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

Nicholas Coles, David March, Fernando Marmolejo-Ramos, Jeff Larsen,
Nwadiogo Arinze, Izuchukwu Ndukaihe, Megan Willis, Francesco Foroni, Niv
Reggev, Aviv Mokady, et al.

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[6], Megan L. Willis[7], Francesco Foroni[7], Niv Reggev[8,9], Aviv Mokady[8], Patrick S. Forscher[10], John F. Hunter[11], Gwenaël Kaminski[12], Elif Yüvrük[13], Aycan Kapucu[13], Tamás Nagy[14], Nandor Hajdu[14], Julian Tejada[15], Raquel M. K. Freitag[16], Danilo Zambrano[17], Bidisha Som[18], Balazs Aczel[14], Krystian Barzykowski[19], Sylwia Adamus[19], Katarzyna Filip[19], Yuki Yamada[20], Ayumi Ikeda[21], Daniel L. Eaves[22,23], Carmel A. Levitan[24], Sydney Leiweke[24,25], Michal Parzuchowski[26], Natalie Butcher[27], Gerit Pfuhl[28], Dana M. Basnight-Brown[29], José A. Hinojosa[30,31], Pedro R. Montoro[32], Lady G. Javela D.[33], Kevin Vezirian[34], Hans IJzerman[35,36], Natalia Trujillo[37], Sarah D. Pressman[38], Pascal M. Gygax[39], Asil A. Özdoğru[40], Susana Ruiz-Fernandez[41,42,43], Phoebe C. Ellsworth[44], Lowell Gaertner[4], Fritz Strack[45], Marco Marozzi[46], Marco Tullio Liuzza[47]

[1]Stanford University, Center for the Study of Language and Information, USA, [2]Florida State University, Department of Psychology, USA, [3]University of South Australia, Center for Change and Complexity in Learning, Australia, [4]University of Tennessee, Department of Psychology, USA, [5]Alex Ekwueme Federal University Ndufu-Alike Nigeria, Department of Psychology, Nigeria, [6]Alex Ekwueme Federal University Ndufu-Alike, Department of Psychology, Nigeria, [7]Australian Catholic University, School of Behavioural and Health Sciences, Australia, [8]Ben-Gurion University, Department of Psychology, Israel, [9]Ben-Gurion University, Zlotowski Center for Neuroscience, Israel, [10]Busara Center for Behavioral Economics, Nairobi, Kenya, [11]Chapman University, USA, [12]CLLE, Université de Toulouse, CNRS-UMR, 5263, Toulouse, France, [13]Ege University, Department of Psychology, Turkey, [14]ELTE Eötvös Loránd University, Institute of Psychology, Hungary, [15]Federal University of Sergipe, Department of Psychology , Brazil, [16]Federal University of Sergipe, Vernacular Languages Department, Brazil, [17]Fundación para el Avance de la Psicología, Colombia, [18]Indian Institute of Technology Guwahati, Department of Humanities and Social Sciences, India, [19]Institute of Psychology, Jagiellonian University, Poland, [20]Kyushu University, Faculty of Arts and Science, Japan, [21]Kyushu University, Graduate School of Human-Environment Studies, Japan, [22]Newcastle University, School of Biomedical, Nutritional and Sport Sciences, UK, [23]Teesside University, School of Health and Life Sciences, UK, [24]Occidental College, USA, [25]Undergraduate Research Center at Occidental College, [26]SWPS University of Social Sciences and Humanities, Sopot, Center of Research on Cognition and Behavior, Poland, [27]Teesside University, Department of Psychology, UK, [28]UiT The Arctic University of Norway, Department of Psychology, Norway, [29]United States International University - Africa, Department of Psychology, Kenya, [30]Universidad Complutense, Spain, [31]Universidad Nebrija, Spain, [32]Universidad Nacional de Educación a Distancia (UNED), Departamento de Psicología Básica 1, Spain, [33]Universidad del Rosario, Programa de Psicología, Colombia, [34]Université Grenoble Alpes, France, [35]Université Grenoble Alpes, LIP/PC2s, France,

[36]Institut Universitaire de France, France, [37]University of Antioquia-UDEA, Colombia, [38]University of California, Irvine, Department of Psychological Science, USA, [39]University of Fribourg, Department of Psychology, Switzerland, [40]Üsküdar University, Department of Psychology, Turkey, [41]FOM University of Applied Sciences, Germany, [42]Leibniz-Institut für Wissensmedien, Tübingen, Germany, [43]Eberhard Karls University, LEAD Research Network, Tübingen, Germany, [44]University of Michigan, Department of Psychology, USA, [45]University of Würzburg, Department of Psychology, Germany, [46]Ca' Foscari University of Venice, Department of Environmental Sciences, Informatics and Statistics, Italy, [47]"Magna Graecia" University of Catanzaro, Department of Medical and Surgical Sciences, Italy

*Corresponding Author. e-mail: ncoles@stanford.edu

Abstract

Following theories of emotional embodiment, the facial feedback hypothesis suggests that an individual's subjective experience of emotion is 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.

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

The facial feedback hypothesis suggests that an individual's emotional experience is influenced by their facial expressions. For example, smiling should typically make an individual feel happier and frowning should make them feel sadder. Researchers suggest that these effects emerge because facial expressions provide sensorimotor feedback that (a) contributes to the sensation of an emotion[1,2], (b) serves as a cue that individuals use to make sense of ongoing emotional feelings[3,4], (c) influences other emotion-related bodily responses[5,6], and/or (d) 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, 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 participants in this pen-in-mouth study were presumably unaware 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 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 behavior[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 only occur when participants are aware 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 only emerge 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 only influences certain types of emotional experiences. Some researchers distinguish between self- versus 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, 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]. 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. Consequently, we 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: (a) specify our beliefs regarding when facial feedback effects, if real, should most reliably emerge, (b) determine the best way(s) to test those beliefs, and (c) 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 vs. neutral expressions. Based on 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) if 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) 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 (see Supplementary Figure 1).

To provide an easy-to-follow task that would produce more prototypical facial expressions, we used a facial mimicry paradigm, wherein 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 technique[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 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 vs. neutral facial expressions. Furthermore, we hypothesized that the magnitude of this effect would be similar across tasks that produce less (voluntary facial action technique and pen-in-mouth tasks) vs. more (mimicry task) prototypical expression of happiness. We also expected that facial feedback effects would be smaller in the absence vs. 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; see Supplementary Information) confirmed these predictions. A third pilot study conducted after initial Stage I acceptance (n = 119; see Supplementary Information) provided preliminary evidence in favor 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 (Version 4.1.2)[75]. For frequentist analyses, we fit mixed-effect models using the *lme4* package[76]. Some of these 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 ANOVA-like *F*-values with Satterthwaite degrees of freedom[77]. When we observed higher-order interactions, we used the *emmeans* package to decompose them using simple effect tests and pairwise contrasts[78]. 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 with medium Cauchy priors (r scale = $1/2$) on the alternative hypotheses and default Markov chain Monte Carlo settings[79]. We also performed sensitivity analyses with wide (r scale = $\sqrt{2}/2$) and ultrawide (r scale = 1) priors, and we thus report a range of Bayes Factors. For tests of main effects, interactions, and simple effects, we computed Bayes Factors by comparing models containing vs. excluding the terms representing the tested effect.

Participants

We made two minor deviations from the pre-registered sampling plan. First, due to constrains 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 vs. online data collection. Second, we had 80 fewer participants than we initially planned for our primary analyses.

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). We stopped data collection once at least 22 research groups had each collected at least 105 participants, totaling 3,878 participants from 26 groups (see Figure 1; M_{age} = 26.6; SD_{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 their posed expression did not match an image of an actor completing the task correctly (3%), indicated they were very distracted (3%), or exhibited any awareness of the study hypothesis (46%). (For country-specific exclusion criteria rates, see 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 (i.e., 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 in (a) the presence vs. absence of emotional stimuli and (b) after posing happy vs. neutral facial expressions. We also predicted that the effect of posed expressions on happiness would be larger in the presence vs. absence of positive stimuli. Following the study design (see Supplementary Information Figure 1), we modeled happiness reports with (a) 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, (b) all higher-order interactions, (c) random intercepts for participants and research groups, and (d) random slopes for research groups.

Participants reported higher levels of happiness in the presence vs. absence of positive images, $M_{diff} = 0.30$, 95% CI [0.12, 0.48], 5% scale range, $F(1, 22.65) = 10.67$, $p = .003$. However, Bayesian analyses were inconclusive, $BF_{10} = 0.71$-1.25. Participants also reported more happiness after posing happy vs. neutral expressions, $M_{diff} = 0.31$, 95% CI [0.21, 0.40], 5.17% scale range, $F(1, 24.34) = 39.86$, $p < .001$, $BF_{10} = 61.06$-102.63. Unexpectedly, the Pose effect was not significantly larger in the presence vs. absence of positive stimuli, $F(1, 29.50) = 1.33$, $p = .26$, $BF_{10} = 0.06$-0.13.

Unexpectedly, there was an interaction between Pose and Facial Movement Task, $F(2, 32.95) = 17.11$, $p < .001$, $BF_{10} = 34.13$-100.14 (see Figure 2). The effect of Pose on self-reported happiness was largest in the facial mimicry, $M_{diff} = 0.49$, 95% CI [0.36, 0.61], 8.17% scale range, $F(1, 28.62) = 57.55$, $p < .001$, $BF_{10} > 100$, and voluntary facial action tasks, $M_{diff} = 0.40$, 95% CI [0.23, 0.56], 6.67% scale range, $F(1, 25.48) = 22.93$, $p < .001$, $BF_{10} = 25.20$-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 vs. 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 (e.g., as a result of boredom)[80]. To test this, we refit the primary analysis model with an effect-coded Pose factor that compared happy pose to *filler* trials that participants completed. We focused on participants who were not exposed to positive images because these images were only shown 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 (see Figure 2).

