and that they would complete four simple movement tasks and math problems. The first and last tasks were randomly presented filler trials that helped ensure the cover story was believable ("Place your left hand behind your head and blink your eyes once per second for 5 seconds" and "Tap your left leg with your right-hand index finger once per second for 5 seconds"). In the two critical tasks, the participants were asked to pose happy and neutral facial expressions in randomized order through the facial mimicry, voluntary facial action or pen-in-mouth procedure. While posing these expressions, some participants were randomly assigned to view positive images. To reinforce the cover story, the participants were provided with an on-screen timer during all tasks.

After each task (including the filler tasks), the participants completed a simple filler arithmetic problem and the Discrete Emotions Questionnaire's four-item happiness subscale, which asked the participants to indicate the degree to which they experienced happiness, satisfaction, liking and enjoyment during the preceding task (1 = 'not at all' to 7 = 'an extreme amount')[82]. The participants also completed two items measuring anxiety (worry and nervous). To further obscure the purpose of the study, the participants also completed one anger, tiredness and confusion filler item. All emotion items were presented in random order. By not referencing the emotional stimuli, this questionnaire better captured self-focused, as opposed to world-focused, emotional experience[57,58]. Afterwards, the participants rated how much they liked the task and how difficult they found the task and arithmetic problem. In the non-filler tasks, an attention check item asking the participants to choose a specific response option was randomly inserted in the questions regarding the task and arithmetic problem difficulty.

In the facial mimicry condition, the participants were shown a 2 × 2 image matrix of actors posing happy expressions. The participants were then instructed to either mimic these expressions (happy condition) or maintain a blank expression (neutral condition). Importantly, having the participants view the happy expression matrix before both the happy and neutral trials ensured that any potentially confounding effects that images of smiling people have on emotional experience were constant across the mimicry trials. The expression matrix was displayed for at least five seconds, and the participants indicated when they were ready to perform the task. In the voluntary facial action condition, the participants were instructed to either move the corners of their lips up towards their ears and elevate their cheeks using only the muscles in their face (happy condition) or maintain a blank facial posture (neutral condition). In the pen-in-mouth condition, the participants received video instructions regarding the correct way to hold the pen in their teeth (happy condition) or lips (neutral condition). During all facial pose tasks, the participants were instructed to maintain the poses for five seconds, the approximate duration of spontaneous happiness expressions[83].

After completing the five movement tasks, the participants answered a variety of open-ended questions regarding their beliefs about the purpose of the experiment via Qualtrics. Each research group recruited two independent, results-blind coders to review the open-ended responses. The coders were provided a written description of the study purpose and methods and subsequently reviewed the participants' open-ended responses in randomized order. On the basis of the open-ended responses, the coders rated the degree to which each participant was aware of the true purpose of the experiment (1 = 'not at all aware' to 7 = 'completely aware').

After answering questions about their beliefs regarding the purpose of the experiment, the participants completed a short demographic form and the Body Awareness Questionnaire[84]. The participants then answered several questions related to the quality of their data. First, the participants were 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, the participants were asked to repeat the task and rate the degree to which it felt like they were expressing happiness (1 = 'not at all' to 7 = 'exactly'). Third, the participants were asked to watch themselves repeat the task (for example, via a mirror or camera phone) and indicate the degree to which their expression matched an image of an individual completing the task correctly (1 = 'not at all' to 7 = 'exactly'). Fourth, the participants were asked to describe any issues that may have compromised the quality of their data (such as distractions). The two coders from each research group reviewed the responses to this last question and rated the degree to which each participant was distracted (1 = 'not at all distracted' to 7 = 'completely distracted'). The participants were told that there would not be a penalty for indicating that they did not complete the task correctly or that there were issues with the quality of their data.

Ideally, the quality of the participants' posed expressions would have been assessed via video recordings or participant-submitted photos. However, many members of our collaboration expressed doubts about receiving ethical approval to collect and share images or recordings. Participants in many of our data collection regions may also have lacked a web camera. Furthermore, researchers are still debating whether awareness of overt video recording interferes with facial feedback effects[49,50,59,85]. Nevertheless, pilot study recordings and self-reports confirmed that almost all participants successfully posed the target facial expressions (Supplementary Information).

Materials

In the facial mimicry task, the participants all viewed the same 2 × 2 image matrix of actors posing happy facial expressions from the Extended Cohn–Kanade Dataset[86]. All four actors posed prototypical facial expressions of happiness, as confirmed by coders trained in the Facial Action Coding System[78]. An image matrix of actors, as opposed to a single image, was used so that the participants had multiple examples of the movement and were provided with more options for a suitable facial model. In the pen-in-mouth task, the instructional videos were adopted from Wagenmakers and colleagues' replication materials[40].

