This preprint is published in:

Bublatzky, F., Guerra, P., & Alpers, G. W. (2020). Watch out, he's dangerous! Electrocortical indicators of selective visual attention to allegedly threatening persons. *Cortex, 131*, 164-178.

Watch out, he´s dangerous!

Electrocortical indicators of selective visual attention to allegedly threatening persons

Florian Bublatzky [1,2]*, Pedro Guerra [3], Georg W. Alpers [2]

[1] Central Institute of Mental Health Mannheim, Medical Faculty Mannheim/Heidelberg University, Germany

[2] Clinical Psychology and Biological Psychology and Psychotherapy, Department of Psychology, School of Social Sciences, University of Mannheim, Germany

[3] University of Granada, Department of Personality, Spain

* Correspondence to: Florian Bublatzky, Central Institute of Mental Health, J5, 68159 Mannheim, Germany, Phone: +49-621-1703-4461, florian.bublatzky@zi-mannheim.de

Abstract

The face of a friend indicates safety, the face of a foe can indicate threat. Here, we examine the effects of verbal instructions ('beware of this person') on the perception of unknown persons. Focusing on visual attention, face identity and facial expression information is examined during instructed threat-of-shock or safety. However, shocks never occurred. Participants quickly acquired instructed threat associations, and electrocortical processing differentiated threat- from safe-identities as well as emotional and neutral facial expressions. Importantly, face encoding varied as a joint function of identity and facial expression, as revealed by pronounced N170 amplitudes to smiling threat-identities. Moreover, instructions readily reversed previously learned affective associations leading to attention allocation and memory updating as reflected by N170, EPN and P3 amplitudes toward new threat-identities displaying angry expressions. These findings demonstrate that person perception flexibly re-adjusts according to minimal information. Intriguingly, perceptual biases occur even though the anticipated aversive consequence does not occur, with implications for research on stereotyping and anxious psychopathology.

Key words: person perception, instructional learning, reversal learning, emotional facial expression, face identity, EEG/ERP

Word counts:

Abstract 160, introduction 1458, methods 1458, results 1802, discussion 1906

5 figures, 2 tables

1. Introduction

Recognizing facial information, such as person's identity or their emotional expressions, is a crucial function for interacting appropriately with other people. This becomes apparent, for instance, in people suffering from facial neglect or prosopagnosia who are unable to recognize faces, even when those are highly familiar (e.g., family members; Young et al., 1990). Moreover, facial expressions are an important source of information about the social environment. For instance, a fearful face might indicate a threatening situation and a smiling person could signal safety. However, depending on who displays a facial expression, the meaning can be ambiguous (e.g., a smiling foe expressing schadenfreude) and flexibly changed by means of new information (e.g., reversal instructions; Atlas, 2019; Bublatzky et al., 2018, 2019). Focusing on the interaction of face identity and facial emotion, the present study examined face and person perception as a function of social learning.

Learning about potential threats is vital for organizing adequate behavior. The same is true for learning about conditions that might signal safety. Based on first hand experiences (e.g., Pavlovian conditioning), various brain structures have been shown to be relevant for the formation and extinction of threat associations (e.g., amygdala, ventro-medial prefrontal cortex; Milad & Quirk, 2012). This so-called fear network is subject to dysregulations, which presumably contribute to the development and maintenance of psychopathological fear and anxiety (Etkin & Wager, 2007) but also prejudices and stereotyping (Amodio, 2014). However, much of our everyday learning rests upon social interactions. Such social learning processes, based on observations or verbal instructions, are less risky because we do not have to go through harmful situations before we learn to avoid them (Askew & Field, 2008; Olsson & Phelps, 2007).

An experimental approach to investigate the effects of verbal instructions is the threat-of-shock paradigm (Grillon et al., 1991). Participants are instructed that they might receive aversive stimuli (e.g., electric shocks) when a specific cue is presented (e.g., blue square), whereas another cue signals safety (e.g., green circle). Such threat instructions have been shown to trigger selective attention toward threat cues (Bublatzky & Schupp, 2012; Robinson et al., 2013) and activate the fear network similar to studies using experiential threat learning (Koban et al., 2017; Mechias et al., 2010; Olsson & Phelps, 2007). Building upon this, instructed threat primes defensive response programs, activating the autonomic nervous system and motor-behavioral reflexes (e.g., enhanced skin conductance responses and potentiated startle reflex; Bradley et al., 2005; Bublatzky et al., 2013, 2014a). Moreover, verbal instructions have proven very effective in reversing threat to safety when a new threat cue is concurrently established (i.e., reversal learning; Atlas, 2019; Schiller & Delgado, 2010). For instance, recent studies observed immediate and complete attenuation of defensive responding when participants were told that a previously instructed threat cue now signaled safety (Bublatzky et al., 2018, 2019; Costa et al., 2015; Mertens & De Houwer, 2017). Thus, an increasing number of studies demonstrate the role that verbal instructions can play in both instantiating and attenuating defensive reactions (Atlas, 2019; Costa et al., 2015; Koban et al., 2017; Mertens et al., 2018). However, the modulation of visual attention to social signals that are instructed and reversed threat or safety cues is not well understood.

