# In-group Social Conformity Updates the Neural Representation of Facial Attractiveness

Danni Chen[1], Ziqing Yao[1], Jing Liu[2],

Haiyan Wu[3] and Xiaoqing Hu[1,4*]

1, Department of Psychology, The State Key Laboratory of Brain and Cognitive Sciences,

The University of Hong Kong, Hong Kong SAR, China

2, Department of Applied Social Sciences,

The Hong Kong Polytechnic University, Hong Kong SAR, China

3, Centre for Cognitive and Brain Sciences and Department of Psychology,

University of Macau, Macau SAR, China

4, HKU-Shenzhen Institute of Research and Innovation, Shenzhen, China

* Correspondence should be sent to:

Xiaoqing Hu

Department of Psychology, The University of Hong Kong,

Pokfulam, Hong Kong SAR, China

Email: xiaoqinghu@hku.hk

# In-group Social Conformity Updates the Neural Representation of Facial Attractiveness

## Abstract

People readily change their behavior to comply with the public. However, to which extent they will internalize the social influence remains elusive. In this pre-registered electroencephalogram (EEG) study, we employed a facial attractiveness social learning paradigm to investigate how learning from one's in-group or out-group members would change attractiveness perception and neural representation. We found that participants changed their explicit attractiveness ratings to both in-group and out-group influences, i.e., public compliance. We next quantified the neural representational similarities of learned faces with prototypical attractive faces during a face perception task without overt social influence and intentional evaluation. We found that the neural representation of facial attractiveness changed only when participants learned from their in-group members, and among those who perceived tighter social norms. These findings provided novel knowledge on how group affiliations and individual differences modulate the impact of social influence on the internalization of social influence.

*Keywords:*

Social conformity, social influence, private acceptance, perceived tightness-looseness, multivariate representational similarity analysis

## Statement of Relevance

Evolutionary-wise, following the crowd is important given its survival benefits. However, blind compliance with group opinions can also result in immoral decisions when group influence is toxic. Therefore, the crucial question is: To what extent, in which context, and for which population do individuals internalize social influence? We found that while people complied with opinions from both in- and out-group members, they only internalized such social influence from their in-group members, as evidenced by updated neural representations even when social influence was no longer present, and when the intentional evaluation was not required. This neural internalization effect was particularly pronounced among those who perceived tighter social norms. Our study suggests that in-group social influence is internalized and explains why in-group social influence is hard to eliminate afterward.

**Introduction**

When observing behaviors or opinions shared by the majority, people often align their behaviors and thoughts to be consistent with others, even though initially they hold opposite views. This phenomenon is termed “social conformity” (Izuma, 2013) and is ubiquitous: from mundane choices (e.g., which movie to watch) to decisions that bear significant personal and societal consequences (e.g., whether to get vaccinated or which candidate to favor). Evolutionary-wise, conformity helps people adapt to and learn about uncertain environments via following the crowds so as to ensure survival and reproduction (Claidière & Whiten, 2012; Morgan et al., 2022). Indeed, social conformity has been documented across different species, ranging from rodents to primates (Hopper et al., 2011; van de Waal et al., 2013), and emerges early along the developmental trajectory (Haun & Tomasello, 2011; Izuma & Adolphs, 2013).

Despite the prevalence of social conformity, to what extent people would internalize social influence remains contentious. Social conformity leads to two outcomes: public compliance and private acceptance. While public compliance refers to changes in one’s external behavior to align with public opinions, private acceptance induces stable changes in one’s attitudes and beliefs (Izuma, 2013). Understanding how social influence induces private acceptance is important because attitudes and beliefs exert a powerful influence on behaviors in various settings, including consumer choices, interpersonal/intergroup relationships (Kurdi et al., 2019), and political voting (Greenwald et al., 2009), among others. On the other hand, self-reported public opinions can be distorted by either lacking introspection of one’s mental processes or impression management strategies (Sassenrath, 2020).

Despite the importance of private acceptance in social influence, a significant yet unresolved challenge is how to distinguish between public compliance and private acceptance, particularly when relying on explicit behavioral changes. Advances are made when research leverages neuroscientific methods to investigate the neural correlates of beliefs and evaluations. One often-adopted approach is that if changes in explicit behavior are accompanied by changes in neural activity implicated in evaluation or value computation, then private acceptance of group influence is inferred (Edelson et al., 2011; Zaki et al., 2011). However, when people make deliberate evaluations, public compliance may nevertheless influence the neural activity implicated in evaluations, which casts doubts on the inference of private acceptance (Berns et al., 2010). Thus, a better understanding of private acceptance and the implicated neural activity changes would entail both behavioral paradigms and analytical approaches that minimize the impact of public compliance on evaluation.

In addition to examining private acceptance, we further investigated one important factor that may modulate social influence: one's group affiliations. People readily see others via the lens of social categorization, and social influence from either in-group or out-group members could lead to different levels of conformity at both behavioral and neural levels. People were more likely to converge their behavior to the in-group than the out-group members and even diverge from disliked out-group members (Izuma & Adolphs, 2013). Additionally, compared to the out-group, the in-group influence exerted more powerful impacts on the neural activity in the reward-related regions (Lin et al., 2018) during undergoing social influence. Evolutionary-wise, in-group conformity can be more adaptive because it increases in-group homogeneity and facilitates coordination (Claidière & Whiten, 2012). However, how in- vs. out-group influence impacts private acceptance and the implicating neural activity remains unclear.

Here, we aimed to address these questions in a pre-registered EEG experiment on facial attractiveness (for pre-registration, see https://osf.io/5e7kr/?view_only=cf903bc29f8543a19272046a45a8349c). In a classic social learning framework (FeldmanHall et al., 2018), participants received normative feedback on facial attractiveness from either in-group or out-group members, introduced via a minimal group paradigm (Goldenberg et al., 2020). We measured public compliance based on the changes in attractiveness ratings. To quantify private acceptance at a neural level, we designed an EEG-based face perception task in the absence of intentional evaluation or ostensible social influence, minimizing the impact of public compliance. Moreover, we applied multivariate neural representation similarity (RSA) analyses to the face-elicited EEG to extract neural representational changes of facial attractiveness. Notably, the task also included prototypical attractive faces, which allowed us to build an individualized neural representation model of prototypical attractiveness. Via computing the neural representation similarities between the learnt faces and the prototypical attractive faces, we could infer whether the learnt faces were indeed perceived as more or less “attractive” as a result of social influence. Indeed, this RSA approach provides a powerful tool to capture the subtle changes of neural representation of specific features of stimuli, in our case, facial attractiveness, even without significant changes in the univariate neural activities (Popal et al., 2019).

