Accepted Manuscript

Title: One plus one is more than two: The interactive influence of group membership and emotional facial expressions on the modulation of the affective startle reflex

Authors: Andrea Paulus, Katrin Renn, Dirk Wentura

PII: S0301-0511(18)30479-4
DOI: https://doi.org/10.1016/j.biopsycho.2018.12.009
Reference: BIOPSY 7634

To appear in:

Received date: 25 June 2018
Revised date: 11 December 2018
Accepted date: 12 December 2018

Please cite this article as: Paulus A, Renn K, Wentura D, One plus one is more than two: The interactive influence of group membership and emotional facial expressions on the modulation of the affective startle reflex, *Biological Psychology* (2019), https://doi.org/10.1016/j.biopsycho.2018.12.009

Running head: Modulation of the startle reflex by group and emotion

One plus one is more than two: The interactive influence of group membership and emotional facial expressions on the modulation of the affective startle reflex

Andrea Paulus[1], Katrin Renn[2] and Dirk Wentura[2]

[1]Technical University of Munich, Munich, Germany
[2]Saarland University, Saarbrücken, Germany

Author note:

Andrea Paulus, Chair of Psychology, Department of Psychology, Technical University Munich, Building 5, D-80333 München, Germany. E-mail: andrea.paulus @wi.tum.de. Katrin Renn, Faculty of Behavioral Sciences, Department of Psychology, Saarland University, Building A2 4, D-66123 Saarbrücken, Germany. E-mail: katrin.renn@uni-saarland.de. Dirk Wentura, Faculty of Behavioral Sciences, Department of Psychology, Saarland University, Building A2 4, D-66123 Saarbrücken, Germany. E-mail: wentura@mx.uni-saarland.de.

Corresponding author for this article is the first author. We would like to thank Ullrich Ecker for his helpful comments on an earlier version of this article. This research was supported by a grant from the German Research Foundation to Dirk Wentura and Andrea Paulus.

Highlights

- Faces typically carry several evaluatively relevant aspects concurrently
- It has not been examined if the startle reflex is sensitive to more than one aspect

- Modulation of the startle reflex by emotional in- and outgroup faces was assessed
- An influence of group membership on the social message of emotions was expected
- Results of two studies show that both factors interactively influence the startle reflex

Abstract

It has been demonstrated repeatedly that the evaluation of a face is not only influenced by the social message that is signaled by the face's emotional expression, but also by other social factors such as ethnicity or group membership. Studies in the field of startle research, however, have hitherto investigated only the effects of one of the two factors—that is, *either* emotional expression *or* group membership—on the startle response. Yet, we propose that the startle reflex is a sensitive marker for the interactive effect of both factors. Specifically, we predicted that group membership influences the social meaning signaled by an emotional expression, leading to an interactive effect. In two experiments, we examined the modulation of the startle response by happy, fearful, and angry expressions shown by ingroup and outgroup members. As predicted, an interaction between group membership and emotional expression emerged, such that happiness expressed by an ingroup member resulted in lower startle responses compared to the same expression shown by an outgroup member; the opposite pattern emerged for fearful and angry expressions. This effect was found in both experiments and independent of the exact stimulus materials employed, pointing to the generalizability of the effect.

Keywords: startle reflex; facial expressions; group membership;

One plus one is more than two: The interactive influence of group membership and emotional

facial expressions on the modulation of the affective startle reflex

The startle eyeblink reflex has been used for decades to shed light on affective processing (e.g., Bradley, Codispoti, Cuthbert, & Lang, 2001; Bradley, Codispoti, & Lang, 2006; Vrana, Spence, & Lang, 1988). This reflex, which is considered to be a primitive, defensive reflex, involves fast closure of the eyelids in response to a sudden visual or auditory pre-leading stimulus, the so-called startle probe (e.g., Lang, Bradley, & Cuthbert, 1990). Of interest for emotion research, the startle reflex is modified by an individual's affective state (Lang et al., 1990; Vrana et al., 1988): Being confronted with negative images (compared to neutral ones) during the presentation of the startle probe typically results in an augmentation of the startle reflex (i.e., a stronger and faster closing of the eyelids), while positive images (compared to neutral ones) attenuate the reflex. Typically, this influence of the affective state on the startle response has been explained by a match or mismatch, respectively, between the individual's activated motivational system and the motivational nature of the startle reflex: If the individual's appetitive system is pre-activated by the presentation of a positive stimuli, a mismatch between the activated appetitive system and the defense system activated by the startle probe occurs, resulting in inhibition of the startle reflex. If, however, the defensive system is activated by a negative stimulus as well as the subsequent startle probe, augmentation of the reflex ensues (Bradley et al., 2001; Cuthbert, Bradley, & Lang, 1996). Typically, the startle reflex is triggered by a burst of white noise played to the participants and the strength and/or latency of the following eye blink is assessed with electrodes attached over the muscle orbicularis oculi beneath the left eye (i.e., electromyography; EMG).

While most studies demonstrating startle modification have used valent-picture stimuli, some studies have also explored the influence of emotional facial expressions on the startle reflex (e.g., Anokhin & Golosheykin, 2010; Paulus, Musial, & Renn, 2014; Springer, Rosas, McGetrick, & Bowers, 2007). With facial-expression stimuli, the result patterns seem to be more nuanced; for instance, Springer and colleagues (2007) found an enhanced startle response with angry (compared to neutral) facial expressions but not with fearful expressions—even though both angry and fearful expressions are negative stimuli. Others found that angry expressions led to startle modification only if the expresser was male (Hess, Sabourin, & Kleck, 2007; Paulus et al., 2014). These results have been explained by the nature of emotional expressions. Emotional faces do not only carry information about the affective state of the expresser, but also communicate a social message: while an angry face indicates the expresser's intention to attack, a fearful face typically only signals indirect danger. Accordingly, angry faces (especially those of men, who on average are physically stronger than women) should be more relevant to the defensive system and lead to a stronger augmentation of the startle reflex. The notion that the startle reflex is influenced by the social message signaled by a face is further supported by studies showing that only maximally angry faces (and not slightly or moderately angry faces) lead to an augmentation of the startle response (Dunning, Auriemmo, Castille, & Hajcak, 2010), and that fearful faces do lead to augmentation if they are diagnostic of direct danger (Grillon & Charney, 2011).