Electrocortical measures are particular well suited to investigate the temporal dynamics of face and person perception as a function of threat or safety. Given the importance to efficiently ‘read and understand’ face identity and facial emotions, such information has been suggested to be processed preferentially. This is revealed by early occipito-temporal brain potentials, which are sensitive to facial stimuli (N170; Bentin et al., 1996), emotional facial expressions (early posterior negativity [EPN]; Schupp et al., 2004), and identity learning (N250; Kaufmann et al., 2009). Moreover, later elaborate face processing (LPP; 400-700 ms; Schupp et al., 2004) has been shown to be biased in social phobia and in participants undergoing aversive anticipation (Bublatzky et al., 2014b; Wieser et al., 2010). These processing differences presumably reflect the activity of distinct brain systems specialized for the analyses of visual facial appearance (e.g., fusiform face area, posterior superior temporal sulcus), and extended systems that mediate processing of face identity and facial expression (e.g., anterior temporal cortex, medial PFC, and amygdala; for a review see Haxby & Gobbini, 2011).

Importantly, face and person perception is no isolated process and changes as a function of contextual settings, social situations and knowledge about a person (e.g., Wieser & Brosch, 2012; Kim et al., 2004). For instance, face identity processing varies with associated facial expressions (Aguado et al., 2012), surrounding people´s facial expressions (Bublatzky et al., 2017), affective background scenes (Righart & de Gelder, 2008), and body posture of the displayed person (Meeren et al., 2005). Moreover, verbal statements about a target person modulate early and late stages of face processing (e.g., EPN and LPP) and are enhanced by negative and positive descriptions (e.g., "She/he thinks your voice is pleasant/annoying"; Wieser et al., 2014). Interestingly, the impact of such affective descriptions is particularly pronounced when they are explicitly directed at the observer (e.g., self-relevant physically or socially threatening; Klein et al., 2015; McCrackin & Itier, 2018; Wieser et al., 2014).

The present study tested the key hypothesis that verbally acquired knowledge about another person, that he/she might be dangerous, amplifies person perception and attentional processing. Participants were told that certain individuals indicated shock threat whereas others signaled safety (e.g., Person A and B cue shocks, and Person C and D cue safety). Building upon previous work on electrocortical processing of facial expressions of emotions and social threat learning, we predicted that threat instructions would modulate face processing as evinced by a range of ERP components. Specifically, enhanced P1 and late positive potential (LPP) amplitudes were expected to reflect vigilance and elaborated stimulus processing towards threat relative to safe identities (Baas et al., 2002; Bublatzky & Schupp, 2012). In addition to the face identity information, which signaled threat or safety, these persons further displayed happy, neutral, and angry facial expressions. Here, face- and emotion-sensitive components (N170, EPN, and LPP) were predicted to be associated with more negative N170 and EPN, as well as more positive LPP amplitudes for emotional compared to neutral faces (Bublatzky et al., 2014; Hinojosa et al., 2015; Schupp et al., 2004; Pourtois et al., 2004).

Of particular interest was the interaction between verbal threat/safety instructions and visual facial expressions on the perception of a person. If face identity and facial expression information are processed in distinct neuronal systems, as suggested by earlier models on face perception (cf. Bruce & Young, 1986), threat effects should occur regardless of the displayed emotional expression. This would support a general threat-sensitization hypothesis suggesting a processing advantage for visual signals of danger (i.e., more pronounced P1 and LPP amplitudes to instructed threat relative to safe identities regardless of their facial expression; Bublatzky & Schupp, 2012). Alternatively, a potential 'aggressor' displaying facial emotions might amplify selective emotion (N170, EPN, LPP) and/or threat cue processing (P1, LPP). This may result in enhanced amplitudes to specifically emotional (but not neutral) threat relative to safe identities. Thus, a significant interaction of person identity by facial expression information would support recent models' assumptions that question the notion of clear-cut independent processing systems in face perception (Young & Bruce, 2011).

Going beyond the initial instantiation of threat or safety, flexibly updating such affective associations by means of verbal communication plays an important social function (Kringelbach & Rolls, 2003). Here, a second instructional manipulation served to examine the reversal of threat and safety associations in person perception (e.g., Person B who was considered threatening before, now becomes safe), and a partial reversal design was used to examine reversed compared to unchanged threat/safety cues (e.g., Person C was previously safe and maintains cueing safety; Costa et al., 2015). By changing preexisting associations from threat to safety and vice versa, reversal learning is important to adapt to new information concerning our fellow human beings. Such reversal learning processes, linked to face identity and emotional expressions, have been shown to involve the prefrontal and anterior cingulate cortices as well as modulations of the P3 amplitudes (Atlas, 2019; Koban et al., 2017; Willis et al., 2010). For instance, reduced P3 amplitudes have been observed when participants were cued to switch associations formed with angry faces (Willis et al., 2010). Moreover, regarding reversal of instructed threat, one previous study showed that the LPP was more pronounced to instructed threat relative to safety cues, even after repeated reversal instructions (Bublatzky & Schupp, 2012). Thus, reversal learning has been suggested to be associated with rather later processing stages (as indicated by P3 and LPP components) involved in motivated attention, memory update, and (re)appraisal processes regarding perceived self-relevance of threatening information (Bublatzky & Schupp, 2012; Blechert et al., 2012; Muench et al., 2016).

2. Methods

2.1 Participants

Thirty-three healthy volunteers (7 males) were recruited from the students of the University of Mannheim, Germany. Participants′ age ranged from 18 to 28 (*M* = 21.7, *SD* = 2.8). Sample size was chosen based on our previous research using facial expressions and instructed threat manipulations (e.g., Bublatzky & Schupp, 2012; Bublatzky et al., 2014b). Moreover, statistical estimations with G*Power (Faul et al., 2007), indicated that a sample size of N = 33 was required to detect instruction by facial expression effects at a medium effect size (power = .95, α error = .05, and assumed correlation of repeated measures in repeated measure ANOVAs = .4). No preliminary analyses were conducted before completing the sample. The sample was within the normal range of state and trait anxiety (STAI, *M* = 36.5 and 38.5, *SD* = 8.1 and 9.9), social anxiety (SPIN, *M* = 16.4, *SD* = 8.0), and depression (BDI, *M* = 7.7, *SD* = 5.9)[1]. All participants were informed about the general study procedures before providing informed consent. This included information about non-painful electrical stimulation and that a full explanation of experimental procedures and objectives would be provided in the debriefing.