Like the primary analyses, there was an interaction between Pose and Facial Movement Task, $F(2, 18.02) = 20.47$, $p < .001$. Participants reported higher levels of happiness after posing happy expressions vs. completing filler tasks in both the facial mimicry, $M_{diff} = 0.48$, 95% CI [0.29, 0.67], 8% scale range, $t(22.4) = 5.23$, $p < .001$, and voluntary facial action tasks, $M_{diff} = 0.20$, 95% CI [0.05, 0.36], 3.33% scale range, $t(19.6) = 2.69$, $p = .01$. In the pen-in-mouth task, participants reported *less* happiness after completing the happy vs. filler task, $M_{diff} = -0.15$, 95% CI [-0.28, 0.02], 2.5% scale range, $t(31.5) = 2.39$, $p = .02$.

Moderating role of pose quality

We next examined the moderating role of three indicators of the quality of posed expressions: participants' reports of the extent to which they followed pose instructions (*compliance ratings*), reported 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 (a) the indicator entered mean-centered, and (b) a term denoting its interaction with Pose. For each quality indicator, there was an interaction with Pose (see Figure 3). The effect of facial poses on happiness was larger among participants with higher compliance, $\beta = 0.08$, 95% CI [0.05, 0.12], $t(1482.63) = 4.33$, $p < .001$, similarity, $\beta = 0.03$, 95% CI [0.01, 0.06], $t(1358.62) = 3.37$, $p < .001$, and genuineness ratings, $\beta = 0.08$, 95% CI [0.06, 0.09], $t(1420.95) = 10.57$, $p < .001$.

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 modeled each quality indicator with (a) Facial Movement Task and Stimuli Presence entered as effect-coded factors, (b) random intercepts for research groups, and (c) 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, $M_{\text{diff}} = 0.03$, 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 ($M = 6.45$, $SD = 1.07$) vs. voluntary facial action ($M = 6.57$, $SD = 0.93$), $M_{\text{diff}} = -0.15$, 95% CI [-0.28, -0.02], 2.5% scale range, $t(23.5) = -2.47$, $p = .02$, and pen-in-mouth tasks ($M = 6.68$, $SD = 1.01$), $M_{\text{diff}} = -0.25$, 95% CI [-0.37, -0.14], 4.17% scale range, $t(22.8) = -4.49$, $p < .001$. Compliance ratings were also slightly higher in the pen-in-mouth vs. voluntary facial action tasks, $M_{\text{diff}} = 0.10$, 95% CI [-0.01, 0.21], or 1.67% scale range, $t(21.9) = 1.96$, $p = .06$.

Likewise, similarity ratings varied by Facial Movement Task, $F(2, 40.12) = 7.35$, $p = 0.002$, but not Stimuli Presence, $M_{\text{diff}} = -0.12$, 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 highest in the facial mimicry ($M = 5.30$, $SD = 1.36$) vs. voluntary facial action ($M = 5.09$, $SD = 1.73$), $M_{\text{diff}} = 0.23$, 95% CI [0.03, 0.43], 3.83% scale range, $t(22.7) = 2.43$, $p = .02$, and pen-in-mouth tasks ($M = 5.07$, $SD = 1.61$), $M_{\text{diff}} = 0.24$, 95% CI [0.11, 0.36], 4% scale range, $t(194) = 3.63$, $p < .001$.

Genuineness ratings strongly varied by Facial Movement Task, $F(2, 13.69) = 82.56$, $p < .001$. Genuineness ratings were substantially lower in the pen-in-mouth ($M = 2.98$, $SD = 1.89$) vs. facial mimicry ($M = 4.15$, $SD = 1.92$), $M_{\text{diff}} = -1.15$, 95% CI [-1.34, -0.97], 19.17% scale range, $t(23.85) = 12.85$, $p < .001$, and voluntary facial action tasks ($M = 3.91$, $SD = 2.00$), $M_{\text{diff}} = -0.89$, 95% CI [-1.12, -0.66], 14.83% scale range, $t(24.92) = 8.00$, $p < .001$. Genuineness ratings were also lower in the voluntary facial action vs. facial mimicry tasks, $M_{\text{diff}} = -0.26$, 95% CI [-0.48, -0.05], 4.33% scale range, $t(6.67) = -2.90$, $p = .02$. Participants also reported higher genuineness ratings in the presence ($M = 3.78$, $SD = 2.00$) vs. absence ($M = 3.57$, $SD = 2.00$) of positive images, $M_{\text{diff}} = 0.23$, 95% CI [0.11, 0.34], 3.83% scale range, $F(1, 1538.52) = 13.66$, $p < .001$.

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 modeled coder ratings of the extent to which they were aware with (a) Facial Movement Task and Stimuli Presence entered as effect-coded factors, (b) random intercepts for research groups, and (c) random slopes for research groups. Awareness scores varied by Facial Movement Task, $F(2, 19.70) = 13.54$, $p < .001$, with participant being less aware in the pen-in-mouth ($M = 1.75$, $SD = 1.41$) vs. voluntary facial action task ($M = 2.28$, $SD = 1.78$) $M_{diff} = -0.48$, 95% CI [-0.67, -0.29], 8.02% scale range, $t(24) = -5.19$, $p < .001$, and facial mimicry task ($M = 2.05$, $SD = 1.52$), $M_{diff} = -0.27$, 95% CI [-0.43, -0.11], 4.48% scale range, $t(24) = -5.19$, $p < .001$. Participants were also less aware in the facial mimicry vs. voluntary facial action task, $M_{diff} = -0.21$, 95% CI [-0.36, -0.07], 3.53% scale range, $t(39.4) = -2.97$, $p = .005$.

To test whether facial feedback effects are amplified by awareness of the study purpose, we modeled happiness reports with (a) Pose, Facial Movement Task, and Stimuli Presence entered as effect-coded factors, (b) awareness scores entered mean-centered, (c) a higher-order interaction term for Pose and awareness scores, (d) random intercepts for participants and research groups, and (e) research group random slopes for all terms other than awareness scores. 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], $t(22.74) = 7.55$, $p < .001$ (see Figure 3).

Body awareness

To examine the moderating role of body awareness, we re-ran our primary analysis model with (a) participants' responses on a body awareness measure entered mean-centered and (b) 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], $t(9.87) = 0.02$, $p = .99$ (see Figure 3).

Between-condition differences in other inclusion criteria

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

The probability that participants used the incorrect device did not vary by Facial Movement Tasks (96%, 96%, 97% pass rate in the facial mimicry, voluntary facial action, and pen-in-mouth tasks), $\chi^2(2) = 3.06$, $p = .22$, or Stimuli Presence (97% pass rate in the absence and presence of positive stimuli), $\chi^2(1) = 0.11$, $p = .74$. Likewise, the probability that participants failed attention checks did not vary between Facial Feedback Tasks (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 rate in the absence and presence of positive stimuli), $\chi^2(1) = 2.54$, $p = .11$.

We also tested for between-condition differences in coder ratings of the extent to which participants were distracted using linear mixed-effect regression with (a) Facial Movement Task and Stimuli Presence entered as effect-coded factors, (b) random intercepts for research groups, and (c) random slopes for research groups. Distraction scores did not significantly vary between the facial mimicry ($M = 2.01$, $SD = 1.17$), voluntary facial action ($M = 1.92$, $SD = 1.14$), and pen-in-mouth ($M = 1.92$, $SD = 1.14$) tasks, $F(2, 18.57) = 2.45$, $p = .11$. Distraction scores also did not vary in the absence ($M = 1.94$, $SD = 1.15$) vs. presence ($M = 1.96$, $SD = 1.16$) of positive stimuli, $F(1, 900.52) = 0.02$, $p = .90$.

Anger and anxiety

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

Happy vs. neutral facial expression poses did not significantly decrease feelings of anger, $M_{diff} = -0.02$, 95% CI [-0.07, 0.03], 0.33% scale range, $F(1, 20.71) = 0.85$, $p = .37$, or anxiety, $M_{diff} = -0.01$, 95% CI [-0.07, 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, $F(2, 58.20) = 5.18$, $p = 0.008$, did differ by Facial Feedback Task. Participants reported higher levels of anger in the pen-in-mouth vs. facial mimicry, $M_{diff} = 0.14$, 95% CI [0.03, 0.24], 2.33% scale range, $t(24.2) = 2.64$, $p = .01$, and voluntary facial action tasks, $M_{diff} = 0.12$, 95% CI [0.02, 0.21], 2% scale range, $t(31.6) = 2.40$, $p = .02$. Similarly, participants reported more anxiety in the pen-in-mouth vs. facial mimicry, $M_{diff} = 0.13$, 95% CI [0.02, 0.24], 2.17% scale range, $t(51.6) = 2.35$, $p = 0.02$, and voluntary facial action tasks, $M_{diff} = 0.17$, 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 (see Supplementary Information).

Exploratory Analyses

For all analyses, we pre-registered 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[81]. Sensitivity analyses without (vs. with) random slopes generally yielded identical inferences, except for the simple effect of Pose in the pen-in-mouth task. After removing 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, $F(1, 1498) = 2.78$, $p = .095$—but an exploratory one-sided test was (one-sided $p < .05$). However, Bayesian analyses were inconclusive, $BF_{10} = 0.46\text{-}0.96$. 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], or 2.33% scale range, $F(1, 3872) = 16.37$, $p < .001$, $BF_{10} > 100$.

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 (vs. in-person); focused on feelings of happiness (vs. amusement); used a different cover story; had participants pose expressions for a relatively short duration (5 seconds); and did not instruct 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 culturally diverse sample (see Supplementary Figure 2 for 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 both in 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 to 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[82].