However, there are other factors apart from emotional expression that influence face evaluations; social factors such as gender, ethnicity, or group membership also carry evaluative connotations and influence the viewer's emotional reaction and behavior. For example, studies using indirect measures like the evaluative priming paradigm or the implicit association test (IAT) have demonstrated that images of outgroup members lead to automatic negative evaluations compared to images of ingroup members (for meta-analyses, see Cameron, Brown-Iannuzzi, & Payne, 2012; Greenwald, Poehlman, Uhlmann, & Banaji, 2009). Thus, it is plausible that neutral facial expressions shown by ingroup and outgroup members may also modify the startle response. Specifically, since ingroup members tend to attract more positive evaluations than outgroup members, it can be hypothesized that neutral ingroup faces should function as positive stimuli resulting in startle-reflex inhibition, whereas neutral outgroup faces should act as negative (threatening) stimuli causing startle-reflex facilitation (Guglielmi, 1999; Vanman, Ryan, Pedersen, & Ito, 2013).

Indeed, the expected ingroup bias has repeatedly been found in studies using the startle reflex. March and Graham (2015), for example, presented Caucasian female participants pictures of Caucasian and Hispanic men. As predicted, the authors found an ingroup bias, that is, the startle reflex was increased for Hispanic faces compared to Caucasian faces.[1] Comparable results were obtained by Amodio, Harmon-Jones, and Devine (2003), who found the expected effect for participants scoring high on explicit prejudice measures. The authors presented images of neutral Caucasian or Black faces to Caucasian participants and found that participants high in explicit prejudice showed an augmentation of the startle reflex if Black faces (compared to White faces) were presented, whereas this effect was absent for participants scoring low in explicit prejudice. In addition, Phelps and colleagues (2000) found a correlation between the startle-reflex amplitude (assessed by electromyography, EMG) and the strength of amygdala activity (taken to reflect evaluative processing) when viewing White and Black faces: the greater participants' amygdala activity, the more they reacted with a blink augmentation to Black compared to White faces.

The studies presented so far show that the startle reflex is influenced by emotional expressions as well as by the group membership of the expresser. However, to the best of our knowledge, no studies in this area have investigated the influence of both these factors concurrently. This is quite surprising for two reasons: First of all, in everyday life faces are almost always characterized by more than one evaluative feature. Second, a number of

[1] It should be noted that Hispanic participants showed a similar pattern to Caucasian participants. However, the absence of an ingroup bias is not uncommon in disadvantaged groups.

studies have examined evaluative reactions other than the startle reflex to faces varying in both emotional expression *and* group membership.

For example, Bourgeois and Hess (2008) examined the influence of minimal group membership on the imitation (mimicry) of emotional faces shown by ingroup and outgroup members. They found that negative emotions (particularly sad faces) were only mimicked if the expresser was an ingroup member, whereas happy expressions were imitated regardless of the expresser's group membership; similar results were found by van der Schalk and colleagues employing expressions shown by individuals with White-Caucasian and MiddleEastern background as stimuli (2011; but see Sachisthal, Sauter, & Fischer, 2016). Using the approach-avoidance paradigm, Paulus and Wentura (2014) examined approach and avoidance reaction to happy and fearful faces shown by ingroup and outgroup members. Group membership was established via ethnicity (White Caucasian and Middle-Eastern young men, Study 1) and a modified minimal group manipulation (Study 2). The authors also found evidence for an interaction between emotional expression (happy and fearful) and group membership: Happy ingroup faces as well as fearful outgroup faces evoked relatively more approach than avoidance behavior. The opposite reaction pattern – relatively more avoidance behavior – was found with happy expressions displayed by outgroup members and fearful faces displayed by ingroup members. An interaction between emotional expression and group membership has also been observed for mood contagion (Epstude & Mussweiler, 2009), emotion or group recognition (Hugenberg, 2005; Hugenberg & Bodenhausen, 2003), and evaluative priming (Weisbuch & Ambady, 2008). These results were generally explained by an influence of group membership on the social message signaled by the emotional expression. Paulus and Wentura as well as Weisbuch and Ambady, for example, argued that happy ingroup faces are interpreted as benevolent (i.e., signalling a safe social environment and/or a desire to affiliate), whereas happy outgroup faces are interpreted as potentially malevolent (i.e., signalling rival superiority). By contrast, fearful ingroup faces are interpreted as signalling an unsafe social environment, whereas fearful outgroup faces are interpreted as signalling rival inferiority. Accordingly, happy ingroup and fearful outgroup faces should trigger a positive evaluative reaction, whereas fearful ingroup and happy outgroup faces should trigger a negative evaluative reaction, which is in line with the observed results.

Based on the above-cited evidence, one might suggest that emotional expression and group membership should also have an interactive influence on the startle reflex. However, a number of studies – again focusing on evaluative measures other than the startle reflex – have examined the simultaneous influence of group membership (employing different ethnicities) and emotional expression and found only main effects of emotion (Craig, Lipp, & Mallan, 2014) or emotion and group (Paulus & Wentura, 2018), but no interaction. Therefore it is not clear whether – and if so, how – the two factors emotional expression and group membership influence the startle response if they are present concurrently. It is the goal of the present studies to examine this question. In order to do so, we presented happy, fearful and angry expressions shown by ingroup and outgroup members to participants while a startle probe was administered. Group membership was manipulated with a modified minimal group paradigm, allowing us to present the same images as ingroup and outgroup members across participants.