During the two facial expression pose tasks, one group of participants viewed an array of four positive photos (for example, photos of dogs, flowers, kittens and rainbows). Multiple photos (as opposed to a single photo) were used to increase the probability that the participants found at least one of the photos emotionally evocative. All photos were drawn from a database comprising 100 images from the internet and the International Affective Picture System[87] that were separately rated on how good and bad they were[88]. The results from the three pilot studies confirmed that these images successfully elicited feelings of happiness (Supplementary Information). Due to potential cross-cultural differences in what types of photos elicit happiness (for example, dog photos can be expected to elicit happiness in many Western cultures but not in all African cultures), each lab was permitted to replace photos with more culturally appropriate positive photos. For non-English-speaking data collection sites, the experiment materials were translated into the local language.

Primary analyses

Due to the nested nature of the data (for example, ratings nested within individuals, which were nested within research groups), we used linear multilevel modelling. More specifically, happiness reports were modelled with (1) Pose, Facial Movement Task and Stimuli Presence entered as factors; (2) random intercepts for research groups and participants; and (3) random slopes for research groups. All hypotheses in Table 1 were examined using both null hypothesis significance testing and Bayesian alternatives.

Participants were excluded from the primary analyses if they (1) exhibited any awareness of the facial feedback hypothesis (that is, received an awareness score over 1 from two independent coders), (2) disclosed that they were very distracted during the study (that is, 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 that they did not follow the pose instructions, (5) indicated that 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. These stringent exclusion criteria were added after we failed to observe the pen-in-mouth effect in pilot study 3.

Secondary analyses

Although our primary analyses were run with the aforementioned exclusion criteria, we also re-ran these analyses to examine whether the exclusion criteria interact with Pose to influence happiness reports. We also examined whether these exclusion criterion variables varied as a function of Facial Movement Task and Stimuli Presence.

To examine the alternative explanation that doing something (for example, posing a happy facial expression) may simply be more enjoyable than doing nothing (for example, posing a neutral facial expression), we also re-ran our primary analyses with a factor contrasting the happy pose and filler trials.

Although previous research has indicated that many psychology studies yield similar effect sizes when completed online versus in a lab[89], we recorded the mode of data collection and planned to re-run our primary analyses with the data collection mode included as a moderator. However, we noted that this analysis may be confounded by (1) whether the research group is a proponent or a critic of the facial feedback hypothesis (that is, proponents may be more likely to collect data in the laboratory) and (2) the region of data collection (that is, research groups in regions with fewer COVID-19 cases may be more likely to collect data in the laboratory).

Although we did not anticipate a Pose by Facial Movement Task interaction, we noted that the pen-in-mouth condition may lead to heightened levels of anxiety in the midst and/or aftermath of COVID-19. Although this is speculative, heightened levels of anxiety may interfere with facial feedback effects. Consequently, as an exploratory analysis, we examined whether anxiety ratings differ as a function of Facial Movement Task.

Power simulation

Power analysis was performed via a linear multilevel modelling simulation. We randomly generated normally distributed data for 96 participants from 22 research groups. Effect size estimates for the hypothesized effects of Pose ($d = 0.39$), Stimuli Presence ($d = 0.68$) and the Pose by Stimuli Presence interaction ($d = 0.29$) were estimated from pilot studies 1 and 2 (Supplementary Information). All other effects were set to zero. Pilot study 3 was run after initial in-principle acceptance was granted and yielded somewhat different effect size estimates. However, this pilot study led to minor refinements in the exclusion criteria that left our original predictions unchanged.

On the basis of two pilot studies, we simulated random intercepts for participants with s.d. = 0.70. We did not simulate random slopes for participants since there are only two observations within each participant, which would probably lead to convergence issues. Random slopes for research groups were simulated on the basis of the values from the previous many-lab failure to replicate[40]. For the hypothesized effects, we specified conservative random slope estimates on the basis of the standard deviation of their meta-analytic effect size from the previous many-lab failure to replicate (s.d. = 0.28). For the effects we expected to be zero, we specified random slopes on the basis of the random slope from the previous many-lab failure to replicate ($\tau^2 \approx 0$). However, due to convergence issues, the research groups random slope for the facial

feedback task factor was removed. Residual variance was set to 0.60 on the basis of the estimates from pilot studies 1 and 2.