We pre-registered our hypotheses that participants would be more likely to publicly comply with in-group than out-group opinions as evidenced by explicit attractiveness rating change (Hypothesis 1). For private acceptance evidenced by the neural representation updates and ERP changes, we aimed to test two competing hypotheses: Participants would privately endorse only in-group members' influence (Hypothesis 2a) or both in- and out-group influence (Hypothesis 2b).

## Methods

### Participants

We pre-registered to recruit 42 participants, which is larger than previous similar EEG experiments on attractiveness (Werheid et al., 2007) and social conformity (Shestakova et al., 2013), and allows us to detect effect sizes in the range of 0.40 to 0.50[1]. Anticipating potential attrition and data exclusion, we recruited 48 participants (37 females; age, *mean* = 23.98, *S.D.* = 3.13) from a local university. Participants received monetary compensation at a rate of 80 HKD/hour. Three participants were excluded from subsequent EEG analysis due to excessive EEG artifacts, resulting in 45 participants who met our pre-registered inclusion criteria: 1) Following artifact rejection, each participant's clean EEG segments should be more than 50% of total trials in the face perception task in both pre- and post-learning phases; and 2) participants should correctly report their assigned group identity. All participants were native Chinese speakers, right-handed, not color blind, had a normal or corrected-to-normal vision, and did not report any history of neurological or psychological disorders. All participants provided written

[1] We conducted the sensitivity analysis with G*Power (Faul et al., 2007). As we focused on the main effect of feedback, we conducted a sensitivity analysis with number of groups = 1, measurements = 3, and sample size = 42. The lowest power was set as 0.80, while the highest power was set as 0.95.

informed consent prior to the participation and were debriefed and compensated after completing the study. This research procedure was approved by the Human Research Ethics Committee of the University of Hong Kong (HREC No. EA1912003).

**Materials**

For experimental faces, we chose 70 photographs of medium-attractive East Asian female faces with their hair and ear removed by PhotoShop.  For prototypical attractive faces, we created 10 faces by morphing four randomly selected faces from the same face database by FunMorph. All photos were round-cropped and manually aligned with size, luminance, lightness, and color using Adobe Lightroom. Data from an independent sample of 18 participants confirmed that prototypical faces were more attractive than experimental faces ($p < .001$, see Data S1).

**Procedure**

All tasks were programmed and presented by PsychoPy (version 2020.1.3, Peirce, 2007). Participants visited the lab twice, separately by seven days. In the first lab visit, participants completed the Positive and Negative Affect Schedule (PANA-SF; Watson et al., 1988), Interpersonal Reactivity Index (IRI; Davis, 1983), Tightness-Looseness Questionnaire (TLQ; Gelfand et al., 2011), Socially Desirable Responding (SDR; Paulhus, 1984), and Social Phobia Inventory (SPIN; Connor et al., 2000), followed by computer-based tasks.

For computer-based tasks, participants completed three phases: 1) pre-learning, 2) learning, and 3) post-learning (Figure 1). Participants performed a face perception task and an explicit rating task in both the pre-learning and post-learning phases, with 70 medium-attractive experimental faces intermixed with 10 prototypical attractive faces. In the face perception task, participants viewed 480 faces intermixed with 144 objects (for trial structure, see Figure 1), divided into 6 blocks. To maintain attention, participants pressed a button on a keyboard when an object was presented on the monitor (target hit rates > 0.99). In the explicit rating task, participants rated each of the 80 faces with a mouse on attractiveness, confidence, perceived competence, and perceived warmness (1 to 11).

The learning phase included a minimal group formation task, an associative learning task, and a social learning task. In the minimal group formation task, participants were randomly assigned to one of two groups (Green or White) and were told that the group assignment was based on the similarity of their personal preferences with the others (Goldenberg et al., 2020). Participants then completed an associative learning task, in which they made a speedy button press when their name was paired with their assigned group labels, among other names and the other group label pairs. This task served to strengthen the group identity and sense of belongingness. Participants indeed showed higher in-group identification and favoritism (*p*s < .002, Data S2). During the social learning task, participants were presented with the face again, together with the attractiveness rating feedback from either in-group or out-group members (i.e., Affiliation), which was either HIGHER, LOWER, or CONSISTENT (i.e., Feedback) than/with their initial ratings, resulting in a 2 (Affiliation) by 3 (Feedback) within-subject design. Assignments of experimental faces to each of the six conditions were counterbalanced across participants, with 10 faces as no-learning control faces. The post-learning phase was the same as the pre-learning phase, except that participants completed a cued recall task on their memories of the feedback before the perception and the rating tasks. Participants then provided their demographic information and answered group identification questions. Details of the minimal group manipulation, cued recall task, and social learning task are provided in Supplemental Methods.

Seven days later, participants visited the lab for the delayed phase. EEGs were not recorded in this phase. Analyses of the learning task, and the cued recall tasks are beyond the scope of current experiment and not reported here.