Given the provided rationale and the findings that the startle reflex is influenced by the social message signaled by an emotional expression, we expected to find an interactive effect of emotional expression and group membership on the startle reflex. In accordance with Paulus and Wentura (2014) and Weisbuch and Ambady (2008), we hypothesized that happy ingroup faces should be seen as positive social signals, while happy outgroup faces should signal dominance and possible threat. In contrast, we expected the opposite pattern for fear:

We predicted that fearful ingroup faces should act as warnings, whereas fearful outgroup faces should signal submission. We thus hypothesized that happy outgroup faces should potentiate the startle reflex and happy ingroup faces should reduce it, whereas the reverse pattern should hold for fearful faces. For expressions of anger, our predictions were less clear regarding the direction of the ingroup/outgroup difference. On the one hand, one might argue that anger expressed by an outgroup member should potentiate the startle reflex to a lesser degree than anger expressed by an ingroup member: Weisbuch and Ambady showed that anger expressed by an outgroup member led to stronger anger responses in participants compared to anger expressed by an ingroup member. As anger is seen as an approach-related emotion (e.g., Carver & Harmon-Jones, 2009), leading to approach rather than avoidance reactions (e.g., Paulus & Wentura, 2016), one might assume that outgroup anger activates an anger response in participants, which activates the approach system. This should lead to a reduced startle response than anger expressed by the ingroup. On the other hand, to the extent that outgroup members are seen as dominant, outgroup anger might lead to a stronger startle response than anger expressed by an ingroup member. This is because anger expressions by dominant individuals in particular lead to an augmentation of the startle response (e.g., Hess et al., 2007; Paulus, Musial, & Renn, 2014). Since both lines of reasoning appeared plausible, we investigated the effects of anger faces in a more exploratory fashion.

Overview

We conducted two experiments in order to test our hypotheses. In both experiments, happy, fearful, and angry expressions of ingroup and outgroup members were presented to participants while an auditory startle probe was played. The startle reflex was assessed by EMG. Group membership was manipulated in a modified minimal group paradigm (see

Paulus & Wentura, 2014), such that across participants, all images were presented as ingroup as well as outgroup members in a counterbalanced fashion. There was only one difference between experiments: Experiment 1 used both male and female face stimuli for all participants; in Experiment 2, we controlled for gender and presented only same-sex images to participants.[2] The experiments' procedure was otherwise identical, hence we report method and results together (but see Appendix B for experiment-wise analyses).

Analyses were performed using linear mixed modeling, for several reasons. First, these analyses can adequately account for stimulus variability (Judd, Westfall, & Kenny, 2012). This was especially pertinent since Experiments 1 and 2 used different materials. Second, preliminary data inspection indicated a habituation effect of the startle reaction across trialsequence position. Linear mixed modeling allowed us to control for the influence of this effect (see below). Linear mixed modeling is akin to conduct multiple regression analyses across trials for each participant (and a subsequent test of mean regression weights on deviance from zero). Thus, we can use trial-sequence position and its quadratic term to account for the habituation and test whether our predictors emotion, group, and their interaction term account for additional variance. Since linear mixed models need rather large data sets, we present an analysis on the combined data of Experiment 1 and 2. This decision was further supported by the fact that there was no indication in the data for a difference in results between the two experiments (see also Appendix B).

All the data as well as the material of the two experiments is openly accessible at https://osf.io/ujf4k/?view_only=cea6507fce2543348e8586381aa602dc. We report all measures, manipulations and exclusions for our study.

---

[2] Preliminary analysis of Experiment 1 data using GLM indicated that face gender moderated the main results. However, the gender effect was not corroborated in the more adequate LMM analysis and it will thus not be discussed further (see also Appendix B). We nevertheless replicated Experiment 1 while controlling for the influence of the factor gender.

Method

Participants

Across both experiments, a total of 87 participants (46 female, 41 male) participated. Participants received a reimbursement of €12. The median age was 24 years, ranging from 19 to 36 years.

In Experiments 1, 36 participants (18 female, 18 male) participated. Data from an additional six participants were discarded prior to analysis: Three participants withdrew during the learning phase (see below); one was unable to correctly distinguish between in- and outgroup members in a final test. As is common for startle experiments, two participants were classified as non-responders as they showed no detectable startle reaction on more than 64 %

of the trials.

Since we know of no study that has examined the interactive influence of group membership and emotional expression on the startle reflex, we based a-priori power calculations for Experiment 1 on the assumption of a medium-sized effect ($d_Z = 0.50$) for the critical 2 (group membership: ingroup vs. outgroup) × 2 (emotional expression: happy vs. fear) interaction effect. Detecting a medium-sized effect with 1- β = .80 (α = .05) requires a sample size of $N = 34$ (actual power with $N = 36$ was 1- β = .83.)

In Experiments 2, 51 participants (28 female, 23 male) participated. Data from an additional nine participants were discarded prior to analysis: One participant withdrew during the learning phase (see below); one withdrew from the startle task due to a pre-existing tinnitus condition; two experienced technical problems with the amplifier. As in Experiment 2, five participants were classified as non-responders as they showed no detectable startle reaction on more than 64 % of the trials.

For Experiment 2, we based our power calculations on the effect size found in Experiment 1, which was in fact a bit smaller than medium-sized ($d_Z = 0.36$). To detect this effect with 1- β = .80 (α = .05, one-tailed),[3] a sample size of $N = 50$ is necessary (actual power with $N = 51$ was 1-β = .81.) Power calculations were done with G*Power (Faul, Erdfelder, Lang, & Buchner, 2007).

Design

The study followed a two (group: ingroup vs. outgroup) by three (emotional expression: happy, fear, angry) within-participants design.