The results from this power simulation indicated that over 95% power for all our hypothesized effects could be obtained with at least 1,584 participants. However, on the basis of pilot study 3, we estimated that 44% of the participants would not meet our strict inclusion criteria, leading to a desired sample of 2,281. We therefore planned to stop collecting data once one of the following conditions was met: (1) 22 labs had collected 105 participants each or (2) at least six months had elapsed since the start of data collection and we had at least 2,281 participants. We planned for a minimum of 22 labs to collect data for this project, although additional labs with pending ethics approval were allowed to join the project later.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The full data are publicly available at https://osf.io/ac3t2/. Source data are provided with this paper.

Code availability

The full analysis code is publicly available at https://osf.io/ac3t2/.

References

1. Zajonc, R. B. The primacy of affect. *Am. Psychol.* **40**, 849–850 (1985).
2. Tomkins, S. *Affect Imagery Consciousness: The Positive Affects* Vol. 1 (Springer, 1962).
3. Laird, J. D. & Crosby, M. in *Thought and Feeling: Cognitive Alteration of Feeling States* (eds London, H. & Nisbett, R. E.) 44–59 (Transaction, 1974).
4. Allport, F. H. A physiological–genetic theory of feeling and emotion. *Psychol. Rev.* **29**, 132–139 (1922).
5. Levenson, R. W., Ekman, P. & Friesen, W. V. Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology* **27**, 363–384 (1990).
6. Coan, J. A., Allen, J. J. B. & Harmon-Jones, E. Voluntary facial expression and hemispheric asymmetry over the frontal cortex. *Psychophysiology* **38**, 912–925 (2001).
7. Scherer, K. R. & Moors, A. The emotion process: event appraisal and component differentiation. *Annu. Rev. Psychol.* **70**, 719–745 (2019).
8. Stepper, S. & Strack, F. Proprioceptive determinants of emotional and nonemotional feelings. *J. Pers. Soc. Psychol.* **64**, 211–220 (1993).
9. Friedman, B. H. Feelings and the body: the Jamesian perspective on autonomic specificity of emotion. *Biol. Psychol.* **84**, 383–393 (2010).
10. James, W. Discussion: the physical basis of emotion. *Psychol. Rev.* **1**, 516–529 (1894).
11. Lange, C. G. *Om Sindsbevaegelser; Et Psyko-Fysiologisk Studie* (Lund, 1885).
12. Cannon, W. The James–Lange theory of emotions: a critical examination and an alternative theory. *Am. J. Psychol.* **39**, 106–124 (1927).
13. Cannon, W. *Bodily Changes in Pain, Hunger, Fear and Rage* (D. Appleton, 1915).
14. Sherrington, C. S. Experiments on the value of vascular and visceral factors for the genesis of emotion. *Proc. R. Soc. Lond.* **66**, 390–403 (1899).
15. Ansfield, M. E. Smiling when distressed: when a smile is a frown turned upside down. *Pers. Soc. Psychol. Bull.* **33**, 763–775 (2007).

16. Kraft, T. L. & Pressman, S. D. Grin and bear it: the influence of manipulated facial expression on the stress response. *Psychol. Sci.* **23**, 1372–1378 (2012).

17. Schmitz, B. *Art-of-Living: A Concept to Enhance Happiness* (Springer Cham, 2016).

18. Lyubomirsky, S. *The How of Happiness: A Scientific Approach to Getting the Life You Want* (Penguin Group, 2008).

19. Alam, M., Barrett, K. C., Hodapp, R. M. & Arndt, K. A. Botulinum toxin and the facial feedback hypothesis: can looking better make you feel happier? *J. Am. Acad. Dermatol.* **58**, 1061–1072 (2008).

20. Alves, M. C., Sobreira, G., Aleixo, M. A. & Oliveira, J. M. Facing depression with botulinum toxin: literature review. *Eur. Psychiatry* **335**, 5290–5643 (2016).

21. Chugh, S., Chhabria, A., Jung, S., Kruger, T. H. C. & Wollmer, M. A. Botulinum toxin as a treatment for depression in a real-world setting. *J. Psychiatr. Pract.* **24**, 15–20 (2018).

22. Finzi, E. Update: botulinum toxin for depression: more than skin deep. *Dermatol. Surg.* **44**, 1363–1365 (2018).

23. Finzi, E. & Rosenthal, N. E. Emotional proprioception: treatment of depression with afferent facial feedback. *J. Psychiatr. Res.* **80**, 93–96 (2016).

24. Finzi, E. & Rosenthal, N. E. Treatment of depression with onabotulinumtoxinA: a randomized, double-blind, placebo controlled trial. *J. Psychiatr. Res.* **52**, 1–6 (2014).

25. Finzi, E. & Wasserman, E. Treatment of depression with botulinum toxin A: a case series. *Dermatol. Surg.* **32**, 645–649 (2006).