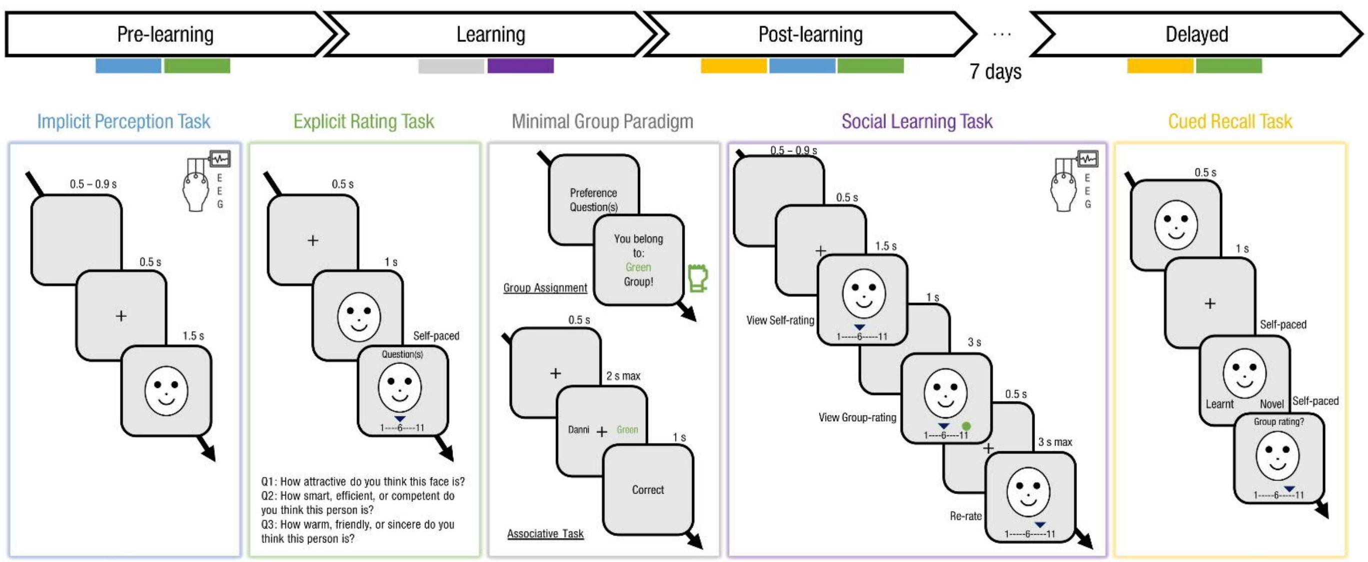


**Figure 1.** Experimental Procedure. The upper row represents the procedural flow, with colored rectangles below to illustrate each task in sequence. In both the pre-learning and post-learning phase, participants completed the same face perception task and the explicit rating task. Participants learned about their affiliation and either in-group or out-group members' attractiveness evaluation of each face in the learning phase. In the delayed phase, participants completed the same cued recall task and the same explicit rating task as in the post-learning phase. EEGs were recorded during the Face Perception Task and Social Learning Task in the first lab visit. The face icon, with hair and ears removed, illustrated the Asian female faces we used in our experiment. Exemplar trials from corresponding tasks were shown in the colored diagram of the lower row.

**EEG Acquisition and Preprocessing**

Continuous EEGs were recorded with an eego amplifier and a 64-channel gel-based waveguard cap based on an extended 10–20 layout (ANT Neuro, Enschede, and Netherlands). The online sampling rate was 500 Hz. The online reference electrode was CPz, and the ground electrode was AFz. The horizontal electrooculogram (EOG) was recorded from an electrode placed 1.5 cm to the left external canthus. The impedance of all electrodes was maintained below 20 kΩ during the recording.

Raw EEG data were processed offline using custom scripts, the EEGLAB toolbox (Delorme & Makeig, 2004), and the ERPLAB toolbox (Lopez-Calderon & Luck, 2014) implemented in MATLAB (MathWorks Inc, Natick, MA, USA). Raw EEG signals were first downsampled to 250 Hz and bandpass-filtered in the frequency range of 0.05–30 Hz using the FIR filter implemented in EEGLab. We removed 50 Hz line noise by applying the CleanLine algorithm (Mullen, 2012). EOG, M1, and M2 electrodes were removed from the EEG data before further processing. Bad channels were visually detected, removed, and then interpolated. To facilitate the independent component analysis (ICA), the EEG data were segmented into [-1000 to 2000 ms] epochs relative to the onset of the face and were then high-pass filtered with a cutoff frequency of 1 Hz (Winkler et al., 2015) before being subjected to ICA. After the ICA, artifacts caused by eye movements and muscle activity were identified and corrected using visual inspection and the ICLabel plugin (Pion-Tonachini et al., 2019) implemented in EEGLAB. In addition, artifacts were automatically identified using the threshold of +/- 100 μV. Note that we pre-registered a threshold of +/- 75 μV to exclude EEG artifacts. However, adopting this stricter threshold resulted in more excluded trials, thus reducing statistical power (Table S1). Trials with artifacts or with incorrect responses (i.e., false alarms) were excluded from further analysis. On average, 475.50 (*S.D.* = 34.13) and 452.99 (*S.D.* = 44.01) trials were included for pre- and post-learning face perception EEG analyses, respectively.

**Event-Related Potential (ERP) Analysis**

For ERP quantifications, continuous EEGs were segmented into [-200 to 1000 ms] epochs and were averaged for ERPs using the -200-0 pre-stimulus as baselines. We pre-registered our intention to analyze the face processing component N170 (120 - 220 ms, Lu et al., 2014) at pre-defined occipitotemporal sites, and the evaluation-related component LPC (300 - 800 ms, Werheid et al., 2007) at the pre-defined central-parietal and frontal-central sites. We conducted statistical analyses on the update of ERP by subtracting the mean amplitude of N170/LPC during the pre-learning phase from the same ERPs during the post-learning phase.

**Multivariate Representational Similarity Analysis (RSA)**

We calculated the neural representation similarity between the experimental face and prototype attractive faces, i.e., the experimental-prototype face similarity (EPS) as a neural index of attractiveness. EEG data were downsampled to 100 Hz to facilitate multivariate similarity analyses. To reduce the effect of univariate activity on the multivariate neural pattern similarity, all individual trials were normalized by subtracting the mean and then dividing by the standard deviation of ERP activities at each time point within each participant (Fellner et al., 2020). Next, the -200 to 1000 ms EEG epochs were continuously segmented into overlapping windows of 200 ms with 10 ms increments. By Spearman Correlation, we calculated the neural pattern similarity between individual time windows of every two trials (experimental face and prototype face) across all 61 channels. To control the temporal proximity effect (i.e., higher similarities would be expected for adjacent trials), we only analyzed the trials with more than four trials apart. The similarity of each face was averaged across all the correlation coefficients between all the trials of this face and all prototypical attractive faces. We conducted the cluster-based non-parametric permutation test by shuffling the subject label and constructing a null distribution 5,000 times with the default functions implemented in FieldTrip (Oostenveld et al., 2011).