Materials

In Experiment 1, the material consisted of sixty-four emotional face images selected from various databases (Langner et al., 2010; Olszanowski et al., 2015; Paulus, Rohr, Neuschwader, Seewald, & Wentura, 2012; Van Der Schalk, Hawk, Fischer, & Doosje, 2011). The complete stimulus set contained 16 angry, happy, fearful, and neutral expressions, respectively. Experiment 1 used a mixed pool of male and female face images; each facial expression was displayed by eight men and eight women (i.e., the same eight male and eight female faces displayed all four expressions). The images were selected based on the rate of accurate emotion recognition (all above 85 %). The 16 neutral expressions were only used in the learning phase. The remaining 48 happy, fear, and angry faces were used in the startle task. To manipulate group membership, we created two stimulus sets containing eight individuals each, one representing the ingroup and one representing the outgroup (counterbalanced across participants). Each set contained happy, fearful, and angry expressions (displayed by four men and four women), respectively. The two stimulus sets did not differ in terms of emotion-recognition rates or attractiveness and intensity ratings, all $|t|$'s

[3] Our hypothesis was that the fear-happy difference is larger for ingroup compared to outgroup faces. Therefore, a plan based on one-tailed testing was appropriate; it allowed for a somewhat smaller sample size, given the rather time-consuming procedure.

< 1. The images were edited in a way that the face and the top of the neck were shown on a white background with a straight orientation and gaze directed towards the viewer. In each set, the images of six individuals (three men and three women) served as critical items, meaning that a startle probe was presented during their presentation. The emotional faces of two further individuals per stimulus set (one man and one woman) were used for filler trials with no startle probe. The startle probe was a burst of white noise (50ms, 100 db[A]) presented binaurally via headphones.

In Experiment 2, stimulus selection and presentation was the same as for Experiment 1. The only difference was that for Experiment 2, two stimulus sets were created, one containing only images of men and one containing only images of women. This modification enabled us to show only same-sex images to participants in Experiment 2, eliminating the influence of expresser's sex as an additional factor. Thus, in Experiment 2, each expression was displayed by 16 men *or* 16 women.

Procedure

The procedure comprised three parts: The modified minimal group manipulation, a learning phase, and the startle task. Everything was equal for the two experiments.

Modified minimal group manipulation. The aim of the first part of the experiment was to use a modified minimal group manipulation (see Paulus & Wentura, 2014) to assign participants to one of two groups. This involved a mock assessment of their alleged personality style: After being seated in front of the computer and providing informed consent, participants were told that the first part of the experiment involved a questionnaire to assess whether they had a *basal* or a *focal* personality style. A set of twenty self-report statements (e.g., "I am moody") were presented on the computer screen. For each statement, participants had to indicate to what extent it described themselves accurately. Participants were then informed of their alleged personality style, apparently based on their self-report responses. In truth, however, all participants were classified as having a *basal* personality style. Finally, participants were given specific information about the two different personality styles. People with a basal personality style were described as sociable, agreeable, socially minded, and balanced, but also occasionally sloppy and forgetful. In contrast, people with a focal personality style were characterized in a more negative way, as selfish, reckless, manipulative, occasionally aggressive, but also skillful and intelligent. The styles were modeled in a specific way to reinsure that the majority of students would identify with the *basal* style (and no one with the *focal* style). The negative features were added to the *basal* style, and the positive features to the *focal* style, to increase the plausibility of the cover story. For the remainder of the experiment, we expected participants to categorize others with the same *basal* personality style as ingroup members and those with a *focal* style as outgroup members.

Learning phase. The goal of the learning phase was for participants to learn the assignment of face stimuli to the ingroup (*basal* style) and outgroup (*focal* style). The cover story was that we were examining the influence of personality style on performance in a facelearning task. To this end, we presented the sixteen neutral-expression faces with first names and an indication of *basal* versus *focal* personality style (i.e., identifying each face as ingroup or outgroup). After presentation of the faces, participants completed an extensive learning phase, during which they had to categorize the face images as ingroup and outgroup members. In order to strengthen the manipulation, we added two silhouettes to the each stimulus set. One silhouette was representing the participant (as an ingroup member) and – for reasons of symmetry – the silhouette of another anonymous participant (as an outgroup member) was added. The learning phase only terminated if the participant correctly classified all faces in three consecutive blocks.

Startle task. The final part of the experiment assessed the startle probe reaction to emotional faces. The cover story (for the application of electrodes) was that gaze direction of emotional facial expressions would be measured. Presentation of noise bursts were mentioned; participants were instructed to ignore them. A female experimenter tested the participants individually. The experiment consisted of two blocks, each block containing 24 trials. In 18 of these trials, a startle probe was administered during stimulus presentation (critical trials) and in the remaining 12 trials no startle probe was presented (filler trials). In critical trials, the probe was randomly administered three, four, or five seconds after image onset. On 11 trials, we additionally administered a startle probe during the inter-trial-interval (ITI). The filler trials, the probe-onset jitter, as well as probe presentation during the ITI aimed to reduce predictability of the startle probe.

Each trial started with a fixation cross displayed in the center of the computer screen for 1,000 ms, followed by the presentation of an emotional facial expression for another 6,000 ms. Five seconds after offset of the facial image, participants were asked to indicate whether the face belonged to a person with a *basal* or a *focal* personality style and which emotion (happiness, fear, anger) it displayed; they were also asked to rate the level of arousal and dominance of the displayed expression, on seven-point scales ranging from 1 (not arousing/submissive) to 7 (arousing, dominant). The main reason for including these questions was to subjectively shorten the relatively long ITI and keep participants focused. Finally, a message saying “Please wait while your data is processed!” was displayed for a variable period of time, resulting in an ITI between 20 s and 25 s in total.