[1] Several covariation effects emerged for questionnaire scores and ERP components. However, as no a priori hypotheses were specified, these effects are not reported, but will serve as piloting data for follow-up studies.

Participants received course credits for participation. The local ethics committee approved the study protocol.

2.2 Materials and presentation

Face pictures of four actors[2] (2 females; 1024 × 768 pixels) displaying happy, neutral, and angry facial expressions were selected from the Karolinska Directed Emotional Faces set (KDEF; Lundqvist et al., 1998). To focus on early attentional processes and to obtain a sufficient number of trials per condition, stimuli were presented as a rapid picture stream for 1 s each directly followed by the next picture (see Figure 1).

In three experimental blocks, the full set of 12 pictures was repeated 30 times amounting to 360 trials per block. Block 1 served as a control condition without specific instructions (passive viewing task). For Block 2 (threat/safety instantiation), two specific face identities (1 female and 1 male actor) were instructed as threat cues and the other two actors as safety cues (e.g., Person A and B indicate threat, C and D safety). In Block 3 (instructed reversal), previous threat and safety associations were partially reversed (Costa et al., 2015). Each one person maintained cueing threat or safety, and the meaning of the other two identities was reversed (e.g., Person A and C maintain cueing threat and safety, but Person B now newly indicates safety and Person D cues threat). Assignment of face actors to condition was balanced across participants. Each participant viewed an individual picture sequence, and several constraints were implemented to account for potential picture sequence effects (e.g., Flaisch et al., 2008ab; Schweinberger & Neumann, 2016). (1) Randomization of face pictures was restricted to no more than three repetitions of the same face identity and facial expression in a row, (2) equal transition probabilities between face identities and facial expression categories, and (3) no immediate repetition of the same face identity displaying the same emotional expression.

[2] KDEF identifiers: af20, af25, am09, am10.

Picture presentation was controlled using Presentation software (Neurobehavioral Systems, Inc., Albany, CA).

Pictures were presented on a 22-inch screen located 1 meter in front of the participant. For the shock work-up procedure, electrical stimulation (maximum 10 mA, 100ms) was manually applied by using Digitimer Stimulator DS-5 (Digitimer Ltd, UK) with electrodes attached to the tip of the non-dominant index finger. It is important to note that no shocks were administered during the main experiment (Blocks 1 - 3). While previous research demonstrated the persistence of instructed threat effects within and across repeated test sessions even without any shock application (Bublatzky et al., 2012, 2013, 2014a), this was done to examine the mere aversive anticipation but not experience of shocks associated with face identities.

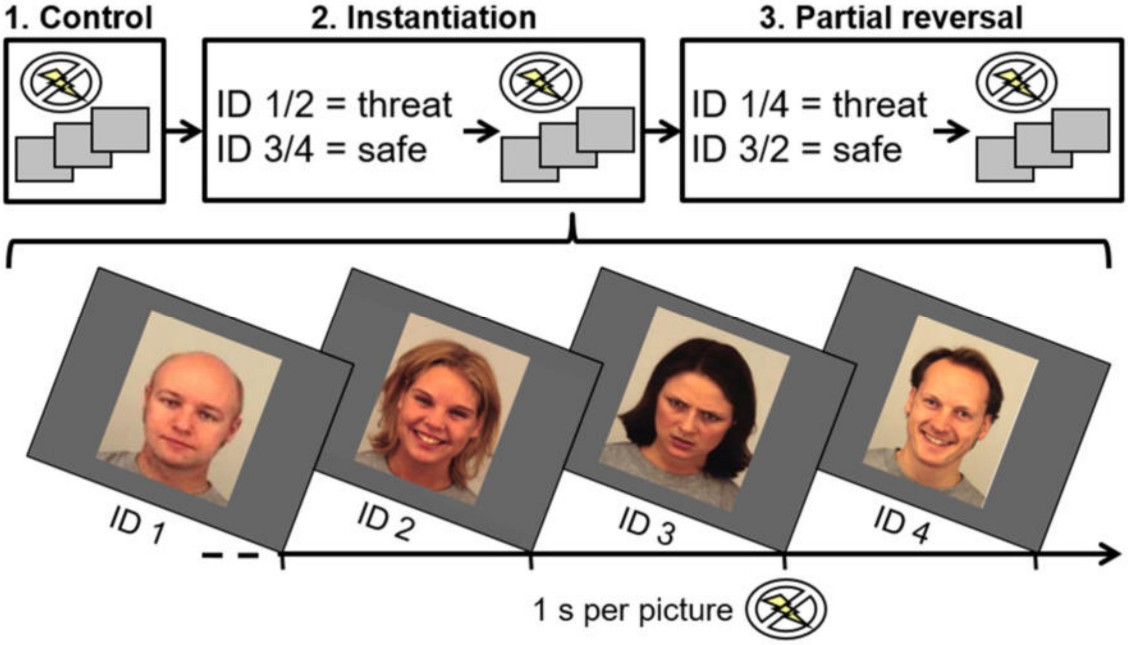


Figure 1. Schematic illustration of the experimental procedure. The first experimental block served as a control condition, in which participants were told to passively view all face pictures. Afterwards verbal instructions were given regarding which face identity (ID) is cueing threat or safety (instantiation block). To this end, two face identities were pointed out as cues for aversive shocks, whereas the other two identities served as instructed safety cues. In the partial reversal block, threat and safety associations were partially changed. Each one identity maintained cueing threat or safety, the associations of the other two identities were reversed