## Results

### Pre-Registered Confirmatory Behavioral Results

First, we tested whether group affiliation interacted with social influence in impacting public compliance. To correct the regression-to-mean effect and the systematic rating difference across phases, we computed the mean-corrected ratings by subtracting the average rating of each phase across conditions from the ratings (Huang et al., 2014). Public compliance was calculated by subtracting the mean-corrected rating of the pre-learning phases from that of the post-learning for each face. An affiliation (in- vs. out-group) by feedback (higher, lower, consistent) repeated measures ANOVA on public compliance revealed a significant feedback effect ($F$ (1.91, 89.75) = 9.17, $p$ < .001, $\eta^2$ = 0.07, $BF_{10}$ = 720.31, Figure 2A). Participants rated the faces less attractive in the lower condition than in the higher ($t$ (94) = 4.08, $p$ < .001, $d$ = 0.60) and in the consistent condition ($t$ (94) = 3.17, $p$ = .006, $d$ = 0.53), while there was no significant difference between the higher and consistent conditions ($t$ (94) = 0.91s, $p$ = 1.000, $d$ = 0.14). However, neither the main effect of affiliation ($F$ (1, 47) = 0.31, $p$ = .579, $\eta^2$ = 0.001, $BF_{10}$ = 0.15) nor the affiliation by feedback interaction was significant ($F$ (1.93, 90.52) = 0.66, $p$ = .513, $\eta^2$ = 0.005, $BF_{10}$ = 0.20), with Bayesian factors strongly favoring the null hypothesis. The same analysis on the delayed public compliance (delayed minus pre-learning) revealed no significant effect (*p*s > .430, Data S3). Taking the regression to mean effect into consideration, we repeated the analyses using faces that were matched on baseline ratings across feedback conditions (Huang et al., 2014; Zaki et al., 2011) and obtained similar results (Data S4). Together, these results suggested that social influences induced public compliance at least in the immediate test. However, in contrast to Hypothesis 1, the group affiliation did not significantly modulate public compliance.

**Pre-Registered Exploratory Behavioral Results**

Next, we aimed to explore the relationships between individual differences and public compliance (Table S2–S3). Among these individual difference measurements, we found that perceived tightness-looseness moderated the effect of social influence. We conducted a linear regression analysis using tightness-looseness, feedback (higher vs. lower), and their interaction as independent variables; and the rating changes as the dependent variable. We found that the regression coefficients of the tightness-looseness scale ($b = 0.13$, $p = .050$), feedback ($b = 0.73$, $p = .057$), and their interaction ($b = -0.24$, $p = .009$) were significant (Figure 2B). The post-hoc analysis revealed that the standardized regression coefficient for the higher condition was significantly higher than for the lower condition: $t$ (188) = 2.65, $p = .009$. These results indicated that individual differences in perceiving the tightness of social norms influenced people's public compliance.

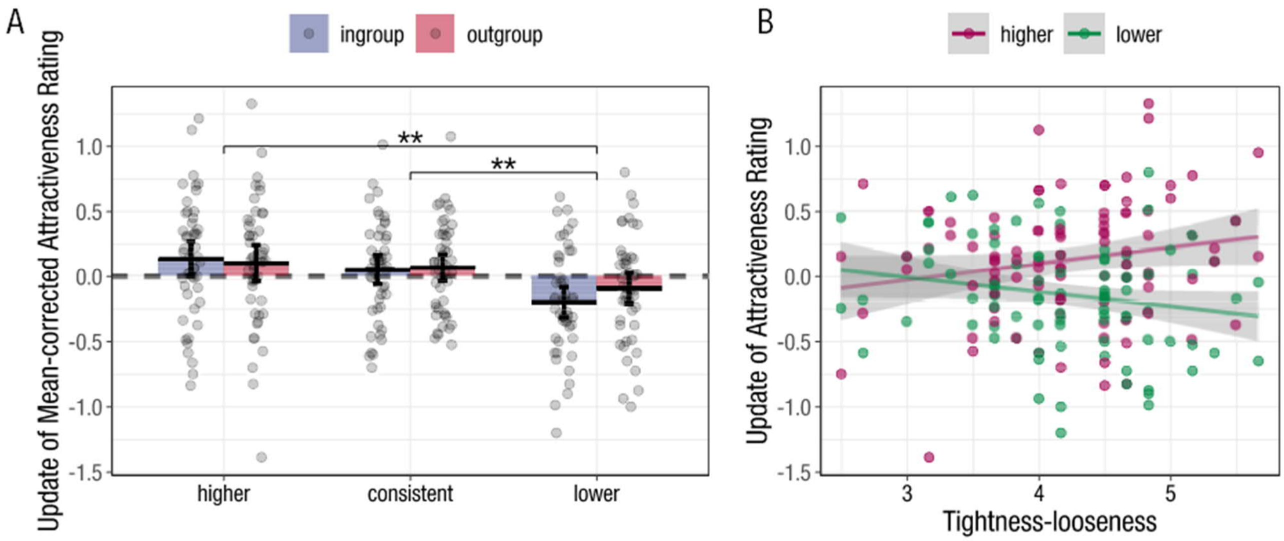


**Figure 2.** Behavioral results from pre-registered analyses. A) Update of mean-corrected attractiveness rating in the immediate test. B) Correlation between the update of mean-corrected attractiveness rating and tightness-looseness.

## Pre-Registered Confirmatory ERP Results

We examined the face- and evaluation-related ERPs (N170, LPC) on each pre-defined ROI by affiliation by feedback repeated measures ANOVAs. No significant main effects of affiliation ($ps > .056$, $\eta^2 < 0.008$), of feedback ($ps > .176$, $\eta^2 < 0.005$), or their interaction ($ps > .088$, $\eta^2 < 0.009$) were found (Figure 3, Table S4). The results remained similar when using the pre-registered threshold (Data S5 and Table S5). Our results suggested that the effect of social influence did not emerge when using univariate ERP analyses.