Participants completed four practice trials before the main experiment. Practice trials used emotional expressions of individuals who were used on filler trials in the main experiment. Practice trials involved three startle probes that were administered during the stimulus presentation and one probe presented during the ITI. Throughout the experiment, each image was presented only once. The images were presented in a pseudorandom order such that the same emotional expression and the same group membership was never presented on more than two consecutive critical trials, and the same individual was never presented on consecutive trials.[4]

Before the practice phase, participants were asked to write their personality style on a piece of paper. The sheet was placed next to the keyboard to ensure participants were aware of their group membership. At the end of the experiment, participants filled in a questionnaire regarding the group manipulation ("To what extent did the result of the personality test correspond with your own impressions regarding your personality? To what extent do the individual characteristics of the other personality style also apply to you?"; scale ranged from 0, totally disagree, to 7, totally agree). Participants were then thanked, paid, and dismissed. They were completely debriefed via email after the completion of data collection.

Physiological Recordings and Data Reduction

The eyeblink component of the startle response was measured using EMG. Two 6-mm Ag/AgCl electrodes with an inter-electrode distance of 1 cm were attached over the orbicularis oculi beneath the left eye. A ground electrode was placed on the mastoid behind the left ear; a reference electrode was placed on the forehead. Before attaching the electrodes, the skin was cleaned with alcohol. Impedance was constantly below 20 kΩ. The blink response was recorded with a V-Amp 16 amplifier (Brain Products Inc.) with a sampling rate of 2,000 Hz.

The EMG data was analyzed offline using Brain Vision Analyzer Software (version 2.0.1; Brain Products Inc.). The data was filtered (30-500 Hz, 50 Hz notch filter, moving average of 50 ms), rectified and visually inspected. We excluded trials with an unstable baseline (i.e., trials with excessive noise, spontaneous blinking) from further analysis (2.81 %

[4] We chose this approach as controlling for emotion, group membership, and individual constrained the degrees of freedom for the trial sequence. However, presenting stimuli in a fixed order is common in startle research (e.g., Anokhin & Golosheykin, 2010; Dillon & LaBar, 2005). We control for the influence of trial sequence by employing linear mixed modeling.

of trials). Startle-response amplitude was defined as the peak muscle activity of the orbicularis oculi in the 20-120 ms time window after probe onset relative to the baseline (the 50-ms period before probe onset). A higher value thereby corresponds to stronger muscle activity (i.e., a stronger startle reflex) and (presumably) indicates a stronger activation of the defensive system. We z-standardized the data of each participant according to Blumenthal et al. (2005). Finally, outlier values (i.e., values at least 1.5 interquartile ranges below the second quartile of the individual participant's distribution and values at least 1.5 interquartile ranges above the third quartile of the individual participant's distribution) were discarded (5.72 % of all trials).

Results

Manipulation Check

Across both experiments, participants correctly identified the group-membership during the startle task in 93.10 % of all critical trials. In order to assess if participants identified with the assigned group (*basal*), we analyzed those questions of the postexperimental questionnaire addressing this issue. The results showed that participants identified highly (M = 5.46, SD = 1.34) with the personality style (*basal*) assigned to them and to a lesser extend (M = 2.49, SD = 1.98) with the personality style of the outgroup (*focal*). The results of the explicit dominance and arousal ratings of the stimuli are reported in the appendix since they were not of particular interest for our study.

Startle Response

To run linear mixed models, we used the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017), which is based on lme4 (Bates, Mächler, Bolker, & Walker, 2015), of the R environment for statistical computing (R-Core-Team, 2016). The lmerTest package allows estimation of degrees of freedom (using Satterthwaite's approximation) and thus *p*-values for the tests of regression weights. In Appendix B, experiment-wise analyses can be found.

The fixed variables of our model were *emotion* (fear vs. happiness vs. anger; coding see below), *group* (ingroup vs. outgroup), *emotion* × *group*, and – because preliminary data inspection indicated a habituation effect of the startle reaction across trial-sequence position (z-standardized; i.e., centered) - its quadratic term. *Emotion* was dummy-coded with happiness as the reference condition; thus, contrast 1 represents the contrast happiness versus fear, which was our a-priori focus, given our earlier studies. Contrast 2 represents the contrast happiness versus anger. We allowed random intercepts and slopes for participants and random intercepts for items.[5] Table 1 presents the results.[6]

Experiment: zSequence: Trial-sequence position (z-standardized); Emotion (Fear): Fear = 1, Happy = 0, [Anger = 0]; Emotion (Anger): Fear = 0, Anger = 1 [Happy = 0]; Group: Ingroup = +1, Outgroup = -1.
The right-most column indicates the *p*-value for the regression weight corresponding to the interaction term [row factor] × experiment.

The results showed an influence of the habituation component (i.e., zSequence plus quadratic term), indicating that the magnitude of the startle reflex declined over the course of the experiment. Most importantly, however, there were significant interactions of emotion and group for the contrasts happiness versus fear (contrast 1) and happiness versus anger (contrast 2). Figure 1 shows the means for the cells of the emotion × group design.

Decomposing the interactions by analyzing the data separately by emotions yielded a main effect of group for happy faces, *B* = -0.043, that was associated with *t*(747.30) = 1.68, *p*

[5] It is common practice to run a model with random slopes and a model without random slopes and to report the latter one if a model comparison indicates no significant difference in overall model fit and to report the former one if there is a difference. This was the case here, $\chi^2(35) = 93.72$, $p < .001$.
[6] We repeated the analysis adding experiment as a further predictor (incl. all interaction terms); the results can be easily summarized: First, the parameter values of Table 1 changed only marginally; second, all terms including experiment were clearly non-significant, all *p*s > .19 (*p* = .327 and *p* = .679 for Emo. [Fear] × Group × Experiment and Emo. [Anger] × Group × Experiment, respectively.)

= .094. As predicted, happy ingroup faces were associated with a numerically lower startle response than happy outgroup faces. The group main effect was reversed for fear, *B* = 0.056, *t*(138.68) = 2.17, *p* = .032; as predicted, fear expressed by an ingroup member led to a significantly stronger startle response than fear expressed by an outgroup member. We also observed a group difference for anger, which had the same direction as the effect for fear: *B* = 0.059, *t*(117.22) = 2.22, *p* = .028; anger expressions by an ingroup member were associated with a stronger startle response than anger expressions by an outgroup member.