(e.g., ID 1 and 3 maintain cueing threat and safety, but ID 2 now newly indicates safety and ID 3 newly cues threat). For each block, all four face identities were presented intermixed displaying happy, neutral and angry expression (360 trials) in a rapid serial picture stream. No shocks were applied throughout the experiment. With permission, face pictures are depicted from the KDEF (http://kdef.se/).

2.3 Procedure

After the EEG sensors were attached, participants completed several questionnaires (on depression, general and social anxiety) and were seated in a dimly lit and sound-attenuated room. A practice run (12 picture trials) served to familiarize participants with the picture viewing procedure. Next, a brief shock work-up was carried out to ensure the credibility of the threat-of-shock instructions and to adjust shock intensity individually (Bublatzky et al., 2010). To this end, up to 8 stimulations were applied ranging from below perceptual threshold until participants rated the intensity as ‘maximally unpleasant but not yet painful’ (maximum of 10 mA, 100 ms). Participants were then instructed that during the experiment the shock intensity would be equal to the most unpleasant test stimulus. Afterward, three experimental blocks were presented with the main instruction to passively view all pictures presented (control Block 1). For threat/safety instantiation (Block 2), participants were instructed that they might receive up to three electrical shocks when viewing instructed threat (e.g. Person A and B), but not safety identities (e.g. Person C and D). Finally, for Block 3, partial reversal instructions were given, stating that threat/safety contingency changed (e.g. Person A and C now cue threat, and Person B and D safety). After each block, participants rated the hedonic valence, arousal, and perceived threat of the four face identities using the Self-Assessment Manikin (SAM; Bradley & Lang, 1994) and a visual analog scale ranging from *not at all* to *highly threatening* (1 to 10). Finally, participants were debriefed.

2.4 EEG recording and data reduction

Electrocortical activity was recorded using a 64-channel actiCap system (BrainProducts, Munich, Germany) with Ag/AgCl active electrodes located according to the 10-10 system. The continuous EEG was sampled at a rate of 500 Hz with FCz as the reference electrode and was filtered on-line (0.1-100Hz) with BrainAmp DC amplifiers and VisionRecorder software (BrainProducts). Sensor impedance was kept below 10 kΩ. For offline data preprocessing, sampling rate was reduced to 250Hz and converted to an average reference. Data was low-pass filtered at 35Hz and eye-movements were screened and interpolated using the inbuilt automated independent component analyses for ocular correction (ICA; using Fp1 and F7 for detecting vertical and horizontal eye-movements, using VisionAnalyzer 2.1 software; BrainProducts). On average 4.4 trials per block were removed (control *M* = 4.15, *SD* = 10.47; instantiation *M* = 5.21, *SD* = 7.06; partial reversal *M* = 3.84, *SD* = 6.57) based on several artifact criteria (i.e. maximal allowed voltage step of 50 µV/ms; maximal allowed difference of values in 200 ms intervals of 200 µV; and minimal/maximal allowed amplitude of +/-100 µV). Building upon this, baseline correction (200 ms prior to picture onset) was applied and stimulus-synchronized epochs were extracted lasting from 200 ms before to 800 ms after picture onset. Finally, separate averaged waveforms were calculated for each experimental block (control, instantiation, reversal), Instruction (threat, safety), and Facial Expression (happy, neutral, angry), for each sensor and participant.

2.5 Data analyses

Self-reported threat, valence, and arousal ratings were analyzed with 3 Block (control[3] vs. instantiation vs. reversal) × 2 Instruction (threat vs. safety) repeated measures ANOVAs; facial expressions were not rated separately. For the reversal block, additional analyses included the

[3] Note: As the control block did not contain threat or safety instructions, an artificial data split (even vs. odd trial numbers) was undertaken to adjust the factor structure.

factor Contingency (maintain vs. reversed) to examine the effects of partial reversal of threat/safety contingencies. Specifically, each one face identity maintained cueing threat and safety (i.e. maintain threat and maintain safe identity), but the meaning of the other two face identities was changed so that one previous safe identity now cued threat and one previous threat identity now cued safety (i.e. safe-to-threat and threat-to-safe reversal).

To examine the effects of threat and safety instructions on the electrocortical processing of face identity and facial expressions, a two-step procedure was used. Based on previous research (Blechert et al., 2012; Bublatzky & Schupp, 2012; Bublatzky et al., 2014b; Pourtois et al., 2004), ERP components, sensors, and time-windows of interest were determined by visual inspection. Following this, repeated measure ANOVAs were based on mean area scores for the selected locations. The P1, N170 and Early Posterior Negativity components were scored over parieto-occipital sensors (PO9 and PO10) between 100-132 ms (P1), 160-210 ms (N170), and 260-360 ms (EPN). In addition, the P3 component was scored over fronto-central sensors (FC1 and FC2) between 280-320 ms, and the Late Positive Potential over central sensor sites (P1 and P2) between 400-600 ms after picture onset.