26. Fromage, G. Exploring the effects of botulinum toxin type A injections on depression. *Aesthet. Nurs.* **7**, 315–317 (2018).

27. Hexsel, D. et al. Evaluation of self-esteem and depression symptoms in depressed and nondepressed subjects treated with onabotulinumtoxinA for glabellar lines. *Dermatol. Surg.* **39**, 1088–1096 (2013).

28. Krüger, T. H. C., Jung, S. & Wollmer, M. A. Botulinumtoxin— ein neuer wirkstoff in der psychopharmakotherapie? *Psychopharmakotherapie* **23**, 2–7 (2016).

29. Lewis, M. B. & Bowler, P. J. Botulinum toxin cosmetic therapy correlates with a more positive mood. *J. Cosmet. Dermatol.* **8**, 24–26 (2009).

30. Magid, M. et al. Treating depression with botulinum toxin: a pooled analysis of randomized controlled trials. *Pharmacopsychiatry* **48**, 205–210 (2015).

31. Magid, M. & Reichenberg, J. S. Botulinum toxin for depression? An idea that's raising some eyebrows. *Curr. Psychiatr.* **14**, 43–56 (2015).

32. Magid, M. et al. Treatment of major depressive disorder using botulinum toxin A: a 24-week randomized, double-blind, placebo-controlled study. *J. Clin. Psychiatry* **75**, 837–844 (2014).

33. Parsaik, A. K. et al. Role of botulinum toxin in depression. *J. Psychiatr. Pract.* **22**, 99–110 (2016).

34. Reichenberg, J. S. et al. Botulinum toxin for depression: does patient appearance matter? *J. Am. Acad. Dermatol.* **74**, 171–173 (2016).

35. Wollmer, M. A., Magid, M. & Kruger, T. H. C. in *Practical Psychodermatology* (eds Bewley, A. et al.) 216–219 (John Wiley & Sons, 2014).

36. Wollmer, M. A. et al. Agitation predicts response of depression to botulinum toxin treatment in a randomized controlled trial. *Front. Psychiatry* **5**, 36 (2014).

37. Wollmer, M. A. et al. Facing depression with botulinum toxin: a randomized controlled trial. *J. Psychiatr. Res.* **46**, 574–581 (2012).

38. Zamanian, A., Jolfaei, A. G., Mehran, G. & Azizian, Z. Efficacy of Botox versus placebo for treatment of patients with major depression. *Iran. J. Public Health* **46**, 982–984 (2017).

39. Finzi, E. *The Face of Emotion: How Botox Affects Our Moods and Relationships* (St. Martin's, 2013).

40. Wagenmakers, E.-J. et al. Registered replication report: Strack, Martin, & Stepper (1988). *Perspect. Psychol. Sci.* **11**, 917–928 (2016).

41. Strack, F., Martin, L. L. & Stepper, S. Inhibiting and facilitating conditions of the human smile: a nonobtrusive test of the facial feedback hypothesis. *J. Pers. Soc. Psychol.* **54**, 768–777 (1988).

42. Buck, R. Nonverbal behavior and the theory of emotion: the facial feedback hypothesis. *J. Pers. Soc. Psychol.* **38**, 811–824 (1980).

43. Zuckerman, M., Klorman, R., Larrance, D. T. & Spiegel, N. H. Facial, autonomic, and subjective components of emotion: the facial feedback hypothesis versus externalizer–internalizer distinction. *J. Pers. Soc. Psychol.* **41**, 929–944 (1981).

44. Ekman, P. & Oster, H. Facial expressions of emotion. *Annu. Rev. Psychol.* **30**, 527–554 (1979).

45. Laird, J. D. Self-attribution of emotion: the effects of expressive behavior on the quality of emotional experience. *J. Pers. Soc. Psychol.* **29**, 475–486 (1974).

46. Laird, J. D. & Bresler, C. in *Review of Personality and Social Psychology: Emotion* (ed. Clark, M. S.) 213–234 (Sage, 1992).

47. Ekman, P. in *Anthropology of the Body* (ed. Blacking, J.) 34–38 (Routledge, 1979).

48. Schimmack, U. & Chen, Y. The power of the pen paradigm: a replicability analysis. *Replicability-Index* https://replicationindex. com/2017/09/04/the-power-of-the-pen-paradigm-a-replicability- analysis/ (2017).

49. Strack, F. Reflection on the Smiling Registered Replication Report. *Perspect. Psychol. Sci.* **11**, 929–930 (2016).