Consistent with our predictions and a previous meta-analysis[59], facial feedback effects, when present, were small (see Supplementary Figure 3 for the distribution of mean difference scores). Nonetheless, these effects were similar in size to the effect of mildly positive photos on happiness—i.e., 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[2,83]. 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[47,84,85].

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 5 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[86]. 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 pre-registered 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 pre-registered that this would be considered a less stringent test of the facial feedback hypothesis.

Although it is less clear if 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[82], but they may involve both inferential processes (e.g., people inferring they are happy because they are smiling)[45,46] and non-inferential processes (e.g., 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 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 (a) amplified when the task activates muscles typically associated with an emotional state, and (b) 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[71], electromyography[87], sonography[88], or thermography[89].

To conclude, our adversarial collaboration was, in part, inspired by conflicting narratives about the validity of the facial feedback hypothesis. We began the collaboration after (a) a large team of researchers failed to replicate a seminal demonstration of facial feedback effects using a pen-in-mouth task[40], but (b) 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 byproduct of emotion processes. Consistent with theories that posit 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 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. At the time of Stage I 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 constrains created by the Coronavirus Disease 2019 (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, participants were asked to confirm that they (a) have a clean pen or pencil nearby that they are willing to place in their mouths, (b) are completing the study on a desktop computer or laptop (details regarding participants' operating systems were automatically recorded to confirm), and (c) are in a setting with minimal distractions.

 Participants were told that the study investigates 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 is 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, participants were asked to pose happy and neutral facial expressions in randomized order through either 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, participants were provided with an on-screen timer during all tasks.

 After each task (including filler tasks), participants completed a simple filler arithmetic problem and the Discrete Emotions Questionnaire's 4-item happiness subscale, which asked 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")[66]. Participants also completed two items measuring anxiety (worry, nervous). To further obscure the purpose of the study, 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, 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 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, participants were shown a 2 x 2 image matrix of actors posing happy expressions. Participants were then instructed to either mimic these expressions (happy condition) or maintain a blank expression (neutral condition). Importantly, having 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 mimicry trials. The expression matrix was displayed for at least 5 seconds, and participants indicated when they are ready to perform the task. In the voluntary facial action technique condition, participants were instructed to either move the corner of their lips up towards their ear 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, 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, participants were instructed to maintain the poses for 5 seconds, the approximate duration of spontaneous happiness expressions[67].

After completing the five movement tasks, 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 participants' open-ended responses in randomized order. Based on the open-ended responses, 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, participants completed a short demographic form and the Body Awareness Questionnaire[68]. Participants then answered several questions related to the quality of their data. First, 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, 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, participants were asked to watch themselves repeat the task (e.g., 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, participants were asked to describe any issues that may have

compromised the quality of their data (e.g., 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"). 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 participants' posed expression 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 region may also have lacked a web camera. Furthermore, researchers are still debating whether awareness of overt video recording interferes with facial feedback effects[49,50,59,69]. Nevertheless, pilot study recordings and self-reports confirmed that almost all participants successfully posed the target facial expressions (see Supplementary Information).

Materials

In the facial mimicry task, participants all viewed the same 2 x 2 image matrix of actors posing happy facial expressions from the Extended Cohn-Kanade Dataset[70]. All four actors posed prototypical facial expressions of happiness, as confirmed by coders trained in the Facial Action Coding System[71]. An image matrix of actors, as opposed to a single image, was used so that participants had multiple examples of the movement and were provided with more options for a suitable facial model. In the pen-in-mouth task, 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 (e.g., photos of dogs, flowers, kittens, and rainbows). Multiple photos (as opposed to a single photo) were used to increase the probability that 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[72] that were separately rated on how good and bad they were[73]. Results from three pilot studies confirmed that these images successfully elicited feelings of happiness (see Supplementary Information). Due to potential cross-cultural differences in what types of photos elicit happiness (e.g., dog photos can be expected to elicit happiness in many Western cultures, but not all African cultures), each lab was permitted to replace photos with more culturally-appropriate positive photos. For non-

English speaking data collection sites, experiment materials were translated into the local language.

Primary Analyses

Due to the nested nature of the data (e.g., ratings nested within individuals, nested within research groups), we used linear multilevel modeling. More specifically, happiness reports were modeled with (a) Pose, Facial Movement Task, and Stimuli Presence entered as factors, (b) random intercepts for research groups and participants, and (c) 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 (i.e., received an awareness score over 1 from two independent coders), (2) disclosed that they were very distracted during the study (i.e., received an average distraction score above 5 from two independent coders), (3) did not complete the study on a desktop computer or laptop, (4) indicated they did not follow the pose instructions, (5) indicated their expression during the happy pose task did not at all match the image of an actor completing the task correctly, or (6) failed attention checks. These stringent exclusion criteria were added after failing 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 if the exclusion criteria interact with Pose to influence happiness reports. We also examined if these exclusion criterion variables varied as a function of Facial Movement Task and Stimuli Presence.

To examine the alternative explanation that doing something (e.g., posing a happy facial expression) may simply be more enjoyable than doing nothing (e.g., posing a neutral facial expression), we also re-ran our primary analysis 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 vs. in a lab[74], we recorded the mode of data collection and planned to re-run our primary analysis with data collection mode included as a moderator. However, we noted that this analysis may be confounded by (a) whether the research group is a proponent vs. critic of the facial feedback hypothesis (i.e., proponents may be more likely to

collect data in the laboratory) and (b) the region of data collection (i.e., research groups in regions with fewer COVID-19 cases may be more likely to collect data in the laboratory).

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

Power Simulation

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 pilot studies 1 and 2 (see 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 exclusion criteria that left our original predictions unchanged.

Based on two pilot studies, we simulated random intercepts for participants with $SD = 0.70$. We did not simulate random slopes for participants since there are only two observations within each participant, which would likely lead to convergence issues. Random slopes for research groups were simulated based on the values from the previous many-lab failure-to-replicate[40]. For hypothesized effects, we specified conservative random slopes estimates based on the standard deviation of their meta-analytic effect size from the 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 the 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 pilot studies 1 and 2.

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, based on pilot study 3, we estimated that 44% of participants would not meet our strict inclusion criteria, leading to a desired sample of 2,281. Consequently, we planned to stop collecting data once one of the following conditions are 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. We planned for a minimum of 22 labs will collect data for this project, although additional labs with pending ethics approval were allowed to join the project later.

Data availability

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

Code availability

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

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 & 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.

Tables

Table 1

Research questions and associated hypotheses.

Research Question	Hypothesis
Does posing happy vs. neutral expression cause people to feel happier?	Participants will report experiencing more happiness when posing happy vs. neutral facial expressions.
Do happy facial poses only influence feelings of happiness if they resemble a natural expression of happiness?	The difference in self-reported happiness when posing happy vs. neutral facial expressions will *not* be larger during tasks that produce more natural expressions of happiness.
Can facial poses initiate emotional experience in otherwise neutral scenarios, or can they only modulate ongoing emotional experiences?	The difference in self-reported happiness when posing happy vs. neutral facial expressions will be significant both in 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 vs. neutral facial.

Figure Legends/Captions

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.
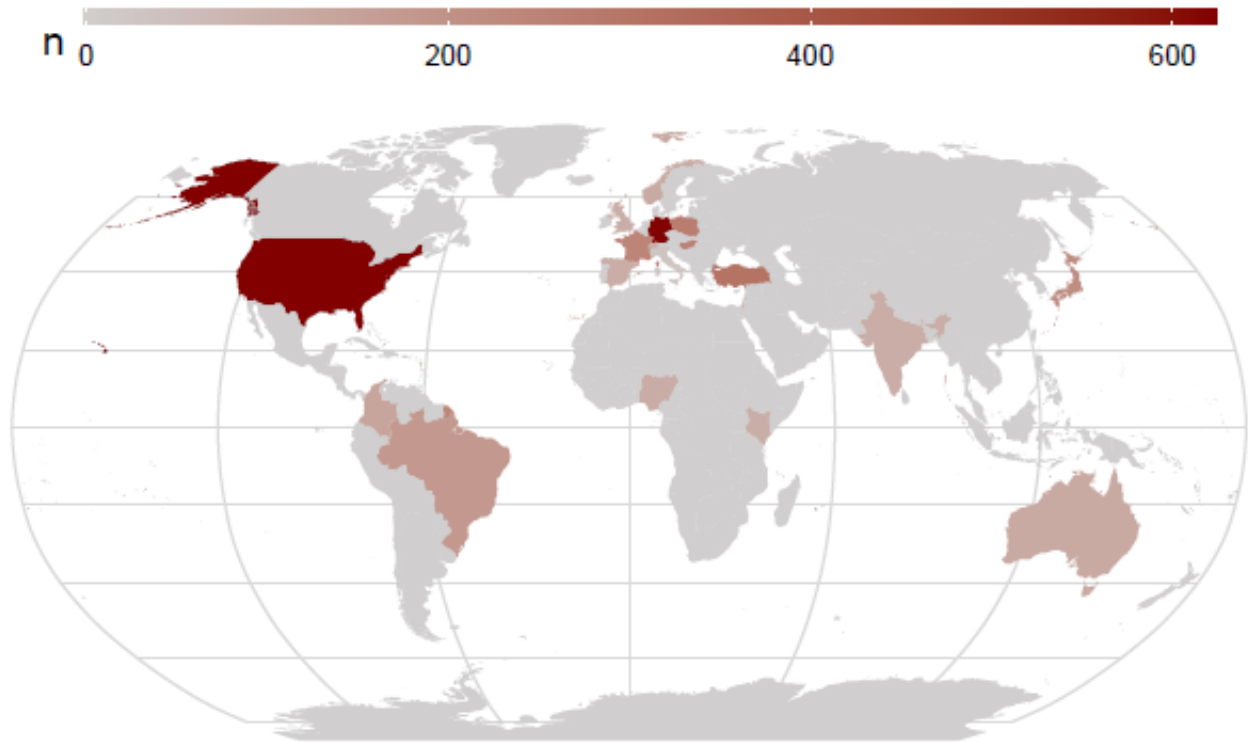
Fig. 2: Effects of facial expression poses and filler tasks on self-reported happiness in each study condition.