For each ERP component, the mean activity of the selected sensors and time windows were entered into separate repeated measures ANOVAs. The most basic design was tested for the control block including the factors Facial Expression (happy vs. neutral vs. angry) and Laterality (left vs. right hemisphere). For the instantiation block, the factor Instruction (threat vs. safety) was added to test verbal learning effects on facial emotion processing. Finally, for the analyses of the reversal block, the additional factor Contingency (maintain vs. reversed) was included to examine the effects of partial reversal of threat/safety contingencies.

Greenhouse-Geisser corrections were used when relevant, and as a measure of effect size the partial η^2 (η^2_P) is reported. To control for Type 1 error, Bonferroni correction was applied for all *t*-test pairwise comparisons.

3. Results

3.1 Rating data

Face identities, which were instructed to signal shock threat, were perceived as more threatening, unpleasant, and arousing as compared to instructed safe identities (*Fs*(1,32) = 23.50, 16.81 and 31.46, *ps* < .001, η_p^2 = .42, .34 and .50; see Figure 2 and Table 1 for *M, SD*, and 95% CI). Picture ratings for threat, valence, and arousal changed across experimental Blocks (*Fs*(2,64) = 4.79, 13.51 and 7.41, *ps* < .05, < .001 and < .01, η_p^2 = .13, .30 and .19) and varied as a joint function of Instruction by Block (*Fs*(2,64) = 10.24, 8.46 and 20.75, *ps* < .01, η_p^2 = .24, .21 and .39). For the control block, pairwise comparisons did not indicate differences regarding threat, valence, or arousal ratings (*ps* = .17, .87 and .38). Instructed threat relative to safe identities were perceived as more threatening, unpleasant, and arousing in the following instantiation (all *ps* < .01) and reversal blocks (all *ps* < .001). Thus, threat and safety instructions effectively changed the evaluation of face identity according to the instructed threat/safety contingencies.

Regarding the impact of reversal instructions, additional analyses for the partial reversal block included the factor Contingency (maintain vs. reverse). Importantly, significant interactions Instruction by Contingency emerged for threat, valence, and arousal ratings (*Fs*(1,32) = 9.98, 14.28, and 8.79, *ps* < .01, η_p^2 = .24, .31, and .22). Pairwise comparisons showed that face identities who maintained cueing threat relative to safety were perceived as more threatening, unpleasant, and arousing (all *ps* < .001). In contrast, reversal instructions flexibly changed face evaluation (i.e. from threat to safety and vice versa) for the threat and arousal ratings (*ps* < .05 and .01) but not significantly for the valence ratings (*p* = .08).