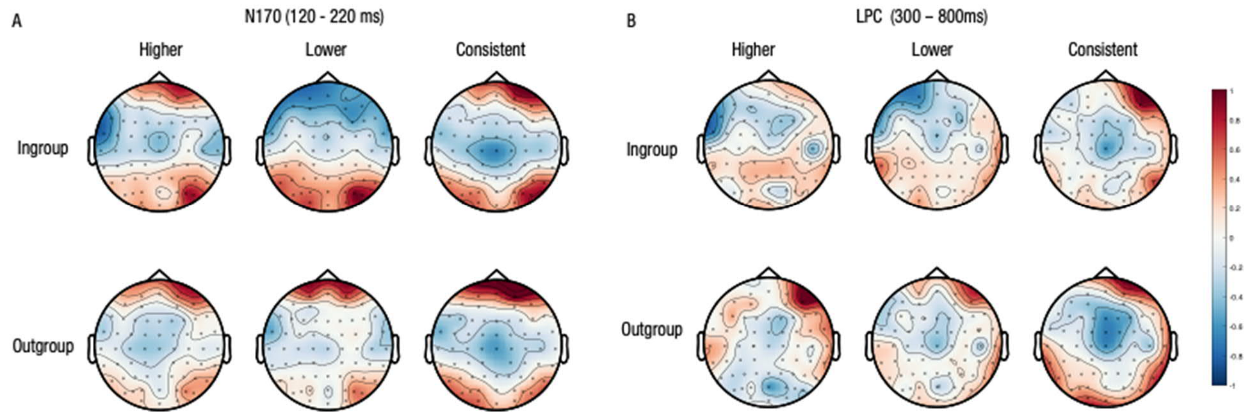


**Figure 3.** ERP Results. The topography of the average amplitude across A) 120 - 220 ms (i.e., N170) and B) 300 - 800 ms (i.e., LPC).

## Pre-Registered Exploratory RSA Results

Given the limitation of univariate analysis in analyzing multidimensional information (Popal et al., 2019), we further examined the neural representations of facial attractiveness using multivariate RSA (Figure 4A). We first showed that in the pre-learning phase, the Experimental-Prototype Similarity (EPS) did not differ between different feedback conditions ($ps > .114$). In contrast, in the post-learning phase, we found a marginally significant cluster in the in-group condition ($p_{corrected} = .069$), while no significant cluster was found in the out-group condition ($ps > .466$). We further conducted a time (pre- vs. post-learning) by feedback (higher vs. lower vs. consistent) ANOVA for in-group and out-group conditions separately to better control the pre-learning baseline difference. We found a significant interaction cluster only in the in-group condition: $p_{corrected} = .048$. Post-hoc analysis (Figure 4B) with the EPS update (i.e., the difference between EPS at the post-learning and pre-learning phases) showed that the higher condition was associated with the higher EPS update than the consistent condition ($p_{corrected} = .016$), and than the lower condition, which was marginal significance ($p_{corrected} = .070$). No significant difference between the CONSISTENT and the LOWER conditions was found ($ps > .260$). Again, no significant clusters were found in the out-group condition ($p_{corrected}$s > .211, Figure S1A). These findings suggested that social influence, especially in-group influence, updated the neural representation of facial attractiveness.

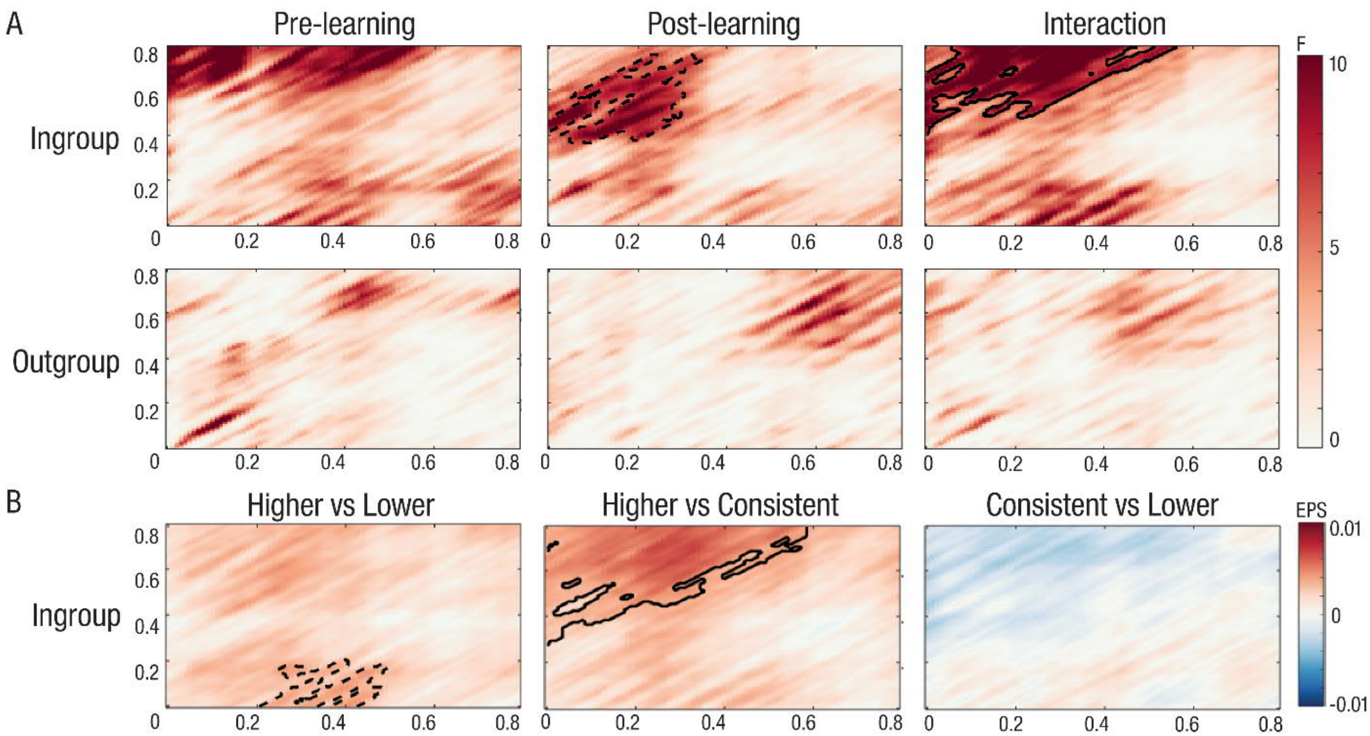


**Figure 4.** RSA Results. A) F-value of the cluster-based permutation analysis testing the difference across higher, lower, and consistent. B) EPS update between different feedbacks in the in-group condition. The x-axis is the timescale of the experimental stimuli, and the Y axis is the timescale of the prototype stimuli. The dashed contour indicates marginally significant clusters ($p < .10$), while the solid contour indicates significant clusters ($p < .05$).