Discussion

The presented experiments examined the interactive influence of two evaluative features of a face – namely emotional expression and group membership – on the startle reflex. Since the presence of more than one evaluative feature in the face is probably the norm rather than exception, we regarded this as an issue worthy of exploration. Previous research had only demonstrated the separate impacts of an emotional expression's social message and group-membership of (neutral) expressers on the startle response. We hypothesized that the startle response should be influenced by the interaction of both features; specifically, we predicted that the social message signaled by an emotional face should modulate the startle reflex, but that the perceived message should be influenced by group membership.

We investigated this issue in two experiments following the exact same procedure: After a modified minimal group manipulation, participants were exposed to happy, fearful, and angry expressions shown by ingroup and outgroup members, while a startle probe was presented. The minimal group manipulation allowed us to present the same images as either ingroup or outgroup stimuli across participants.

The results supported our assumption of an interaction between group membership and emotional expression. The effect was in the predicted direction: Happiness expressed by an ingroup member led to a reduced startle response compared to happiness expressed by an outgroup member. The opposite pattern emerged for fearful expressions, with fear expressed by an ingroup member leading to a greater startle response compared to fear expressed by an outgroup member. Responses to anger expressions followed the pattern observed for expressions of fear. Results were not moderated by the experiment factor, indicating that both experiments yielded essentially the same results. This is noteworthy because the two experiments employed different stimulus materials, thus suggesting some generalizability of results.

We interpret our results in terms of the social message signaled by an emotional expression: While happiness expressed by an ingroup member signals the positive intention to affiliate, happiness expressed by an outgroup member might be seen as a sign of dominance; this is in line with the observed results. A fear expression might be seen as a warning if shown by an ingroup member, but as a sign of submission if shown by an outgroup member. Accordingly, the pattern of results should be opposite to that of happy expressions, which is precisely what we observed.

Our reasoning regarding the influence of group membership on the social message signaled by happy and fearful expressions is supported by other studies examining similar research questions, albeit with different dependent variables (Paulus & Wentura, 2014; Weisbuch & Ambady, 2008). The correspondence between the present results and those observed by Paulus and Wentura (2014) is especially noteworthy: Paulus and Wentura examined manual approach and avoidance reactions to happiness and fear expressed by ingroup and outgroup members. They observed faster approach reactions to ingroup happiness and outgroup fear, compared to outgroup happiness and ingroup fear. Thus, despite marked surface-level differences between dependent variables across studies, the underlying theoretical assumptions regarding the activation of a motivational approach/avoidance system by relevant stimuli were identical, and so were the observed result patterns.

For expressions of anger, our predictions were less definite than those for happy and fearful faces. Based on previous research, two alternative hypotheses were proposed: On the one hand, anger is often seen as an approach-related affect (Carver & Harmon-Jones, 2009), and it has been shown that expressions of anger can lead to approach reactions (e.g., Paulus & Wentura, 2016). Furthermore, outgroup anger tends to lead to stronger anger responses than ingroup anger (Weisbuch & Ambady, 2008). Therefore, we reasoned that anger expressed by an outgroup member should activate the approach system, resulting in a reduced startle response compared to anger expressed by an ingroup member. On the other hand, it has been argued that anger expressions potentiate the startle response if they are expressed by dominant individuals (Hess et al., 2007; Paulus et al., 2014). Since we characterized ouroutgroup as dominant and potentially aggressive, it seemed plausible to assume that anger expressed by an outgroup member might be seen as more threatening than anger expressed by an ingroup member, thus leading to a stronger startle response. The results, however, supported the first hypothesis: Outgroup anger led to a reduced startle response compared to ingroup anger. As argued above, this pattern can be explained by participants showing a stronger anger response to the negative social message signaled by outgroup anger compared to ingroup anger. Therefore, this finding can likewise be subsumed under the general social message hypothesis that guided our hypotheses regarding the happy and fearful faces.

A skeptical reader might ask whether our results could also be explained by the valence of the presented emotional expressions instead of their associated social message: The negative emotional expressions anger and fear both produced the same pattern of results, whereas happiness, a positive emotional expression, led to the opposite pattern of results. However, we believe it is unlikely that expression valence alone is responsible for the observed effects: First of all, numerous studies examining the modulation of the startle reflex by emotional expressions have shown that different negative expressions influence the startle reflex in different ways (e.g., Springer et al., 2007), pointing to a different process than valence. Second, previous research (Paulus et al., 2014) has shown that gender differentially influences the modulation of startle reactions by expressions of anger and fear. This finding indicates that this modulation does not arise from the valence of the expressions (which is the same) but rather from a different aspect of the emotional expression. Third, research examining the influence of group membership on reactions to emotional expressions – but employing dependent variables other than the startle reflex (e.g., Van der Schalk et al., 2011; Weisbuch & Ambady, 2008) – has found that group membership differentially influences responses to negative emotional expressions. Therefore, we believe that even though anger and fear expressed by ingroup and outgroup members lead to a comparable modulation of the startle reflex , the processes by which they modulated the reflex differ. We assume that startle-reflex modulations are only indicate of an activation of the approach or the avoidance system, respectively, and cannot shed light on the reason underlying that activation. For example, both the experience of anger and joy can activate the approach system, leading to an inhibition of the startle reflex. The underlying emotional experience is, however, quite obviously different.

A further point to discuss is that Figure 1 seem to indicate that emotional expression only modulates the startle reaction for ingroup faces, whereas outgroup faces seem to homogeneously produce a relatively neutral effect, irrespective of emotional expression. This interpretation (i.e., that emotional expressions shown by outgroup members do not modulate the startle response) might be seen as contradicting our hypothesis that group membership influences the social meaning of an emotional expression. However, we believe that on a general level, such an interpretation still fits our “social meaning” hypothesis: Group status modifies the processing of the social message of emotional expressions.