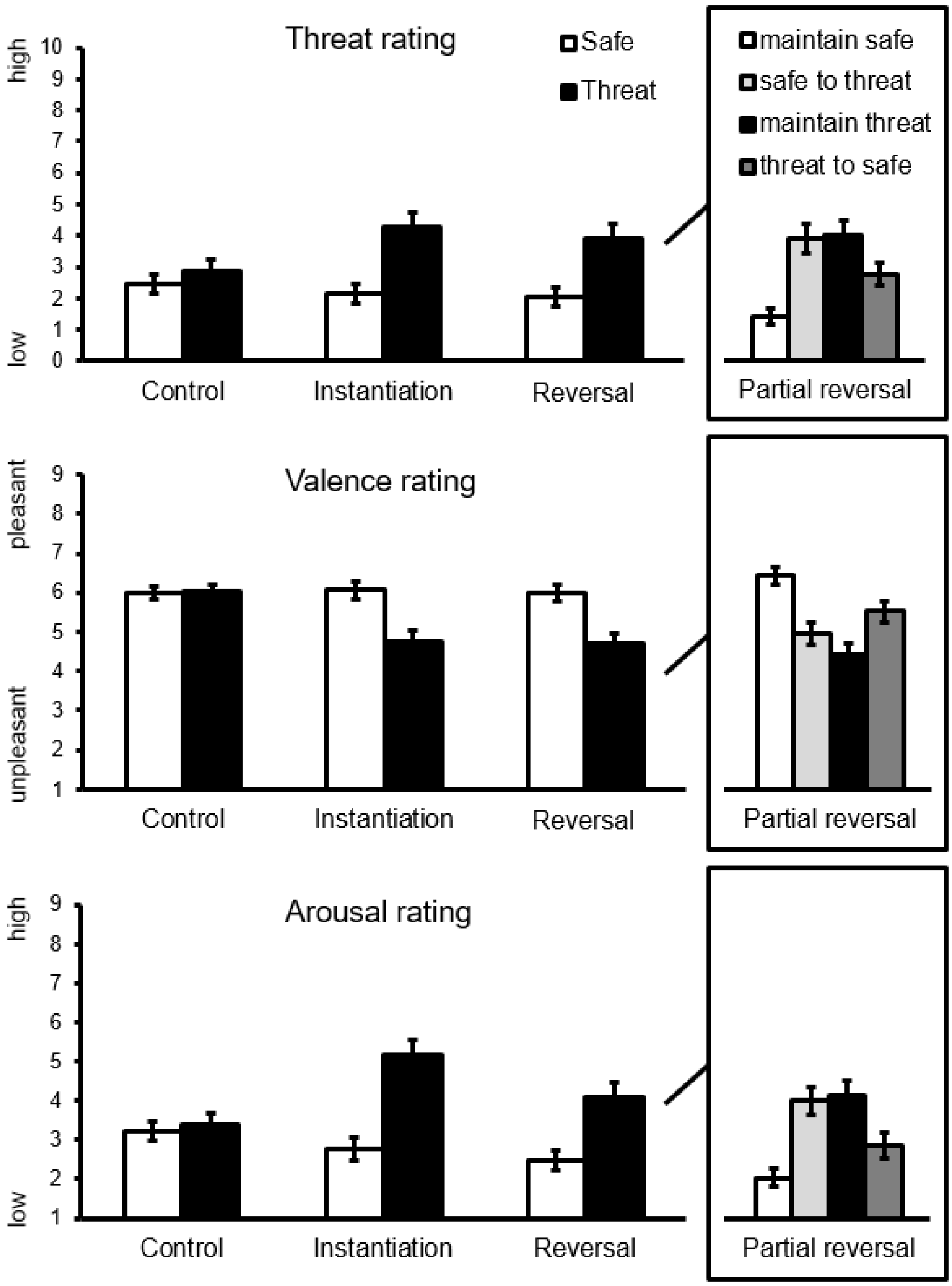


Figure 2. Self-reported threat, valence, and arousal as a function of experimental condition for the control, instantiation, and reversal block. Illustrated are averaged ratings (SEM) for face identities that were instructed as threat or safety cues, no separate facial expression ratings were obtained. For the control conditions (no threat instruction), means are based on an artificial data split. To illustrate the effects of partial reversal instructions, the significant interactions Instruction by Contingency (maintain vs. reversed) are displayed on the right side.

3.2 Event-related brain potentials

3.2.1 Passive picture viewing control block

During the passive viewing control block, electrocortical face processing varied as a function of Facial Expression (see Figure 3, Table 2 for *M, SD*, and 95% CI). An emotional modulation of the ERP waveforms emerged as early as 100 ms after picture onset for the P1 ($F(2,64) = 5.32$, $p < .01$, $\eta_p^2 = .14$) and later for the N170 components over parieto-occipital sensor sites ($F(2,64) = 8.50$, $p < .01$, $\eta_p^2 = .21$). Specifically, angry facial expressions were associated with larger P1 amplitudes compared to neutral ($F(1,32) = 9.47$, $p < .01$, $\eta_p^2 = .23$), but not significantly compared to happy faces ($F(1,32) = 4.14$, $p = .05$, $\eta_p^2 = .12$); happy and neutral expressions did not differ ($F(1,32) = 1.69$, $p = .20$, $\eta_p^2 = .05$). Similarly, the N170 component differentiated angry from both happy and neutral faces ($Fs(1,32) = 8.10$ and 10.75, $ps < .01$, $\eta_p^2 = .20$ and $.25$), which did not differ from each other ($F(1,32) = 1.80$, $p = .19$, $\eta_p^2 = .05$). Neither the P1 nor N170 revealed effects involving Laterality ($Fs < 1.22$, $ps > .28$, $\eta_p^2 < .04$).

Moreover, the early posterior negativity (EPN) and fronto-central P3 components confirmed previous findings indicating selective emotion processing ($Fs(2,64) = 11.88$ and 9.72, $ps < .001$ and $< .01$, $\eta_p^2 = .39$ and $.23$). Pronounced differences were found for both happy and angry faces relative to neutral expressions for the EPN ($Fs(2,64) = 20.18$ and 16.87, $ps < .001$, $\eta_p^2 = .27$ and $.35$) as well as for the P3 component ($Fs(2,64) = 12.42$ and 16.28, $ps < .01$ and $< .001$, $\eta_p^2 = .28$ and $.34$). No differences emerged between happy and angry faces regarding EPN and P3 components ($Fs(1,32) = 1.43$ and 2.99, $ps = .24$ and $.09$, $\eta_p^2 = .04$ and $.09$). For the EPN a more pronounced negativity was observed over the left compared to the right hemisphere ($F(1,32) = 10.46$, $p = .003$, $\eta_p^2 = .25$). No further effects including Laterality reached significance for none of the other components ($Fs < 2.13$, $ps > .13$, $\eta_p^2 < .06$).

Regarding more elaborate later stimulus processing, the late positive potential (LPP) varied as a function of Facial Expressions ($F(2,64) = 4.98$, $p < .05$, $\eta_p^2 = .14$). A pronounced positivity was found for angry compared to both happy and neutral faces ($Fs(1,32) = 13.23$ and

5.66, *ps* < .01 and < .05, η_p^2 = .29 and .15), which did not differ from each other (F(1,32) = .34, p = .57, η_p^2 = .01). Thus, as in previous research, passively viewing facial expressions was associated with a pattern of selective emotion processing which varied across the visual processing stream.