Having shown that the perceived tightness-looseness influenced the attractiveness rating change at the behavioral level, we further examined whether the perceived tightness-looseness would also influence the implicit perception at the neural level. To this end, we divided participants into high ($n = 22$) and low ($n = 23$) tightness-looseness groups (TLQ) based on whether their perceived tightness-looseness was higher than the median tightness-looseness score. As we found that the private acceptance was only in response to the in-group influence, we focused on updating in the in-group condition. The cluster-based permutation test found that for the high-TLQ group, the EPS update (Figure 5) from the pre- to the post-learning phases of the higher condition was significantly higher than that of the LOWER condition ($p_{corrected}$ = .028) and the consistent condition ($p_{corrected}$ = .028). No significant cluster was found between the consistent and the lower condition. Contrarily, for the low-TLQ group, no significant cluster was found when testing the difference among all conditions (*p*s > .296). Again, no significant cluster was found in the out-group condition (Figure S1B). We confirmed that no significant difference between the high- and low-TLQ group in the in-group favoritism (Data S6). These results suggested that the significant main effect of social influence on attractiveness neural representation updates was mostly driven by participants who perceived social norms as tight. Again, this effect was particularly evident in the in-group condition, corroborating an in-group advantage effect.

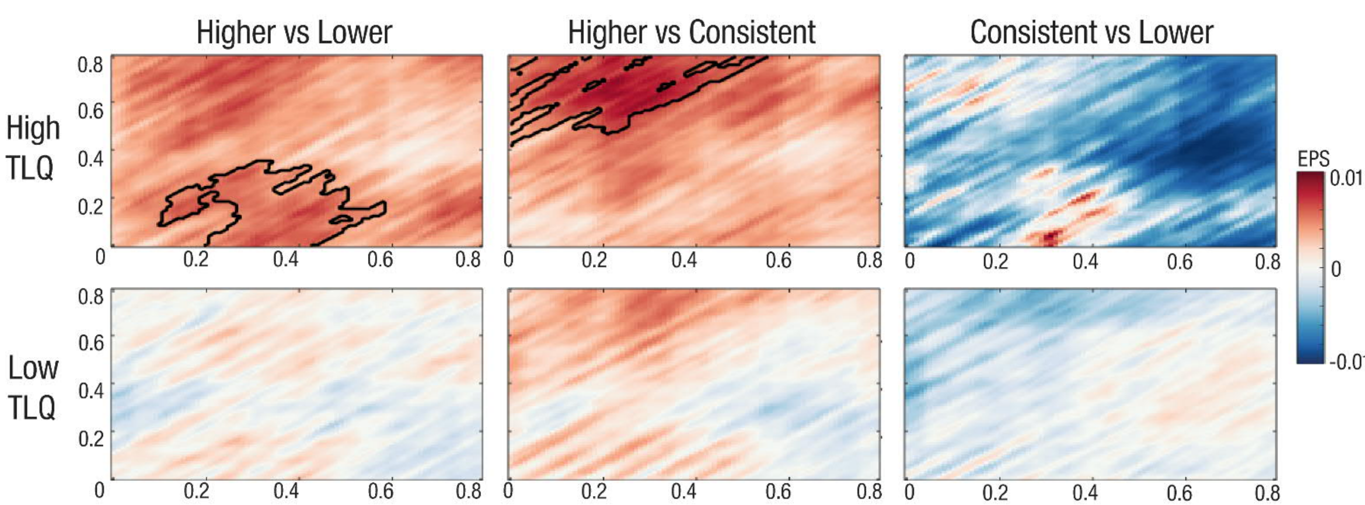


**Figure 5.** EPS update in the in-group condition of the high and low TLQ group. The upper row is the EPS update in the High TLQ group in which we could observe the significant clusters in the HIGHER vs. LOWER and the HIGHER vs. CONSISTENT conditions, but not in the CONSISTENT vs. LOWER condition. The lower row is the EPS update in the Low TLQ group

in which no significant clusters were found. The solid contour indicates the significant clusters ($p < .05$).

## Discussion

Social influences profoundly shape our beliefs and behaviors. Here, we offer new insights into when private acceptance emerges when people learn from either in-group or out-group consensus. To quantify private acceptance in face attractiveness evaluation, we leveraged the computation power of representational similarity analysis (RSA) to examine the neural representation of face attractiveness perception when external social influence is no longer salient. We showed that only in-group influences updated the neural representations of facial attractiveness, and such an update was more pronounced among those who perceived tighter social norms.

### Public Compliance to Both In-Group and Out-Group Influence (Hypothesis 1)

We first replicated previous findings that people have a universal tendency to comply with others publicly (Huang et al., 2014; Zaki et al., 2011). However, contrary to pre-registered Hypothesis 1, we did not observe significant differences in public compliance in response to in- and out-group influence. Using a similar minimal group manipulation, previous research also showed that people did not exclusively conform to in-group members when no oxytocin was given (Stallen et al. 2012). Moreover, Ivanchei et al. (2019) found that participants would comply with feedback regardless of whether it came from humans or computers. Together, these results provide converging evidence that public compliance is not susceptible to the source of influence.