Self-reported happiness (1 = "not at all" to 7 "an extreme amount") after participants posed happy facial expressions, posed neutral facial expressions, or completed filler tasks. Panel columns indicate whether participants completed the facial mimicry, voluntary facial action, or pen-in-mouth task. Panel rows indicate whether positive images were absent or present during the facial pose tasks. Grey points represent jittered participant observations. Blue error bars represent mean ± 1 standard error. Condition-specific sample sizes (n), means (M), and standard deviations (SD) are reported.

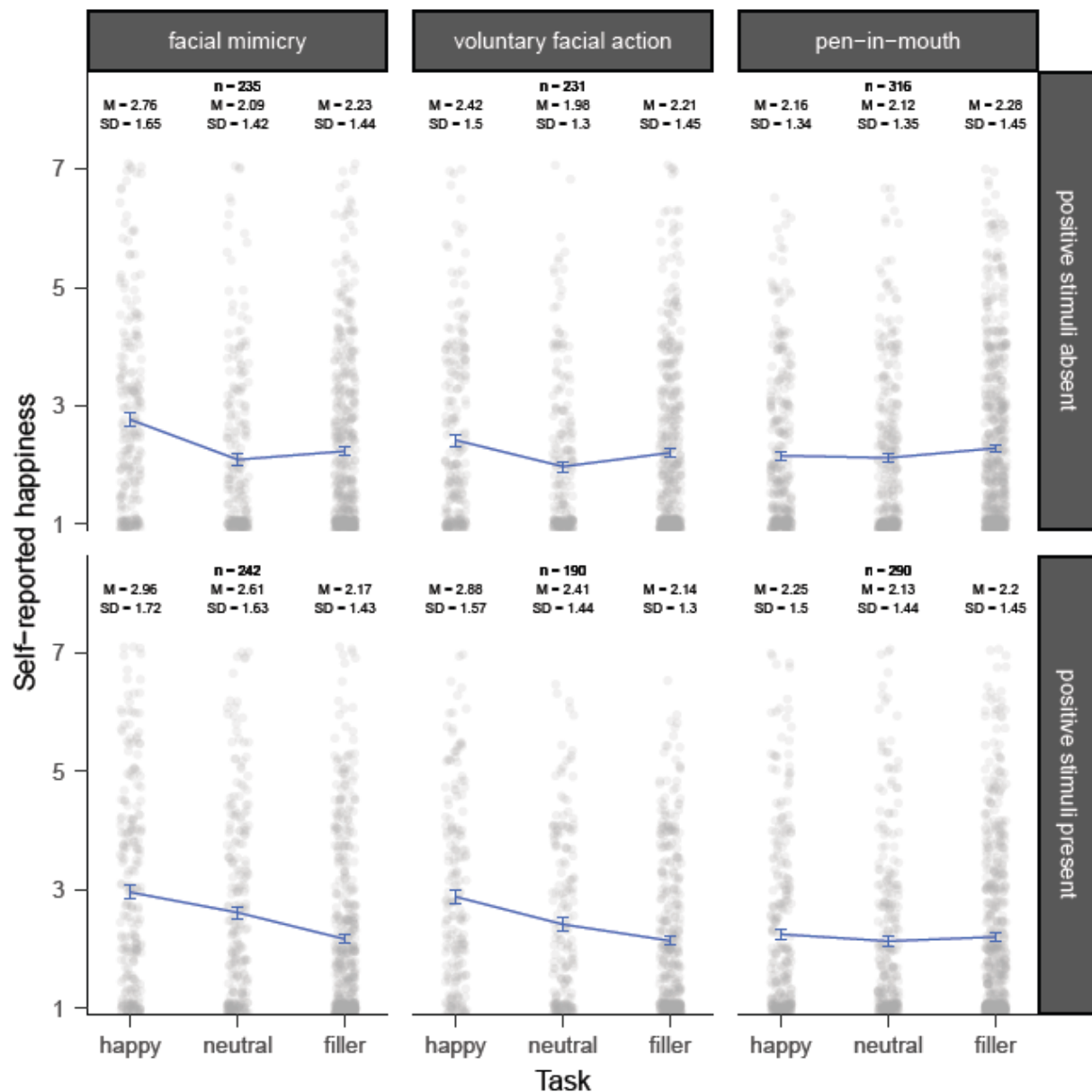**Fig. 3: Potential moderators of facial feedback effects.**

The change in happiness (y-axis) when participants posed happy vs. neutral expressions were moderated by compliance, similarity, genuineness, and hypothesis awareness ratings—but not body awareness ratings (x-axes). Grey points represent jittered participant observations. Blue lines represent estimated linear relationships.
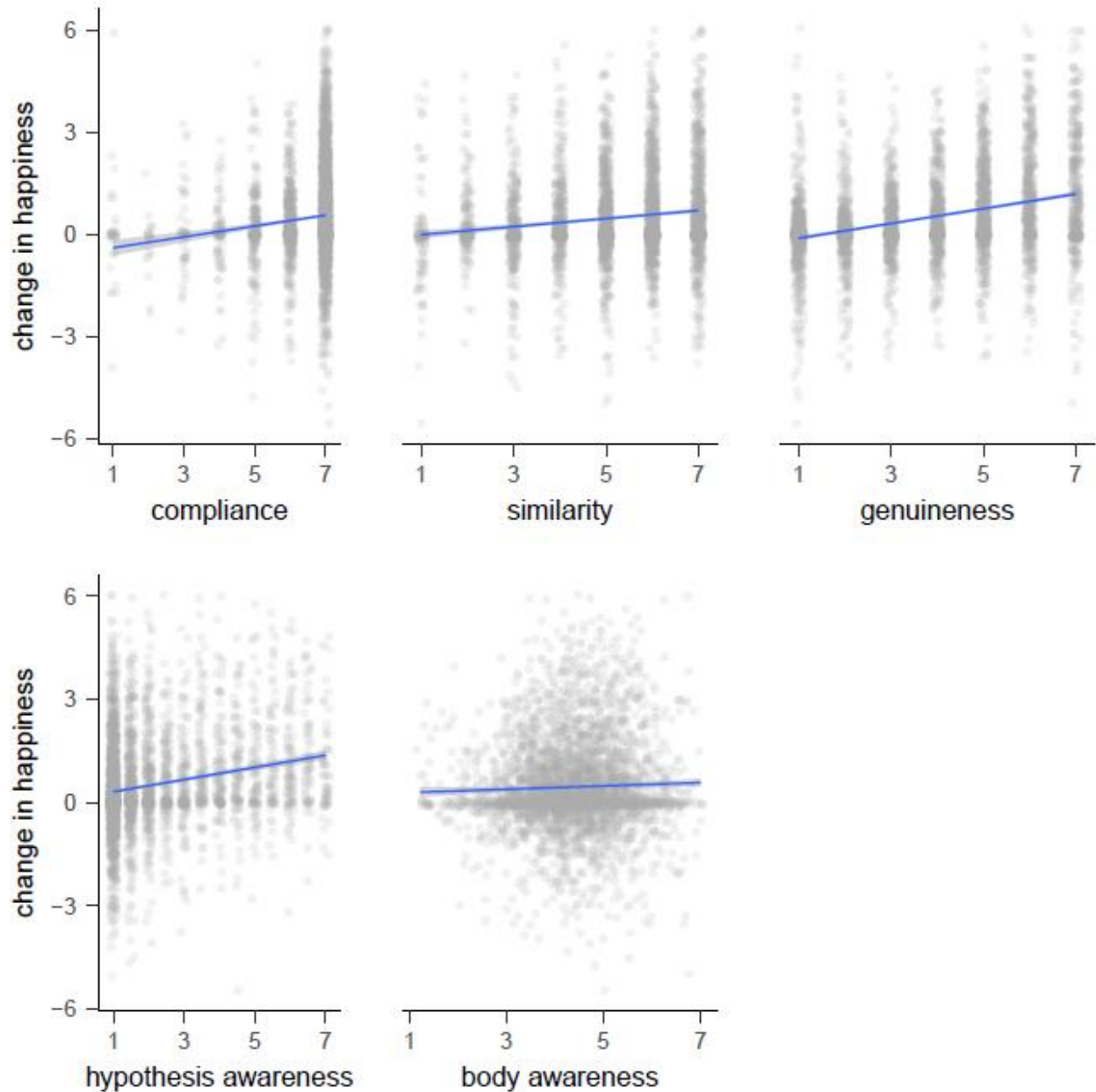
References

1. Zajonc, R. B. The primacy of affect. *Am. Psychol.* **40**, 849–850 (1985).

2. Tomkins, S. *Affect Imagery Consciousness: Vol 1 The Positive Affects*. (Springer, 1962).

3. Laird, J. D. & Crosby, M. Individual differences in self-attribution of emotion. in *Thought and Feeling: Cognitive Alteration of Feeling States* (eds. London, H. & Nisbett, R. E.) 44–59 (Transaction Publishers, 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 and Company, 1915).

14. Sherrington, C. S. Experiments on the value of vascular and visceral factors for the

genesis of emotion. *Proc. R. Soc. London* **66**, 390–403 (1899).

15. Ansfield, M. E. Smiling when distressed: When a smile is a frown turned upside down. *Personal. 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. *Dermatologic 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. *Dermatologic Surg.* **32**, 645–649 (2006).