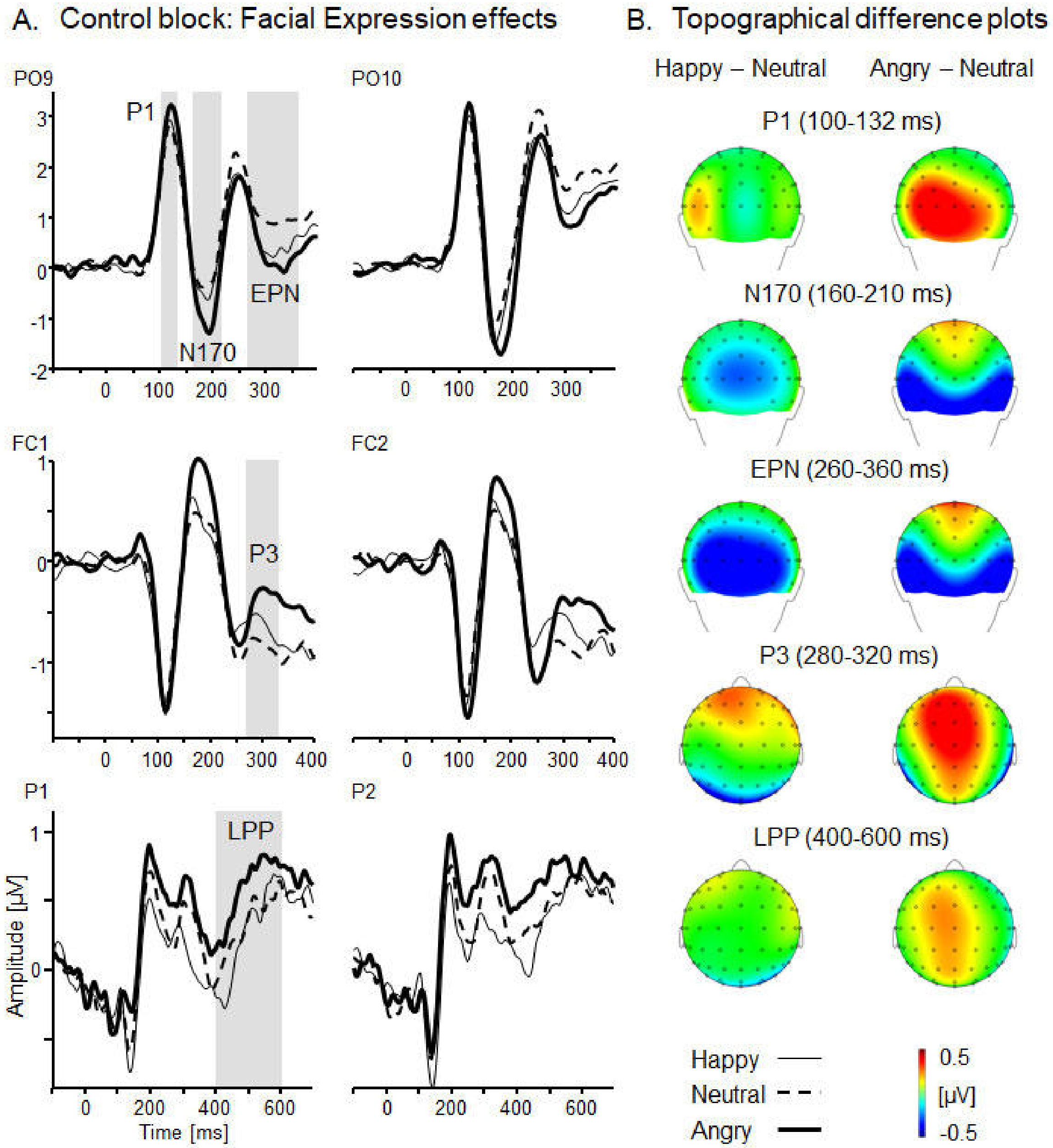


Figure 3. (A) Event-related brain potential waveforms as a function facial expression for the passive viewing control block (left and right sensors). Grey-shaded area markers highlight the time windows, which were used for statistical calculations. (B) Topographical difference plots (happy – neutral and angry – neutral) displaying emotion effects for the P1, N170, EPN, P3 and LPP components. Waveform differences are displayed on the backside or top of a model head.

3.2.2 Instantiation of threat/safety associations

For the instantiation block, participants received instructions that specific face identities now would serve as cues for shock threat while other identities would cue safety, and all faces were presented displaying happy, neutral, and angry facial expressions. Whereas the P1 did not reveal any main or interaction effects (*Fs* < 2.76, *ps* > .07, η_p^2 < .08), importantly, the N170 component showed a significant interaction between Instruction and Facial Expression (*F*(2,64) = 3.47, *p* < .05, η_p^2 = .10; see Figure 4). Follow-up tests indicated that instructed threat relative to safe face identities were associated with more pronounced occipital negativity when they displayed happy facial expressions (*F*(1,32) = 5.75, *p* < .05, η_p^2 = .15), but no instruction effects were observed for neutral or angry facial expressions (*Fs*(1,32) < .01, *ps* = .96 and .76, η_p^2 < .01). Moreover, the N170 component was modulated by Facial Expression (*F*(2,64) = 22.64, *p* < .001, η_p^2 = .41). Relative to neutral faces, a pronounced negativity was observed for angry and happy expressions (*F*(1,32) = 14.9 and 6.20, *p* < .001 and < .05, η_p^2 = .32 and .16), which did not differ from each other (*F*(1,32) = 3.58, *p* = .067, η_p^2 = .10). These emotion effects were more pronounced over the right hemisphere (*F*(2,64) = 3.96, *p* < .05, η_p^2 = .11). For the N170 component, no further effects reached significance (*Fs* < 3.8, *ps* > .06, η_p^2 < .11).

Whereas no further ERP component revealed an interaction of Instruction by Facial Expression (*Fs*(2,64) < 1.81, *ps* > .17, η_p^2 < .05), Instruction main effects were observed for the EPN and LPP components (*Fs*(1,32) = 4.17 and 8.88, *ps* = .05 and <.01, η_p^2 = .12 and .22). Specifically, viewing instructed threat relative to safety face identities was associated with more negative EPN and more positive LPP amplitudes, suggesting selective attention to threat cues and in-depth stimulus elaboration. No Instruction effect was observed regarding fronto-central P3 (*F*(1,32) = .89, *p* = .353, η_p^2 = .03).