50. Noah, T., Schul, Y. & Mayo, R. When both the original study and its failed replication are correct: feeling observed eliminates the facial-feedback effect. *J. Pers. Soc. Psychol.* **114**, 657–664 (2018).

51. Hager, J. C. & Ekman, P. Methodological problems in Tourangeau and Ellsworth's study of facial expression and experience of emotion. *J. Pers. Soc. Psychol.* **40**, 358–362 (1981).

52. Tomkins, S. The role of facial response in the experience of emotion: a reply to Tourangeau and Ellsworth. *J. Pers. Soc. Psychol.* **37**, 1519–1531 (1981).

53. Matsumoto, D. The role of facial response in the experience of emotion: more methodological problems and a meta-analysis. *J. Pers. Soc. Psychol.* **52**, 769–774 (1987).

54. Levenson, R. W., Ekman, P., Heider, K. & Friesen, W. V. Emotion and autonomic nervous system activity in the Minangkabau of West Sumatra. *J. Pers. Soc. Psychol.* **62**, 972–988 (1992).

55. Ekman, P. Facial expression and emotion. *Am. Psychol.* **48**, 384–392 (1993).

56. Soussignan, R. Duchenne smile, emotional experience, and autonomic reactivity: a test of the facial feedback hypothesis. *Emotion* **2**, 52–74 (2002).

57. Lambie, J. A. & Marcel, A. J. Consciousness and the varieties of emotion experience: a theoretical framework. *Psychol. Rev.* **109**, 219–259 (2002).

58. Frijda, N. H. Emotion experience. *Cogn. Emot.* **194**, 473–497 (2010).

59. Coles, N. A., Larsen, J. T. & Lench, H. C. A meta-analysis of the facial feedback literature: effects of facial feedback on emotional experience are small and variable. *Psychol. Bull.* **145**, 610–651 (2019).

60. Carter, E. C., Schönbrodt, F. D., Gervais, W. M. & Hilgard, J. Correcting for bias in psychology: a comparison of meta- analytic methods. *Adv. Methods Pract. Psychol. Sci.* **2**, 115–144 (2019).

61. Macaskill, P., Walter, S. D. & Irwig, L. A comparison of methods to detect publication bias in meta-analysis. *Stat. Med.* **20**, 641–654 (2001).

62. Stanley, T. D. Limitations of PET-PEESE and other meta-analysis methods. *Soc. Psychol. Pers. Sci.* **8**, 581–591 (2017).

63. Eysenck, H. J. An exercise in mega-silliness. *Am. Psychol.* **33**, 517 (1978).

64. Kleinke, C. L., Peterson, T. R. & Rutledge, T. R. Effects of self-generated facial expressions on mood. *J. Pers. Soc. Psychol.* **74**, 272–279 (1998).

65. Dimberg, U. & Söderkvist, S. The voluntary facial action technique: a method to test the facial feedback hypothesis. *J. Nonverbal Behav.* **35**, 17–33 (2011).

66. R Core Team. R: A Language and Environment for Statistical Computing v.4.1.2 https://www.Rproject.org/ (R Foundation for Statistical Computing, 2021).

67. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4: mixed-effects modeling with R. *J. Stat. Softw.* **67**, 1–48 (2015).

68. Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. lmerTest package: tests in linear mixed effects models. *J. Stat. Softw.* **82**, 1–26 (2017).

69. Lenth, R. V. emmeans: Estimated marginal means, aka least-squares means. R package version 1.7.2 (2022).

70. Morey, R. D. & Rouder, J. N. BayesFactor: Computation of Bayes factors for common designs. R package version 0.9.12-4.3 (2021).

71. Wilson, T. D. et al. Just think: the challenges of the disengaged mind. *Science* **345**, 75–77 (2014).

72. Brown, V. A. An introduction to linear mixed-effects modeling in R. *Adv. Methods Pract. Psychol. Sci.* **4**, 1–19 (2021).

73. Coles, N. A., Gaertner, L., Frohlich, B., Larsen, J. T. & Basnight-Brown, D. Fact or artifact? Methodological artifacts moderate, but do not fully account for, the effects of facial feedback on emotional experience. *J. Pers. Soc. Psychol.* 1–24 (2022).

74. Izard, C. E. *The Face of Emotion* (Appleton-Century-Crofts, 1971).

75. James, W. What is an emotion? *Mind* **9**, 188–205 (1884).

76. Laird, J. D. & Lacasse, K. Bodily influences on emotional feelings: accumulating evidence and extensions of William James's theory of emotion. *Emot. Rev.* **6**, 27–34 (2014).