At a more specific level, however, our preferred hypothesis is to assume a reversal of the ingroup emotion effects for outgroups (“reversal” hypothesis; see *Introduction*) instead of a null effect for outgroups (“null effect” hypothesis). A null effect for outgroup faces would imply that outgroup status neutralizes the effect of the social message signaled by facial expressions instead of changing it. Even though this interpretation would fit our reasoning at a more general level, we would have to refine our specific conclusion in that case. For example, one might hypothesize that ingroup faces are more socially relevant and emotions communicated by ingroup members will therefore be processed with more attention whereas social signals on outgroup faces are ignored.

However, there exists a different explanation of the apparent null effect: The homogeneity of the three means for outgroup expressions might simply be an arbitrary side effect of a general emotion main effect which indicates that – on average – startle responses are smaller with happy faces compared to angry and fearful faces. To be more specific about this, assume that for a portion of trials (and/or participants) the group category is not automatically processed and will therefore not influence the processing of the emotional expression. If these trials produce a standard emotion main effect whereas those trials with group categorization produce a perfectly disordinal interaction pattern, results as observed will emerge.

If we focus solely on the present study, we might arrive at a standoff in weighing the pros and cons for the two specific hypotheses (i.e., the "reversal" hypothesis and "null effect" hypothesis). On the one hand, methodologically Ockham's razor might be an argument in favor of the "null effect" hypothesis instead of our "reversal" hypothesis because in the latter case an additional emotion main effect must be assumed. On the other hand, we should not disregard that our hypothesis was a priori and well founded, whereas the at first sight simpler hypothesis is post hoc and is not very plausible on second thought (e.g., to ignore the angry expression on an outgroup face might not have been an adaptive process in the long run). Moreover, if we put the present study into context, we can concede that the interpretation in terms of the "reversal" hypothesis fits into a larger body of research and is therefore the to be preferred one. Taken together, our results provide evidence that group membership and emotional expression interactively influence the modulation of the startle reflex. These results can be explained by an influence of the respective group membership on the social message signaled by emotional expressions. They therefore converge with results from studies employing other dependent variables (Paulus & Wentura, 2014; Weisbuch & Ambady, 2008) and point to the importance of considering more than one evaluative feature of a face when examining the modulation of the startle reflex.

References

Amodio, D. M., Harmon-Jones, E., & Devine, P. G. (2003). Individual differences in the activation and control of affective race bias as assessed by startle eyeblink response and self-report. *Journal of Personality and Social Psychology, 84*, 738–753.

Anokhin, A. P., & Golosheykin, S. (2010). Startle modulation by affective faces. *Biological Psychology*, *83*, 37–40.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390-412. doi:10.1016/j.jml.2007.12.005

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Blumenthal, T. D., Cuthbert, B. N., Filion, D. L., Hackley, S., Lipp, O. V., & Van Boxtel, A. (2005). Committee report: Guidelines for human startle eyeblink electromyographic studies. *Psychophysiology*, *42*, 1–15.

Bourgeois, P., & Hess, U. (2008). The impact of social context on mimicry. *Biological Psychology, 77*, 343–352.

Bradley, M. M., Codispoti, M., Cuthbert, B. N., & Lang, P. J. (2001). Emotion and motivation I: defensive and appetitive reactions in picture processing. *Emotion*, *3*, 276–298.

Bradley, M. M., Codispoti, M., & Lang, P. J. (2006). A multi-process account of startle modulation during affective perception. *Psychophysiology, 43*, 486–497.

Cameron, C. D., Brown-Iannuzzi, J. L., & Payne, B. K. (2012). Sequential priming measures of implicit social cognition: A meta-analysis of associations with behavior and explicit attitudes. *Personality and Social Psychology Review*, *16*, 330–350.

Carver, C. S., & Harmon-Jones, E. (2009). Anger is an approach-related affect: evidence and implications. *Psychological Bulletin*, *135*, 183–204.

Craig, B. M., Lipp, O. V., & Mallan, K. M. (2014). Emotional expressions preferentially elicit implicit evaluations of faces also varying in race or age. *Emotion*, *14*, 865–877.

Cuthbert, B. N., Bradley, M. M., & Lang, P. J. (1996). Probing picture perception: Activation and emotion. *Psychophysiology*, *33*, 103–111.

Dillon, D. G., & LaBar, K. S. (2005). Startle modulation during conscious emotion regulation is arousal-dependent. *Behavioral Neuroscience*, *119*, 1118–1124.

Dunning, J. P., Auriemmo, A., Castille, C., & Hajcak, G. (2010). In the face of anger: Startle modulation to graded facial expressions. *Psychophysiology*, *47*, 874–878.

Epstude, K., & Mussweiler, T. (2009). What you feel is how you compare: How comparisons influence the social induction of affect. *Emotion, 9*, 1-14. doi:10.1037/a0014148

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

Fox, J. (2003). Effect displays in R for generalised linear models. *Journal of Statistical Software, 8*, 1–27.

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*, 17–41.

Grillon, C., & Charney, D. R. (2011). In the face of fear: anxiety sensitizes defensive responses to fearful faces. *Psychophysiology, 48*, 1745–1752.

Guglielmi, R. S. (1999). Psychophysiological assessment of prejudice: Past research, current status, and future directions. *Personality and Social Psychology Review, 3*, 123–157.

Hess, U., Sabourin, G., & Kleck, R. E. (2007). Postauricular and eyeblink startle responses to facial expressions. *Psychophysiology, 44*, 431–435.

Hugenberg, K. (2005). Social categorization ant he perception of facial affect: Target race moderates the response latency advantage for happy faces. *Emotion, 5*, 267-276.

Hugenberg, K., & Bodenhausen, G. V. (2003). Facing prejudice: Implicit prejudice and the perception of facial threat. *Psychological Science, 14*, 640-643. doi:10.1046/j.0956-7976.2003.psci_1478.x

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*, 54–69.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*, 1–26.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1990). Emotion, attention, and the startle reflex. *Psychological Review, 97*, 377–395.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H., Hawk, S. T., & Van Knippenberg, A. D. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and Emotion*, *24*, 1377–1388.