### In-Group Advantage in Private Acceptance (Hypothesis 2a)

Consistent with Hypothesis 2a, we observed an in-group advantage for private acceptance, even if group formation was arbitrary. Indeed, previous research observed increased mentalizing- and reward-related neural activities when viewing social influences from in-group members than from out-group members (Lin et al., 2018). Extending this research, we found that in-group influence modulated neural representations of facial attractiveness even when ostensible social influence is no longer present, i.e., private acceptance. Specifically, applying RSA to the EEG elicited by prototypical attractive faces allowed us to build an individualized neural representation model of prototypical attractiveness, to be compared with the experimental faces evaluated by in- and out-group peers. The analytical power of RSA allows us to detect subtle changes in the neural representation of facial attractiveness before and after exposure to social influence, providing a sensitive indicator of private acceptance at a neural level. Moreover, updated neural representations happened at an early time window, which captured attractive evaluation (Kaiser & Nyga, 2020), and overlapped with face perception and attractiveness perception time window. Broadly speaking, our results provide a mechanistic explanation for continued social influence, such that the effect of social influence tends to linger even after debriefing (Edelson et al., 2011). If social influence, especially when from in-group members, readily changes the neural representation of stimuli, the influence will be evident even when the explicit social influence and norms are no longer salient.

**Norms Sensitivity Affects Both Public Compliance and Private Acceptance**

People differ in how they perceive social norms. One dimension of the perception is the perceived tightness-looseness of social norms, which represents how an individual perceives society as having tight or loose norms and a low or high tolerance for norm-deviant behaviors

(Gelfand et al., 2011). Even within the same cultural context, the perceived tightness-looseness would vary across individuals and impact how they react to social influence (Harrington & Gelfand, 2014). For public compliance, we found that perceived tightness-looseness positively predicted explicit attractiveness rating changes: Tighter perceived norms were associated with increased levels of public compliance. Intriguingly, perceived tightness-looseness norms also modulated private acceptance at a neural level: Tighter perceived norms were associated with increased changes in neural representations of facial attractiveness. Recent research found that perceived tightness-looseness predicted the amplitude of N400, an ERP component sensitive to semantic incongruity when participants viewed various norm-deviant behavior (Goto et al., 2022). Extending prior research, our results suggested that those who perceived society as tight would be more intrinsically motivated to follow social influence, as evidenced by both behavioral changes (public compliance) and neural representation changes towards in-group influence (private acceptance).

**Limitations and Future Directions**

First, we did not record EEG during the delayed test, which restricted us from investigating the longevity of private acceptance. As attitude and behavior are not always aligned, future research should focus on the long-term effect of private acceptance. Second, our research focused on facial attractiveness, an evaluation might be easier to challenge than topics that are central to one's values and worldview, such as moral values and political views. Future research could apply this social learning framework, combined with the neural representation approach, to examine how social influence would change attitudes and beliefs that are core to one's worldviews and values. Finally, while we measured individual differences in perceived tightness-looseness among participants from the same culture, future research is warranted to examine whether public compliance vs. private acceptance of social norms may vary as a function of tight-looseness culture in cross-cultural studies.

**Conclusions**

Following the crowd has survival benefits. Particularly, in-group conformity could lead to in-group homogeneity, with private acceptance resulting in long-lasting behavioral change and internalized norms abided by all group members. Our pre-registered EEG study found that the individuals would comply with others in the laboratory setup but further expand to other contexts when overt normative feedback and intentional evaluation were absent. Our results suggest that individuals would indeed internalize the social influence, especially from the in-group members and those who perceive tight social norms. These findings emphasized that while people widely follow social influence publicly, they may only internalize the social influence in limited circumstances.

## Author Contributions

**Danni Chen:** Conceptualization, Investigation, Formal Analysis, Data Curation, Software, Methodology, Writing - Original Draft, Writing - Review & Editing, Visualization; **Ziqing Yao:** Conceptualization, Validation, Writing - Review & Editing; **Jing Liu:** Methodology, Writing – Review & Editing; **Haiyan Wu:** Conceptualization, Writing – Review & Editing; **Xiaoqing Hu:** Conceptualization, Writing - Original Draft, Writing - Review & Editing, Supervision, Project Administration, Funding Acquisition.

## Acknowledgments

We thank Hui Xie for providing the stimuli dataset, Winny W.Y. Yue for her assistance in data collection, and Ruoying Zheng for her comments on the early draft.

## Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest concerning the publication of this article.

## Fundings

The research was supported by the Ministry of Science and Technology of China STI2030-Major Projects (No. 2022ZD0214100), National Natural Science Foundation of China (No. 32171056), General Research Fund (No. 17614922) of Hong Kong Research Grants Council, and the Key Realm R&D Program of Guangzhou (No. 20200703005) to X. H.

**Open Practices**

The pre-registration, data, and analysis scripts are publicly accessible at OSF and can be accessed at https://osf.io/5e7kr/?view_only=cf903bc29f8543a19272046a45a8349c. Deviations from pre-registration and corresponding reasoning can be found in Table S1.

**Reference**

Berns, G. S., Capra, C. M., Moore, S., & Noussair, C. (2010). Neural mechanisms of the influence of popularity on adolescent ratings of music. *NeuroImage*, *49*(3), 2687–2696. https://doi.org/10.1016/j.neuroimage.2009.10.070

Claidière, N., & Whiten, A. (2012). Integrating the study of conformity and culture in humans and nonhuman animals. *Psychological Bulletin*, *138*(1), 126–145. https://doi.org/10.1037/a0025868

Connor, K. M., Davidson, J. R., Churchill, L. E., Sherwood, A., Weisler, R. H., & Foa, E. (2000). Psychometric properties of the Social Phobia Inventory (SPIN): New self-rating scale. *The British Journal of Psychiatry*, *176*(4), 379–386.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*(1), 113–126. https://doi.org/10.1037/0022-3514.44.1.113

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Edelson, M., Sharot, T., Dolan, R. J., & Dudai, Y. (2011). Following the Crowd: Brain Substrates of Long-Term Memory Conformity. *Science*, *333*(6038), 108–111. https://doi.org/10.1126/science.1203557

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.