77. Funder, D. C. & Ozer, D. J. Evaluating effect size in psychological research: sense and nonsense. *Adv. Methods Pract. Psychol. Sci.* **2**, 156–168 (2019).

78. Ekman, P. & Rosenberg, E. L. *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)* (Oxford Univ. Press, 1997).

79. Larsen, J. T., Norris, C. J. & Cacioppo, J. T. Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii. *Psychophysiology* **40**, 776–785 (2003).

80. Alfen, N. Van, Gilhuis, H. J., Keijzers, J. P., Pillen, S. & Van Dijk, J. P. Quantitative facial muscle ultrasound: feasibility and reproducibility. *Muscle Nerve* **48**, 375–380 (2013).

81. Clay-Warner, J. & Robinson, D. T. Infrared thermography as a measure of emotion response. *Emot. Rev.* **7**, 157–162 (2014).

82. Harmon-Jones, C., Bastian, B. & Harmon-Jones, E. The Discrete Emotions Questionnaire: a new tool for measuring state self-reported emotions. *PLoS ONE* **11**, e0159915 (2016).

83. Ekman, P. Darwin, deception, and facial expression. *Ann. N. Y. Acad. Sci.* **1000**, 205–221 (2003).

84. Shields, S. A., Mallory, M. E. & Simon, A. The body awareness questionnaire: reliability and validity. *J. Pers. Assess.* **53**, 802–815 (1989).

85. Marsh, A. A., Rhoads, S. A. & Ryan, R. M. A multi-semester classroom demonstration yields evidence in support of the facial feedback effect. *Emotion* **19**, 1500–1504 (2019).

86. Lucey, P. et al. The Extended Cohn–Kanade Dataset (CK+): a complete dataset for action unit and emotion-specified expression. In *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 94–101 (IEEE, 2010).

87. Lang, P. & Bradley, M. M. in *Handbook of Emotion Elicitation and Assessment* (eds Coan, J. A. & Allen, J. J. B.) 29–46 (Oxford Univ. Press, 2007).

88. March, D. S., Gaertner, L. & Olson, M. A. In harm's way: on preferential response to threatening stimuli. *Pers. Soc. Psychol. Bull.* **43**, 1519–1529 (2017).

89. Klein, R. A. et al. Many Labs 2: investigating variation in replicability across samples and settings. *Adv. Methods Pract. Psychol. Sci.* **1**, 443–490 (2018).

Author contributions

Conceptualization: N.A.C., D.S.M., F.M.-R., J.T.L., J.F.H., P.M.G., P.C.E., L.G. and F.S. Data curation: N.A.C., B.S., Y.Y. and S.R.-F. Formal analysis: N.A.C., L.G., M.M. and M.T.L. Funding acquisition: N.A.C., Y.Y. and N.B. Investigation: N.A.C., D.S.M., J.T.L., N.C.A., I.L.G.N., M.L.W., F.F., N.R., A.M., J.F.H., G.K., E.Y., A.K., N.H., J.T., R.M.K.F., D.Z., B.A., K.B., S.A., K.F., Y.Y., A.I., D.L.E., C.A.L., S.L., M.P., N.B., G.P., D.M.B.-B., J.A.H., P.R.M., L.G.J.D., K.V., H.IJ., N.T., S.D.P., P.M.G., A.A.Ö., S.R.-F. and M.T.L. Methodology: N.A.C., D.S.M., F.M.-R., P.S.F., J.F.H., G.K., K.B., D.L.E., S.R.-F., P.C.E. and L.G. Project administration: N.A.C., M.L.W., F.F., P.S.F., J.F.H., J.T., K.B., K.F., D.L.E., M.P., H.IJ., S.D.P. and A.A.Ö. Resources: N.A.C., D.S.M., I.L.G.N., E.Y., A.K., T.N., R.M.K.F., B.A., K.B., S.A., M.P., G.P., J.A.H., P.R.M., H.IJ., P.M.G., A.A.Ö. and S.R.-F. Software: N.A.C., J.T. and M.M. Supervision: N.A.C., N.C.A., F.F., N.R., J.F.H., B.A., K.B., C.A.L., N.B., H.IJ. and S.D.P. Validation: N.A.C., P.S.F., N.H., J.T., M.P., N.T., M.M. and M.T.L. Visualization: N.A.C., P.S.F., J.A.H. and L.G. Writing—original draft: N.A.C., D.S.M., A.A.Ö. and L.G. Writing—review and editing: N.A.C., D.S.M., F.M.-R., J.T.L., N.C.A., I.L.G.N., M.L.W., F.F., N.R., A.M., P.S.F., J.F.H., G.K., T.N., N.H., D.Z., B.A., K.B., Y.Y., D.L.E., N.B., G.P., D.M.B.-B., J.A.H., P.R.M., L.G.J.D., H.IJ., N.T., S.D.P., P.M.G., A.A.Ö., S.R.-F., P.C.E., L.G., F.S., M.M. and M.T.L.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-022-01458-9.