March, D. S., & Graham, R. (2015). Exploring implicit ingroup and outgroup bias toward Hispanics. *Group Processes & Intergroup Relations*, *18*, 89–103.

Olszanowski, M., Pochwatko, G., Kuklinski, K., Scibor-Rylski, M., Lewinski, P., & Ohme, R. K. (2015). Warsaw set of emotional facial expression pictures: a validation study of facial display photographs. *Frontiers in Psychology*, *5*, 1516.

Paulus, A., Musial, E., & Renn, K. (2014). Gender of the expresser moderates the effect of emotional faces on the startle reflex. *Cognition and Emotion*, *28*, 1493–1501.

Paulus, A., Rohr, M., Neuschwader, M., Seewald, M., & Wentura, D. (2012). Validation Data of the Saarbrücken Emotion Database. (Unpublished data)., *Saarland University, Germany*.

Paulus, A., & Wentura, D. (2014). Threatening joy: Approach and avoidance reactions to emotions are influenced by the group membership of the expresser. *Cognition and Emotion*, *28*, 656–677. https://doi.org/10.1080/02699931.2013.849659

Paulus, A., & Wentura, D. (2016). It depends: Approach and avoidance reactions to emotional expressions are influenced by the contrast emotions presented in the task. *Journal of Experimental Psychology: Human Perception and Performance*, *42*, 197.

Paulus, A., & Wentura, D. (2018). Implicit evaluations of faces depend on emotional expression and group membership. *Journal of Experimental Social Psychology*, *77*, 143-154.
https://doi.org/10.1016/j.jesp.2018.04.004

Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Funayama, E. S., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. *Journal of Cognitive Neuroscience*, *12*, 729–738. R-Core-Team. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Sachisthal, M. S., Sauter, D. A., & Fischer, A. H. (2016). Mimicry of ingroup and outgroup emotional expressions. *Comprehensive Results in Social Psychology*, *1*, 86–105.

Springer, U. S., Rosas, A., McGetrick, J., & Bowers, D. (2007). Differences in startle reactivity during the perception of angry and fearful faces. *Emotion*, *7*, 516–525.

Van der Schalk, J., Fischer, A., Doosje, B., Wigboldus, D., Hawk, S., Rotteveel, M., & Hess, U. (2011). Convergent and divergent responses to emotional displays of ingroup and outgroup. *Emotion*, *11*, 286–298.

Van Der Schalk, J., Hawk, S. T., Fischer, A. H., & Doosje, B. (2011). Moving faces, looking places: validation of the Amsterdam Dynamic Facial Expression Set (ADFES). *Emotion*, *11*, 907–920.

Vanman, E. J., Ryan, J. P., Pedersen, W. C., & Ito, T. A. (2013). Probing prejudice with startle eyeblink modification: A marker of attention, emotion, or both. *International Journal of Psychological Research*, *6*, 30–41.

Vrana, S. R., Spence, E. L., & Lang, P. J. (1988). The startle probe response: a new measure of emotion? *Journal of Abnormal Psychology*, *97*, 487–491.

Weisbuch, M., & Ambady, N. (2008). Affective divergence: automatic responses to others' emotions depend on group membership. *Journal of Personality and Social Psychology*, *95*, 1063–1079.

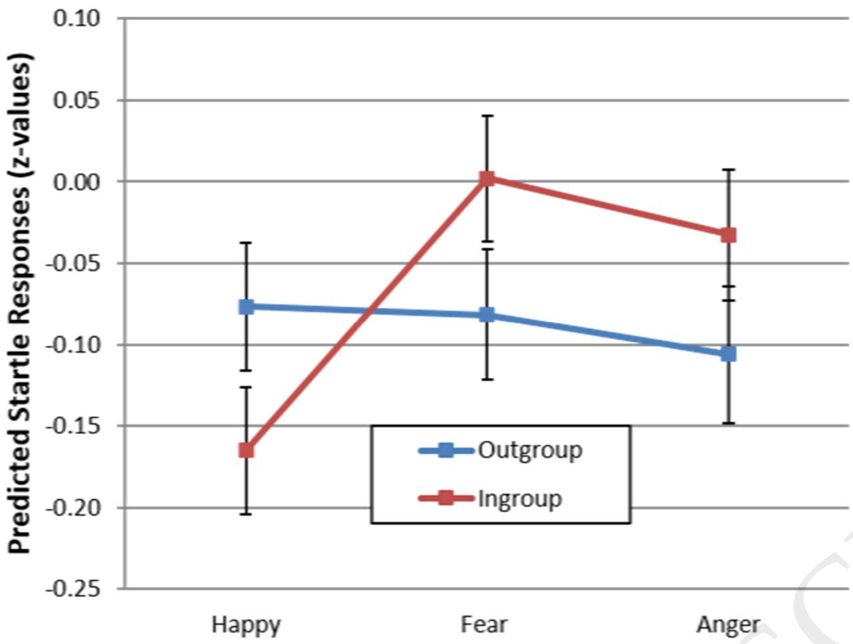

Figure 1. Estimated startle values for emotion × group design. *Note:* Means and whiskers (+/- one standard error of the mean) were computed by R procedure effect (Fox, 2003).

Table 1 Results of the linear mixed model analysis

Fixed Factor	Weight	*SE*	*df*	*t*	*p*
Intercept	-0.254	0.036	90.9	7.05	< .001
zSequence	-0.274	0.025	92.1	10.84	< .001
$zSequence^2$	0.143	0.020	104.3	7.25	< .001
Emotion (Fear)	0.079	0.044	65.6	1.81	.075
Emotion (Anger)	0.049	0.045	68.0	1.10	.277
Group	-0.044	0.025	715.0	1.79	.075