26. Fromage, G. Exploring the effects of botulinum toxin type A injections on depression. *Aesthetic 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. *Dermatologic 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. Botulinum toxin treatment in depression. in *Practical Psychodermatology* (eds. Bewley, A., Taylor, R. E., Reichenberg, J. S. & Magid, M.) 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**, 1–5 (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 Press, 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. The process of emotional experience: A self-perception theory. in *Review of Personality and Social Psychology: Emotion* (ed. Clark, M. S.) 213–234 (1992).

47. Ekman, P. Biological and cultural contributions to body and facial movement. 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. Personal. Soc. Psychol.* **114**, 1–26 (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**, 769–74 (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. Personal. 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. 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).

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

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

69. 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).

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

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

72. Lang, P. & Bradley, M. M. The International Affective Picture System (IAPS) in the study of emotion and attention. in *Handbook of Emotion Elicitation and Assessment* (eds. Coan, J. A. & Allen, J. J. B.) 29–46 (Oxford University Press, USA, 2007).

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

74. 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).

75. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2021). URL https://www.Rproject.org/.

76. 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).

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

78. Russell V. Lenth. *Estimated marginal means, aka least-squares means. R package version 1.7.2* (2022).

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

80. Wilson, T. D. *et al.* Just think: The challenges of the disengaged mind. *Science (80-.).* **345**, 75–77 (2014).

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

82. 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).

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

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

85. 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).

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

87.   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).

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

89.   Clay-Warner, J. & Robinson, D. T. Infrared thermography as a measure of emotion response. *Emot. Rev.* **7**, 1–6 (2014).

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

Supplementary Information

Pilot Studies 1 and 2

Pilot studies 1 and 2 were run prior to initial Stage 1 acceptance and featured a 2 (Pose: happy or neutral) x 2 (Facial Movement Task: facial mimicry or voluntary facial action technique) design. The studies differed from the procedures described in the main manuscript in the following ways: (1) the pen-in-mouth task was not included because this condition was added after peer-review, (2) participants were run in a laboratory and interviewed by an experimenter, (3) participants were covertly recorded in pilot study 1, (4) participants did not complete attention check items, and (5) there were three (as opposed to two) filler trials.

The two pre-registered pilot studies differed in whether participants viewed emotional stimuli while engaging in the critical poses, with the first pilot testing initiating effects (i.e., no stimuli present) and the second pilot testing modulating effects (i.e., stimuli present; see Supplementary Figure 1). Together, data were collected from 206 participants (67% female, 33% male; age $M = 18.52$, $SD = 0.96$) who completed the experiment in exchange for partial course credit. Patterns of data and inferences were identical across the two studies, so pooled analyses are reported unless otherwise noted. All participants provided informed consent.

Supplementary Fig. 1: Experimental design of the main and pilot studies.

The main study and pilot study 3 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. Blue boxes represent each factor in the experimental design, and white and grey boxes represent the cross between factors. Pilot studies 1 and 2 did not examine the pen-in-mouth task, which was added after Stage 1 external review.
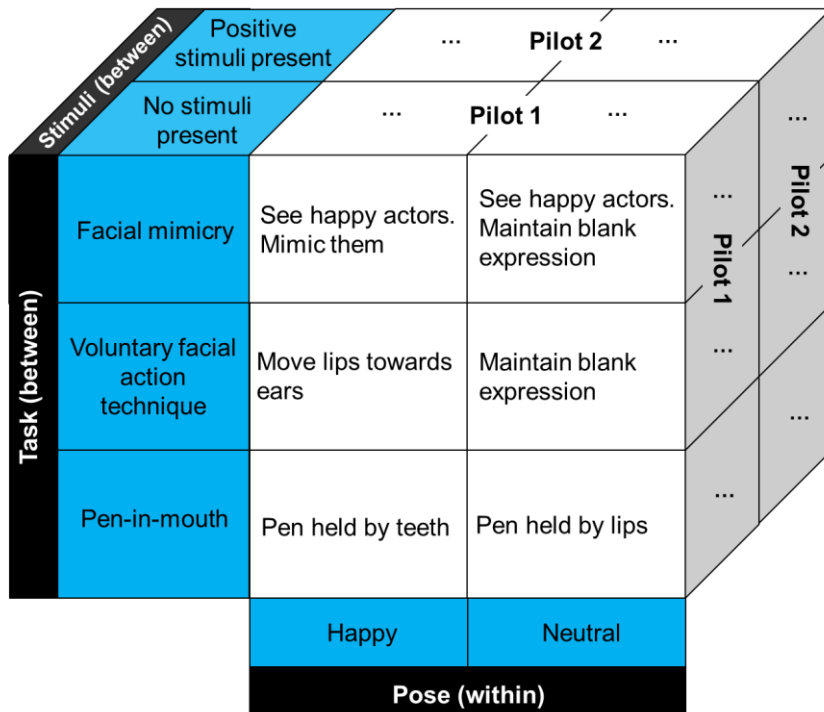
Test of the Facial Feedback Hypothesis

To examine whether facial feedback impacted self-reported happiness, a 2 (Pose: happy or neutral) x 2 (Facial Movement Task: facial mimicry or voluntary facial action technique) x 2 (Stimuli Presence: absent or present) mixed-effect ANOVA was fitted, with Pose included as a within-participant factor. Consistent with the facial feedback hypothesis, participants reported more happiness after posing happy ($M = 2.47$, $SD = 1.48$) vs. neutral expressions ($M = 1.93$, $SD = 1.18$), $F(1, 202) = 43.65$, $p < .001$, $\eta^2_G = 0.04$, 95% CI [0.01, 0.11]. No main effect for Facial Movement Task was detected, $F(1, 202) = .38$, $p = .54$, $\eta^2_G = 0.00$, 95% CI [0.00, 0.03]. There was a main effect for Stimuli Presence, wherein participants reported more happiness when

positive stimuli were present ($M = 2.65$, $SD = 1.54$) vs. absent ($M = 1.72$, $SD = 0.94$), $F(1, 202)$ $= 36.06$, $p < .001$, $\eta^2_G = 0.08$, 95% CI [0.02, 0.17].

Results also revealed an interaction between Pose and Stimuli Presence, $F(1, 202) = 5.79$, $p = .02$, $\eta^2_G = 0.01$, 95% CI [0.00, 0.05]. To decompose this interaction, 2 (Pose: happy or neutral) x 2 (Facial Movement Task: facial mimicry or voluntary facial action technique) mixed ANOVAs were separately fitted for each study. Effects of Pose were obtained both when stimuli were absent, $F(1, 98) = 15.56$, $p < .001$, $\eta^2_G = 0.03$, 95% CI [0.00, 0.13], and present, $F(1, 104) = 29.40$, $p < .001$, $\eta^2_G = 0.06$, 95% CI [0.00, 0.17]. These findings indicate that smiling both initiated and modulated happiness, and that smiling had an especially strong modulating effect. Some may argue that the inclusion of the mimicry task prevented a true test of the initiation hypothesis because images of smiling actors may elicit happiness. This seems unlikely because participants did not report more happiness in the mimicry vs. voluntary facial action technique tasks. Nevertheless, follow-up analyses excluding the mimicry condition confirmed that facial feedback can initiate emotional experience, $F(1, 50) = 6.76$, $p = .01$, $\eta^2_G = 0.04$, 95% CI [0.00, 0.19].

Although results are consistent with the facial feedback hypothesis, one alternative explanation is that participants found the smiling task to be less boring than the neutral task (perhaps because participants do not do anything with their bodies in this latter task). The filler trials allowed us to provide a post hoc test of this possibility because they required participants to perform affectively neutral bodily movements. We therefore compared happiness ratings during the happy vs. filler movement trials using a 2 (Movement Trial: smile or filler) x 2 (Stimuli Presence: absent or present) mixed-effect ANOVA. Consistent with the facial feedback hypothesis, results indicated that participants reported greater happiness after posing happy expressions ($M = 2.47$, $SD = 1.48$) vs. engaging in filler movements ($M = 1.86$, $SD = 1.01$), $F(1, 204) = 56.42$, $p < .001$, $\eta^2_G = 0.06$, 95% CI [0.01, 0.14]. Results also revealed an interaction between Movement Trial and Stimuli Presence, $F(1, 204) = 35.52$, $p < .001$, $\eta^2_G = 0.04$, 95% CI [0.00, 0.11]. To decompose this interaction, we re-examined the main effect of Movement Trial separately for each study. Results provided significant evidence of a Movement Trial effect when stimuli were present, $F(1, 105) = 63.47$, $p < .001$, $\eta^2_G = 0.14$, 95% CI [0.04, 0.27], but not absent, $F(1, 99) = 2.34$, $p = .13$, $\eta^2_G = 0.00$, 95% CI [0.00, 0.06].

Participant Awareness

Based on in-person funnel debriefings, experimenters rated the degree to which participants were aware of the purpose of the experiment (1 = "not at all aware" to 5 = "completely aware"). Results indicated that participants generally exhibited low awareness of the purpose of the experiment ($M = 1.54$, $SD = 0.96$), with 85% of participants characterized as not at all or slightly aware. To examine whether participant awareness varied across conditions, awareness ratings were modeled using a 2 (Facial Movement Task: facial mimicry or voluntary facial action technique) x 2 (Stimuli Presence: present or absent) ANOVA. Contrary to our prediction, results did not indicate that participants were more aware of the purpose of the experiment in the facial mimicry ($M = 1.47$, $SD = 0.85$) vs. voluntary facial action technique ($M = 1.61$, $SD = 1.05$) condition, $F(1, 202) = 1.04$, $p = .31$, $\eta^2_G = 0.00$, 95% CI [0.00, 0.03]. Unexpectedly, participants exhibited more awareness of the experiment's purpose when emotional stimuli were present ($M = 1.68$, $SD = 1.03$) as opposed to absent ($M = 1.39$, $SD = 0.86$), $F(1, 202) = 4.75$, $p = .03$, $\eta^2_G = 0.01$, 95% CI [0.00, 0.06]. No interaction between awareness and stimuli presence was detected, $F(1, 202) = 0.47$, $p = .49$, $\eta^2_G = 0.00$, 95% CI [0.00, 0.03].