FeldmanHall, O., Dunsmoor, J. E., Tompary, A., Hunter, L. E., Todorov, A., & Phelps, E. A. (2018). Stimulus generalization as a mechanism for learning to trust. *Proceedings of the National Academy of Sciences*, *115*(7), E1690–E1697. https://doi.org/10.1073/pnas.1715227115

Fellner, M.-C., Waldhauser, G. T., & Axmacher, N. (2020). Tracking Selective Rehearsal and Active Inhibition of Memory Traces in Directed Forgetting. *Current Biology*, *30*(13), 2638-2644.e4. https://doi.org/10.1016/j.cub.2020.04.091

Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., Duan, L., Almaliach, A., Ang, S., Arnadottir, J., Aycan, Z., Boehnke, K., Boski, P., Cabecinhas, R., Chan, D., Chhokar, J., D'Amato, A., Ferrer, M., Fischlmayr, I. C., … Yamaguchi, S. (2011). Differences Between Tight and Loose Cultures: A 33-Nation Study. *Science*, *332*(6033), 1100–1104. https://doi.org/10.1126/science.1197754

Goldenberg, A., Sweeny, T. D., Shpigel, E., & Gross, J. J. (2020). Is this my group or not? The role of ensemble coding of emotional expressions in group categorization. *Journal of Experimental Psychology: General*, *149*, 445–460. https://doi.org/10.1037/xge0000651

Goto, S. G., Cho, H. J., Park, G., Coyiuto, S. M., & Lewis, R. S. (2022). The neural processing of social norms in biculturals: The relation between cultural tightness and semantic processing. *Biological Psychology*, *170*, 108321. https://doi.org/10.1016/j.biopsycho.2022.108321

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, *97*(1), 17–41. https://doi.org/10.1037/a0015575

Harrington, J. R., & Gelfand, M. J. (2014). Tightness–looseness across the 50 united states. *Proceedings of the National Academy of Sciences*, *111*(22), 7990–7995. https://doi.org/10.1073/pnas.1317937111

Haun, D. B. M., & Tomasello, M. (2011). Conformity to Peer Pressure in Preschool Children: Peer Pressure in Preschool Children. *Child Development*, *82*(6), 1759–1767. https://doi.org/10.1111/j.1467-8624.2011.01666.x

Hopper, L. M., Schapiro, S. J., Lambeth, S. P., & Brosnan, S. F. (2011). Chimpanzees' socially maintained food preferences indicate both conservatism and conformity. *Animal Behaviour*, *81*(6), 1195–1202. https://doi.org/10.1016/j.anbehav.2011.03.002

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

Ivanchei, I. I., Moroshkina, N., Tikhonov, R., & Ovchinnikova, I. (2019). Implicit learning in attractiveness evaluation: The role of conformity and analytical processing. *Journal of Experimental Psychology: General*, *148*(9), 1505–1516. https://doi.org/10.1037/xge0000633

Izuma, K. (2013). The neural basis of social influence and attitude change. *Current Opinion in Neurobiology*, *23*(3), 456–462. https://doi.org/10.1016/j.conb.2013.03.009

Izuma, K., & Adolphs, R. (2013). Social Manipulation of Preference in the Human Brain. *Neuron*, *78*(3), 563–573. https://doi.org/10.1016/j.neuron.2013.03.023

Kaiser, D., & Nyga, K. (2020). Tracking cortical representations of facial attractiveness using time-resolved representational similarity analysis. *Scientific Reports*, *10*(1), 16852. https://doi.org/10.1038/s41598-020-74009-9

Kurdi, B., Mann, T. C., Charlesworth, T. E. S., & Banaji, M. R. (2019). The relationship between implicit intergroup attitudes and beliefs. *Proceedings of the National Academy of Sciences*, *116*(13), 5862–5871. https://doi.org/10.1073/pnas.1820240116

Lin, L. C., Qu, Y., & Telzer, E. H. (2018). Intergroup social influence on emotion processing in the brain. *Proceedings of the National Academy of Sciences*, *115*(42), 10630–10635. https://doi.org/10.1073/pnas.1802111115

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*, 213.

Lu, Y., Wang, J., Wang, L., Wang, J., & Qin, J. (2014). Neural responses to cartoon facial attractiveness: An event-related potential study. *Neuroscience Bulletin*, *30*(3), 441–450. https://doi.org/10.1007/s12264-013-1401-4

Morgan, T. J. H., Suchow, J. W., & Griffiths, T. L. (2022). The experimental evolution of human culture: Flexibility, fidelity and environmental instability. *Proceedings of the Royal Society B*. https://doi.org/10.1098/rspb.2022.1614

Mullen, T. (2012). CleanLine EEGLAB plugin. *San Diego, CA: Neuroimaging Informatics Toolsand Resources Clearinghouse (NITRC)*.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, *2011*, 1–9. https://doi.org/10.1155/2011/156869

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, *46*(3), 598.

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, *198*, 181–197.

Popal, H., Wang, Y., & Olson, I. R. (2019). A guide to representational similarity analysis for social neuroscience. *Social Cognitive and Affective Neuroscience*, *14*(11), 1243–1253.

Sassenrath, C. (2020). “Let Me Show You How Nice I Am”: Impression Management as Bias in Empathic Responses. *Social Psychological and Personality Science*, *11*(6), 752–760. https://doi.org/10.1177/1948550619884566

Shestakova, A., Rieskamp, J., Tugin, S., Ossadtchi, A., Krutitskaya, J., & Klucharev, V. (2013). Electrophysiological precursors of social conformity. *Social Cognitive and Affective Neuroscience*, *8*(7), 756–763. https://doi.org/10.1093/scan/nss064

Stallen, M., De Dreu, C. K. W., Shalvi, S., Smidts, A., & Sanfey, A. G. (2012). The Herding Hormone: Oxytocin Stimulates In-Group Conformity. *Psychological Science*, *23*(11), 1288–1292. https://doi.org/10.1177/0956797612446026

van de Waal, E., Borgeaud, C., & Whiten, A. (2013). Potent Social Learning and Conformity Shape a Wild Primate’s Foraging Decisions. *Science*, *340*(6131), 483–485. https://doi.org/10.1126/science.1232769

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063.

Werheid, K., Schacht, A., & Sommer, W. (2007). Facial attractiveness modulates early and late event-related brain potentials. *Biological Psychology*, *76*(1–2), 100–108. https://doi.org/10.1016/j.biopsycho.2007.06.008

Winkler, I., Debener, S., Müller, K.-R., & Tangermann, M. (2015). *On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP*. 4101–4105.

Zaki, J., Schirmer, J., & Mitchell, J. P. (2011). Social Influence Modulates the Neural Computation of Value. *Psychological Science*, *22*(7), 894–900. https://doi.org/10.1177/0956797611411057