Correspondence and requests for materials should be addressed to Nicholas A. Coles.

Peer review information *Nature Human Behaviour* thanks David Mellor, Rainer Reisenzein, Jared McGinley and Quentin Gronau for their contribution to the peer review of this work.

[1]Center for the Study of Language and Information, Stanford University, Palo Alto, CA, USA. [2]Department of Psychology, Florida State University, Tallahassee, FL, USA. [3]Center for Change and Complexity in Learning, University of South Australia, Adelaide, South Australia, Australia. [4]Department of Psychology, University of Tennessee, Knoxville, TN, USA. [5]Department of Psychology, Alex Ekwueme Federal University Ndufu-Alike, Abakaliki, Nigeria. [6]School of Behavioural and Health Sciences, Australian Catholic University, Strathfield, New South Wales, Australia. [7]Department of Psychology, Ben-Gurion University, Be'er Sheva, Israel. [8]Zlotowski Center for Neuroscience, Ben-Gurion University, Be'er Sheva, Israel. [9]Busara Center for Behavioral Economics, Nairobi, Kenya. [10]Chapman University, Orange, CA, USA. [11]CLLE, Université de Toulouse, Toulouse, France. [12]Department of Psychology, Ege University, Izmir, Turkey. [13]Institute of Psychology, ELTE Eötvös Loránd University, Budapest, Hungary. [14]Department of Psychology, Federal University of Sergipe, São Cristóvão, Brazil. [15]Vernacular Languages Department, Federal University of Sergipe, São Cristóvão, Brazil. [16]Fundación para el Avance de la Psicología, Bogotá, Colombia. [17]Department of Humanities and Social Sciences, Indian Institute of Technology Guwahati, Guwahati, India. [18]Institute of Psychology, Jagiellonian University, Kraków, Poland. [19]Faculty of Arts and Science, Kyushu University, Fukuoka, Japan. [20]Graduate School of Human-Environment Studies, Kyushu University, Fukuoka, Japan. [21]School of Biomedical, Nutritional and Sport Sciences, Newcastle University, Newcastle upon Tyne, UK. [22]School of Health and Life Sciences, Teesside University, Middlesbrough, UK. [23]Occidental College, Los Angeles, CA, USA. [24]Center of Research on Cognition and Behavior, SWPS University of Social Sciences and Humanities, Warsaw, Poland. [25]Department of Psychology, Teesside University, Middlesbrough, UK. [26]Department of Psychology, UiT The Arctic University of Norway, Tromsø, Norway. [27]Department of Psychology, United States International University—Africa, Nairobi, Kenya. [28]Universidad Complutense, Madrid, Spain. [29]Universidad Nebrija, Madrid, Spain. [30]Departamento de Psicología Básica 1, Universidad Nacional de Educación a Distancia (UNED), Madrid, Spain. [31]Programa de Psicología, Universidad del Rosario, Bogotá, Colombia. [32]LIP/PC2s, Université Grenoble Alpes, Grenoble, France. [33]Institut Universitaire de France, Paris, France. [34]University of Antioquia-UDEA, Medellín, Colombia. [35]Department of Psychological Science, University of California, Irvine, CA, USA. [36]Department of Psychology, University of Fribourg, Fribourg, Switzerland. [37]Department of Psychology, Üsküdar University, İstanbul, Turkey. [38]FOM University of Applied Sciences, Essen, Germany. [39]Leibniz-Institut für Wissensmedien, Tübingen, Germany. [40]Department of Psychology, University of Michigan, Ann Arbor, MI, USA. [41]Department of Psychology, University of Würzburg, Würzburg, Germany. [42]Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice, Venice, Italy. [43]Department of Medical and Surgical Sciences, 'Magna Graecia' University of Catanzaro, Catanzaro, Italy. ✉e-mail: ncoles@stanford.edu

nature portfolio

	Corresponding author(s):	Nicholas A. Coles
	Last updated by author(s):	August 24, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Data were collected using Qualtrics. Materials for recreating the data collection survey in Qualtrics are publicly available at https://osf.io/ac3t2/.
Data analysis	Data were analyzed in R. All analysis code is publicly available at https://osf.io/ac3t2/

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Full data are publicly available at https://osf.io/ac3t2/

Human research participants

Policy information about <u>studies involving human research participants and Sex and Gender in Research.</u>