Although results are consistent with the facial feedback hypothesis, it is possible that observed effects were driven by participants' awareness of the purpose of the experiment (e.g., demand characteristics). Confirmatory analyses were re-run using linear mixed-effect modeling with participant awareness included as a continuous moderator. Results from this analysis provided some evidence that facial feedback effects are larger when participants were more aware of the purpose of the experiment, pooled $F(1, 198) = 5.48$, $p = .02$; Study 1 $F(1, 96) = 0.66$, $p = .42$; Study 2 $F(1, 102) = 6.27$, $p = .01$. Consequently, all confirmatory analyses we re-run excluding participants who exhibited any degree of awareness (i.e., had an awareness score higher than 1). All aforementioned results were robust except the interaction between Facial Movement Task and Stimuli Presence. Taken together, the observation of a significant facial feedback effect in participants who were completely unaware of the purpose of the experiment indicates that awareness of the purpose of the experiment does not fully account for our results. At the same time, results from the moderator analysis suggest that being aware of the hypothesis can amplify facial feedback effects.

Quality of Posed Expressions

In pilot study 1, participants were covertly recorded to assess the quality of their posed expressions. For participants who consented for their videos to be analyzed (n = 80), video recordings of their happy and neutral posing trials were processed through Noldus FaceReader 7.0, which provided moment-to-moment ratings of expressed happiness (0 to 1)[91]. FaceReader failed to code videos from two participants, leaving a final sample of 78 pairs of videos.

Expressed happiness ratings were modeled using a 2 (Pose: happy or neutral) x 2 (Facial Movement Task: facial mimicry or voluntary facial action technique) mixed-effect ANOVA, with Pose included as a within-participant factor. As expected, participants expressed more happiness during the happy (*M* = .65, *SD* = .27) vs. neutral (*M* = .03, *SD* = .05) trials, $F(1, 76) = 454.61$, $p < .001$, $\eta^2_G = 0.75$, 95% CI [0.63, 0.83]. Participants also expressed more happiness in the facial mimicry (*M* = .39, *SD* = .39) vs. voluntary facial action technique (*M* = .28, *SD* = .34) condition, $F(1, 76) = 14.59$, $p < .001$, $\eta^2_G = 0.09$, 95% CI [0.01, 0.23]. These main effects were qualified by a significant interaction, wherein the difference in expressed happiness between the happy and neutral trials was larger in the facial mimicry condition, $F(1, 76) = 14.68$, $p < .001$, $\eta^2_G = 0.09$, 95% CI [0.01, 0.24]. These patterns of results are consistent with our prediction that the facial mimicry condition would produce more prototypical expressions of happiness than the voluntary facial action technique condition, although our results so far provide no evidence that the quality of poses influence the magnitude of facial feedback effects. In addition, all but one participant expressed more happiness during the happy vs. neutral trial, indicating that most participants can successfully execute these two facial feedback tasks.

Individual Differences in Bodily Awareness

Unfortunately, no measure of facial awareness has been developed[59]. As a proxy, we included the Multidimensional Assessment of Interoceptive Awareness (MAIA). The MAIA does not contain questions about facial activity, but it does distinguish between multiple aspects of body awareness, four which may be relevant for facial feedback effects: (1) noticing body sensations, (2) attending to body sensations, (3) being aware of the association between body sensations and emotional states, and (4) trusting body sensations[90]. Evidence that any of these aspects moderate facial feedback effects would provide some evidence that facial feedback effects rely in part upon conscious processes. We caution, however, that the converse is not true. Given both (a) the usual challenges in accepting the null hypothesis and (b) the fact that the

MAIA does not assess awareness of facial activity, the absence of evidence for moderation should not constitute evidence that facial feedback effects rely entirely upon unconscious processes.

To examine whether individual differences in body awareness moderated facial feedback effects, we post-hoc modeled happiness ratings using a linear mixed-effect model with (a) type of pose (happy or neutral) entered as a categorical predictor, (b) the four subscales of the MAIA entered as continuous predictors, and (c) interactions between type of pose and each of the four subscales. We did not find evidence that any of these four subscales moderated facial feedback effects, noticing $F(1, 200) = 0.00$, $p = .97$; attending $F(1, 200) = 0.49$, $p = .48$; aware $F(1, 200) = 0.29$, $p = .59$; trust $F(1, 200) = 0.00$, $p = .99$.

Pilot Study 3

Based in part on results from pilot studies 1 and 2, a Registered Report to run an in-person study containing the conditions outlined in Supplementary Materials Figure 1 was accepted in-principle. However, due to the emergence of the Coronavirus Disease 2019 pandemic, many participating research groups had to suspend in-person data collection. Consequently, a third pre-registered pilot study was designed to examine the feasibility of conducting the study online. Results from this pilot study led to a minor amendment of the registered report.

Pilot study 3 featured the full 2 (Pose: happy or neutral) x 3 (Facial Movement Task: facial mimicry, voluntary facial action technique, or pen-in-mouth task) x 2 (Stimuli Presence: present or absent) design. The studies differed from the procedures described in the main manuscript in the following ways: participants were not asked to (1) confirm they were completing the study on a computer (although details regarding participants' operating systems were automatically recorded), (2) confirm that they were in a setting with minimal distractions, (3) rate the degree to which they felt like they were expressing happiness during the happy pose, or (4) rate the degree to which their expression matched an image of an individual completing the happy pose task correctly. These features were added due to unexpected findings in pilot study 3.

We recruited 161 participants through the online data collection platform, Prolific, and paid participants $3 to complete the 15-minute study. Based on pre-registered criteria, we removed participants who (1) opened, but did not complete the survey (n = 7), (2) failed to pass both attention checks (n = 18), (c) failed to accurately answer the filler math problems (n = 5), (d) did not follow instructions to complete the survey on a desktop computer (n = 9), (e) self-reported that they did not accurate complete the happy pose task (n = 1), and/or (f) indicated an unwillingness to place a pen in their mouth (n = 2). After exclusions, there were 119 participants (45% female, 52% male, 2% transgender male, 1% gender variant/non-conforming; age $M = 31.30$, $SD = 11.30$). Forty-three participants completed the facial mimicry task, 42 completed the voluntary facial action technique task, and 34 completed the pen-in-mouth task. All participants provided informed consent.

Test of the Facial Feedback Hypothesis

To examine whether facial feedback impacted self-reported happiness, a 2 (Pose: happy or neutral) x 3 (Facial Movement Task: facial mimicry, voluntary facial action technique, or pen-in-mouth task) x 2 (Stimuli Presence: present or absent) mixed-effect ANOVA was fitted, with Pose included as a within-participant factor. Consistent with the facial feedback hypothesis, participants reported more happiness after posing happy ($M = 2.51$, $SD = 1.52$) vs. neutral expressions ($M = 2.16$, $SD = 1.48$), $F(1, 113) = 10.04$, $p = .002$.

No main effect for Facial Movement Task was detected, $F(2, 113) = 0.43$, $p = .65$. There was a main effect for Stimuli Presence, wherein participants reported more happiness when positive stimuli were present ($M = 2.65$, $SD = 1.55$) vs. absent ($M = 2.02$, $SD = 1.41$), $F(1, 113) = 6.26$, $p = .01$. Unexpectedly, no significant interaction between Pose and Stimuli Presence was detected, $F(1, 113) = 0.22$, $p = .64$. These results suggest that posing happy expressions can both initiate and modulate feelings of happiness. However, contrary to our prediction, the modulation effects were not larger than the initiation effects. Also unexpectedly, results revealed an interaction between Pose and Facial Movement Task, $F(2, 113) = 4.15$, $p = .02$. Follow-up contrasts indicated that participants reported more happiness after posing happy vs. neutral expressions in the facial mimicry, $F(1, 113) = 15.54$, $p < .001$, voluntary facial action technique, $F(1, 113) = 4.69$, $p = .03$, but not pen-in-mouth task, $F(1, 113) = 0.12$, $p = .73$.

Although results are somewhat consistent with the facial feedback hypothesis, one alternative explanation is that participants found the smiling task to be less boring than the neutral task. Similar to pilot studies 1 and 2, we compared happiness ratings during the happy vs. filler movement trials using a 2 (Movement Trial: smile or filler) x 3 (Facial Movement Task: facial mimicry, voluntary facial action technique, or pen-in-mouth task) x 2 (Stimuli Presence: absent or present) mixed-effect ANOVA. Consistent with the facial feedback hypothesis, results indicated that participants reported greater happiness after posing happy expressions ($M = 2.51$, $SD = 1.52$) vs. engaging in filler movements ($M = 2.03$, $SD = 1.31$), $F(1, 232) = 29.49$, $p < .001$. Results also revealed an interaction between Movement Trial and Facial Movement Task, $F(2, 113) = 10.69$, $p < .012$. Follow-up contrasts indicated that participants reported more happiness after posing happy expressions vs. engaging in filler movements in the facial mimicry, $F(1, 232) = 37.15$, $p < .0001$, voluntary facial action technique, $F(1, 232) = 19.14$, $p < .0001$, but *not* pen-in-mouth task, $F(1, 232) = 0.36$, $p = .55$. Results also revealed an interaction between Movement

Trial and Stimuli Presence, $F(1, 232) = 9.33$, $p = .003$. Follow-up contrasts indicated that participants reported more happiness after posing happy expressions vs. engaging in filler movements when positive images were present, $F(1, 232) = 35.71$, $p < .0001$. However, when positive images were absent, this difference was not statistically significant, $F(1, 232) = 2.85$, $p = .09$.