Reporting on sex and gender	Sex and gender were not considered in the development of the study or analysis plan because our team did not specify any a-priori hypotheses about these variables.
	For descriptive purposes, participants were asked to self-report their gender. Of the 3,878 participants who provided complete responses, 71% reported they are a male and 28% reported they are female. 1% of participants chose a different response option (i.e., choosing to decline to respond or indicating that they are transgender male, transgender female, gender variant/non-conforming, or an unlisted option.) These data are publicly available https://osf.io/ac3t2/.
Population characteristics	3,878 participants from 19 countries provided complete responses (age M = 18.52, SD = 0.96). Self-reported gender is described above.
Recruitment	26 research teams recruited participants using varied methods. Depending on the research site, participants completed the study on a completely volunteer basis, for partial course credit, extra credit, entrance into lottery (e.g., for a gift box), a prize (e.g., a pen), or money ($0.75-5 USD).
	Our sample is best conceptualized as a multi-country convenience sample. Thus, it is not clear if findings will generalize to other populations.
Ethics oversight	Each research group received approval from their local Ethics Committee or IRB to conduct the study (e.g., University of Tennessee IRB-19-05313-XM), indicated that their institution does not require approval for the researchers to conduct this type of research, or indicated that the current study is covered by a preexisting approval.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Quantitative experimental
Research sample	We collected a non-representative sample of 3,878 participants from 19 countries (Mage = 26.6; SDage = 10.6; 71% women, 28% men, 1% other).
	Sample size was determined--a-priori--by a power analysis (described below)
Sampling strategy	Convenience sampling was used, and sample size was determined--a-priori--by a power analysis.
	The power analysis was performed via a linear multilevel modeling simulation. We randomly generated normally-distributed data for 96 participants from 22 research groups. Effect size estimates for the hypothesized effects of pose (d = 0.39), stimuli presence (d = 0.68), and the pose by stimuli presence interaction (d = 0.29) were estimated from two pilot studies. All other effects were set to zero.
	Based on two pilot studies, we simulated random intercepts for participants with SD = 0.70. We did not simulate random slopes for participants. Random slopes for research groups were simulated based on the values from a previous many-lab failure-to-replicate. For hypothesized effects, we specified conservative random slopes estimates based on the standard deviation of the meta-analytic effect size from a previous many-labs failure-to-replicate (SD = 0.28). For effects we expected to be zero, we specified random slopes based on the random slope from a previous many-labs failure-to-replicate ($\tau 2 \approx 0$). However, due to convergence issues, the research groups random slope for the facial feedback task factor was removed. Residual variance was set to 0.60 based on the estimates from two pilot studies.
	Results from this power simulation indicated that over 95% power for all the hypothesized effects could be obtained with at least 1,584 participants. However, based on a third pilot study, we estimated that 44% of participants would not meet our strict inclusion criteria. Consequently, we planned to stop collecting data once one of the following conditions were met: (a) 22 labs had collected 105 participants, or (b) at least six months had elapsed since the start of data collection and we had at least 2,281 participants.

Data collection	Data were collected online via Qualtrics.
Timing	Data were collected between January 11 and May 17, 2021

Data exclusions

Exclusion criteria were specified a-priori.

We had complete responses from 3,878 participants. For the primary analyses, we excluded participants if they: failed an attention check (n = 660; 17% fail rate), completed the study on a mobile device (n = 132; 3% fail rate), reported deviating from the pose instructions (n = 35; 1% fail rate), reported that their posed happy expression did not match an image of an actor completing the task correctly (n = 108, 3% fail rate), were judged to be very distracted (n = 115, 3% fail rate), and/or exhibited any awareness of the study hypothesis (n = 1803, 46% fail rate). Some participants were excluded for failing to meet more than one inclusion criteria.

Non-participation

The online study link was opened 5,633 times--but complete responses were only provided in 3,878 of those cases. Because the study was conducted online, we could not collect information regarding why people dropped out of the study. In most cases, however, it appears that people dropped out while viewing the consent form and/or additional information about the study requirements.

Randomization

The study used a 2 (Pose: happy or neutral) x 3 (Facial Movement Task: facial mimicry, voluntary facial action, 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.

Random assignment (for factors manipulated between-participants) and ordering (for factors manipulated within-participants) was conducted through Qualtric's random assignment feature.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	Antibodies
☒	Eukaryotic cell lines
☒	Palaeontology and archaeology
☒	Animals and other organisms
☒	Clinical data
☒	Dual use research of concern

Methods

n/a	Involved in the study
☒	ChIP-seq
☒	Flow cytometry
☒	MRI-based neuroimaging