Taken together, these results provide evidence that certain facial movement tasks—i.e., facial mimicry and the voluntary facial action technique—can modulate feelings of happiness. These results also provide preliminary evidence that these facial movement tasks can initiate feelings of happiness.

Participant Awareness

Towards the end of the experiment, participants provided written responses to a text-based funnel debriefing. Two independent raters reviewed these written responses and coded whether the participant was aware of the purpose of the study (0 = unaware, 1 = aware; 90.30% agreement; Cohen's $\kappa = 0.65$). Participants were classified as aware if either of the coders indicated that the participant seemed aware of the purpose of the study. In total, 76% of participants were coded as unaware of the purpose of the study.

To examine whether participant awareness differed across conditions, a logistic regression was fitted with facial feedback task entered as an effects-coded factor. Results indicated that awareness differed by condition, $\chi^2(2, N = 119) = 11.28$, $p = .004$. Follow-up pairwise comparisons indicated that participants were more likely to be aware of the purpose of the study in the pen-in-mouth (6% aware) vs. facial mimicry (28% aware), $z = 2.27$, $p = .02$, and voluntary facial action technique conditions (36% aware), $z = 2.74$, $p = .006$. Participant awareness did not significantly differ between the facial mimicry and voluntary facial action technique conditions, $z = 0.77$, $p = .44$.

Although pilot study 3 provides some evidence in favor of the facial feedback hypothesis, it is possible that these effects are driven by participants' awareness of the purpose of the experiment (e.g., demand characteristics). To examine if participant awareness moderated facial feedback effects, we fit a linear mixed-effect model with (a) Pose, Facial Movement Task, Image Presence, and Awareness entered as categorical predictors, (b) two-way Pose by Facial Movement Task, Pose by Image Presence, Pose by Awareness, and Facial Movement Task by Image Presence interactions, and (c) a three-way Pose by Facial Movement Task by Image

Presence interaction. Similar to our earlier tests of the facial feedback hypothesis, results confirmed that participants reported more happiness after posing happy vs. neutral expressions, $F(1, 112) = 14.41$, $p < .001$, and when stimuli were present vs. absent, $F(1, 112) = 6.10$, $p = .01$. Results also provided evidence of a Pose by Facial Movement Task interaction, wherein the effect of Pose was significant in the mimicry, $F(1, 112) = 19.65$, $p < .0001$, voluntary facial action technique, $F(1, 112) = 6.56$, $p = .01$, but *not* pen-in-mouth condition, $F(1, 112) = 0.52$, $p = .47$). Results indicated that the effect of Pose was larger when participants were aware of the purpose of the study, $F(1, 112) = 4.10$, $p = .045$. Subsequent contrasts indicated that participants who were aware of the purpose of the study reported more happiness after posing happy vs. neutral expressions, $F(1, 112) = 10.81$, $p = .001$. However, this effect was not statistically significant among participants who were unaware of the purpose of the study, $F(1, 112) = 3.18$, $p = .08$.

Quality of Posed Expressions

Overall, participants reported following the instructions closely (overall $M = 6.69$, $SD = 0.69$; facial mimicry $M = 6.65$, $SD = 0.78$; voluntary facial action technique $M = 6.60$, $SD = 0.67$; pen-in-mouth $M = 6.85$, $SD = 0.56$). To examine if task adherence varied across conditions, adherence ratings were modeled using a 3 (Facial Movement Task: facial mimicry, voluntary facial action technique, or pen-in-mouth task) x 2 (Stimuli Presence: present or absent) ANOVA. Results did not indicate that adherence ratings varied as a function of Facial Movement Task, $F(2, 113) = 1.43$, $p = .24$, or Stimuli Presence, $F(1, 113) = 1.60$, $p = .21$ Taken together, these results suggest that participants can successfully execute all facial feedback tasks well in an online format.

Summary

The primary purpose of pilot study 3 was to examine the feasibility of online data collection. Although participants had no financial incentive to exaggerate their performance, participants reported completing the facial feedback tasks accurately. Ideally the validity of these reports would be confirmed via remote video recording. However, overt video recording would raise concerns about whether video recording interferes with facial feedback effects[49,50]. In pilot studies 1 and 2, we addressed these concerns by covertly recording participants, but it is not possible to do so in an online format without introducing privacy concerns. Nevertheless, facial

feedback effects were observed in most—but not all—conditions, suggesting that many participants completed the facial feedback tasks sufficiently well.

Although pilot study 3 was primarily designed to examine the feasibility of online data collection, the results provide preliminary but inconclusive evidence in favor of the facial feedback hypothesis. Results indicated that the facial mimicry and voluntary facial action technique—but not pen-in-mouth—tasks produced significant facial feedback effects. Results indicated that facial feedback modulated feelings of happiness, but provided only preliminary evidence that facial feedback can initiate feelings of happiness. Results provided mixed evidence regarding whether facial feedback effects can persist when participants are unaware of the purpose of the study. A marginally significant facial feedback effect was detected in participants judged to be unaware of the purpose of the experiment, but participants who completed the task that best masked the purpose of the experiment (the pen-in-mouth task) did not exhibit significant facial feedback effects.

Reflections on the pen-in-mouth task

Results indicated that the facial mimicry, voluntary facial action technique—but not pen-in-mouth—tasks produced significant facial feedback effects. One interpretation of these results is that the pen-in-mouth task does not reliably produce facial feedback effects. If true, it may suggest that facial feedback effects are driven by demand characteristics[42,43,92]. Results from the three pilot studies, however, are generally inconsistent with this explanation. Participants who were unaware of the purpose of the study exhibited significant facial feedback effects in pilot studies 1 and 2. In pilot study 3, participants who were unaware of the purpose of the study did not exhibit significant facial feedback effects, but the effect was marginally significant despite a smaller sample size. Another interpretation of the results is that facial feedback effects cannot occur outside of awareness—the mechanism the pen-in-mouth task is designed to eliminate[41]. However, inconsistent with this explanation, results from pilot studies 1 and 2 did not indicate that individual differences in bodily awareness moderated facial feedback effects. Nevertheless, to assess this further, the main study will include the Body Awareness Questionnaire[68]. Another possibility is that the pen-in-mouth effect does not produce an expression that feels like a smile. To examine this possibility, participants at the end of the main study will be asked to rate the degree to which they feel like they are expressing happiness during the happy pose.

It is possible that the pen-in-mouth effect is valid, but that participants did not accurately complete the task in pilot study 3. Although participants in the pen-in-mouth condition reported the highest adherence to instructions, they may not have understood how to correctly complete the task. To address this concern, participants at the end of the main study will rate the degree to which their expression matches an image of an individual completing the happy pose task correctly. It is also possible that the pen-in-mouth effect is not reliable in the midst and/or aftermath of Coronavirus Disease 2019. Although speculative, placing a pen in one's mouth may elicit anxiety that interferes with the facial feedback effect. Consistent with this possibility, participants reported slightly increased levels of nervousness in the pen-in-mouth vs. facial mimicry, $t(113) = 2.21$, $p = .07$, and voluntary facial action technique conditions, $t(113) = 2.39$, $p = .048$. In the main study, we will more comprehensively measure anxiety using the Discrete Emotions Questionnaire[66].

Main Study

Did increases an anxiety obfuscate a true pen-in-mouth effect?

We ran two exploratory analyses to examine whether the heighted levels of anxiety in the pen-in-mouth task obfuscated a true facial feedback effect. First, we examined whether feelings of happiness were generally attenuated by anxiety. We re-ran our primary analysis model with (a) random slopes removed (due to convergence issues) and (b) anxiety added as a continuous predictor. Results did not indicate that anxiety decreased happiness, $\beta = 0.01$, $t(2958.82) = 0.38$, $p = 0.71$.

Second, we examined whether facial movement tasks can impact happiness in the absence of anxiety. We re-ran our primary analysis model with (a) random slopes removed, and (b) excluding participants who reported that they experienced even mild levels of anxiety ($n = 1101$). We found evidence of Pose effects in the facial mimicry, $M_{\text{diff}} = 0.53$ or 8.83% scale range, $F(1, 871.73) = 66.13$, $p < .001$, and voluntary facial action tasks, $M_{\text{diff}} = 0.53$ or .53% scale range, $F(1, 866.97) = 57.91$, $p < .001$. However, the evidence was once again less clear in the pen-in-mouth task, $M_{\text{diff}} = 0.10$ or 1.67% scale range, $F(1, 887.45) = 2.47$, $p = .12$. Taken together, these exploratory analyses do not provide compelling evidence that increases in anxiety in the pen-in-mouth condition obfuscated a facial feedback effect.

Cross-country comparisons of facial feedback effects

As an exploratory analysis, we examined cross-country differences in the effect of each facial feedback task. As discussed in the main text, we initially attempted to do so by modeling random slopes in the mixed-effects models used in our primary analyses. However, these models often yielded convergence issues. Consequently, we performed a random-effects meta-analysis on the primary analysis data with Cohen's standardized difference (d_{rm}) as the effect size index[93,94]. Before describing the results, three notes of caution are warranted. First, our experiment was not powered to provide high-power tests of the effects of facial feedback in each country sampled. Thus, even if facial feedback effects are real, many country-specific effect size estimates could still be statistically non-significant. Second, although between-country variability in estimated facial feedback effects could have theoretically-interesting causes (e.g., true cross-cultural variability), they could also have theoretically-uninteresting causes (e.g., difference in the age of participants sampled across countries; differences in how well

participants completed the task across countries). Third, we did not engage in representative sampling, so it is unclear if findings generalize to the populations being examined.

Random-effects meta-analysis estimated that the overall effect of posed happy expressions medium on happiness was $d = 0.16$, 95% CI [0.17, 0.32], $p < .001$. This overall effect was significantly moderated by the facial movement task ($z = 4.52$, $p < .001$). As shown in Supplementary Figure 2, posed happy expressions had similar-sized effects on happiness in the facial mimicry ($d = 0.24$, 95% CI [0.17, 0.32], $p < .001$) and voluntary facial action conditions ($d = 0.26$, 95% CI [0.14, 0.38], $p < .001$)—but not the pen-in-mouth condition ($d = 0.04$, 95% CI [-0.01, 0.08], $p = .10$).

Estimated between-country variability in facial feedback effects was medium in size, but not always significant (facial mimicry $Q(18) = 25.59$, $p = .11$, $\tau = 0.09$; voluntary facial action $Q(18) = 30.51$, $p = .03$, $\tau = 0.15$; pen-in-mouth $Q(18) = 25.51$, $p = .11$, $\tau = 0.04$). As shown in Supplementary Figure 2, the estimated effects of facial feedback in the facial mimicry and voluntary facial action tasks were similar in size across countries (and positive in almost all countries sampled). However, these effect size estimates were not statistically significant in all countries (e.g., Nigeria and Italy), possibly due to relatively small country-specific sample sizes.

Supplementary Fig. 2: Forest plot of the effect of posed facial expressions and exposure to positive images on happiness.

Forest plot of the effects of (a) posed happy vs. neutral expressions (white panels) and (b) exposure to positive images (tan panel) on happiness, using Cohen's d_rm as the effect size index. Grey boxes represent country-level effect size estimates, and grey boxes represents their 95% confidence interval. Blue diamonds/bars represent overall effect size estimates abd 95% confidence intervals. Countries are denoted by International Organization for Standardization codes. Three metrics of effect size heterogeneity are reported: Cochrane's Q (Q) and its accompanying p-value, I^2, and τ.
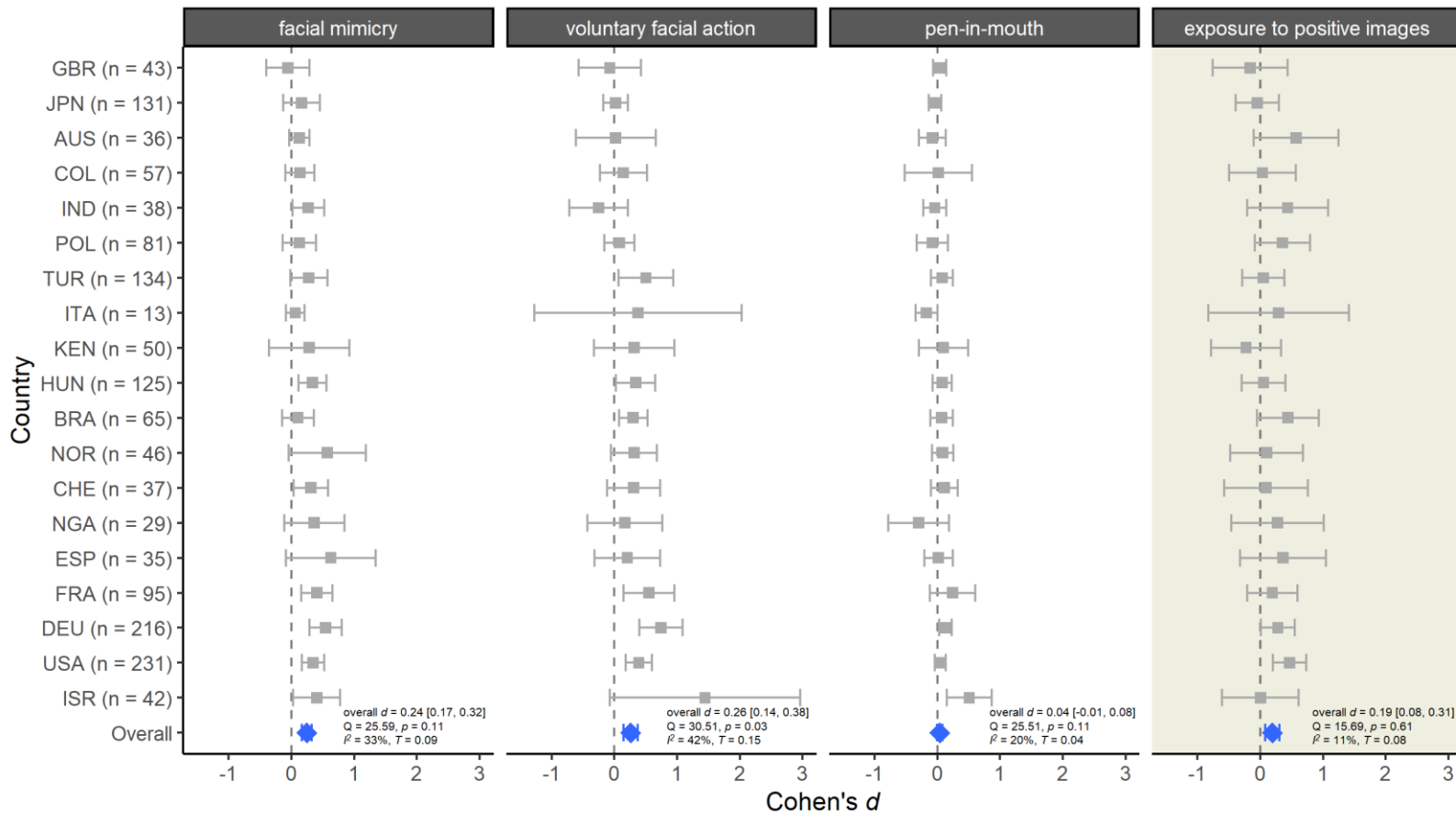
Distribution of difference scores

A simple approach to visualizing the effects observed in the main study is to plot the frequency distribution of the happiness difference scores. To do so, we calculated a happiness difference score for each participant by subtracting their neutral pose happiness ratings from their happy pose happiness ratings. Positive difference scores indicate that the participant reported higher levels of happiness when posing happy expressions. As shown in Supplemental Figure 3, 53% of participants in the facial mimicry and voluntary facial action conditions reported higher levels of happiness when posing a happy, as opposed to neutral, expression. 32% did so in the pen-in-mouth condition.

Supplementary Fig. 3: Frequency distribution of happiness difference scores in the facial mimicry, voluntary facial action, and pen-in-mouth condition.

Positive values indicate that the participant reported higher levels of happiness when posing happy vs. neutral expressions. The proportions of participants with positive or negative happiness difference scores are displayed in each panel.
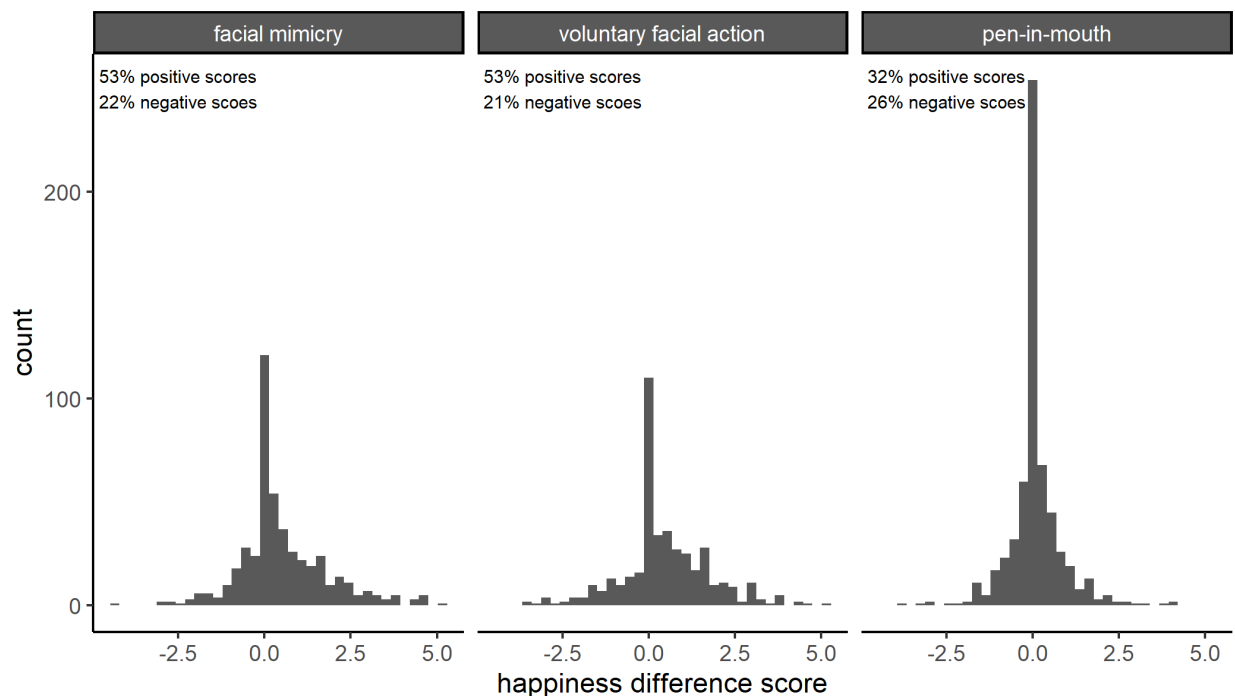
Additional information on the Open Science Framework

We refer the interested reader to the project page on the Open Science Framework: https://osf.io/ac3t2/. On this page, readers can find: (1) all study materials, including translated surveys, images, task videos, and coding manuals, (2) all study data, except the video recordings from Pilot Studies 1 and 2 (we did not have permission to share these recordings, but do share the processed data), (3) a supplemental table showing country-specific demographic and inclusion criteria information, and (4) information about how the collaboration was structured and coordinated.