Selective emotion processing was indicated by pronounced EPN and fronto-central P3 effects (*Fs*(2,64) = 11.90 and 13.21, *ps* < .001, η_p^2 = .27 and .29), but not significantly for the LPP (*F*(2,64) = 2.77, *p* = .073, η_p^2 = .08). Specifically, the EPN was more negative for happy and angry compared to neutral faces (*F*(1,32) = 8.04 and 7.78, *ps* < .01, η_p^2 = .20); happy and angry faces did not differ (*F*(1,32) = .15, *p* = .71, η_p^2 < .01). Moreover, EPN amplitudes were more negative over the left hemisphere (*F*(1,32) = 4.58, *p* < .05, η_p^2 = .13), although Laterality did not interact with either Facial Expression or Instruction (*Fs*(2,64) = 2.25 and .78.21, *ps* = .12 and .38, η_p^2 = .07 and .02). Larger P3 amplitudes were observed for both happy and angry compared to neutral faces (*Fs*(1,32) = 7.16 and 24.56, *ps* < .05 and < .001, η_p^2 = .18 and .43), but no significant differences emerged for angry faces compared to happy ones (*F*(1,32) = 3.57, *p* = .068, η_p^2 = .10). Taken together, in the instantiation block, ERPs differentiated instructed threat from safe identities and emotional compared to neutral facial expressions. Importantly, the N170 component revealed interactive effects of Instruction and Facial Expression, showing most pronounced processing differences for smiling threat identities.

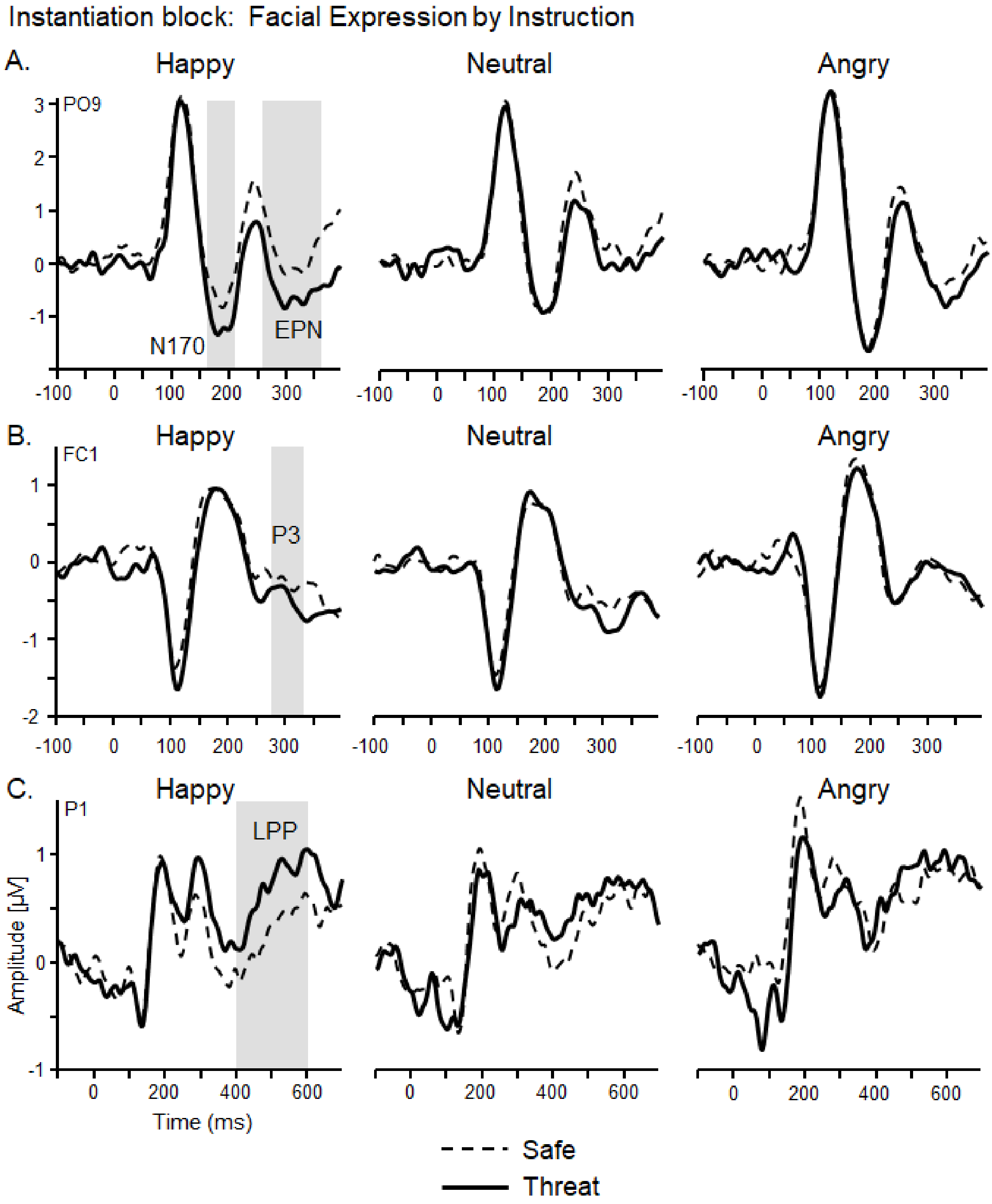


Figure 4. Event-related brain potential waveforms as a function of Facial Expression (happy, neutral, angry) and Instruction (threat, safety) for the instantiation block. (A) Illustration of the N170 interaction effect and main effect Facial Expression for the N170, EPN and (B) for fronto-central P3, as well as the Instruction main effects for the EPN and LPP (C). Exemplary left sensors are depicted. Grey-shaded area markers highlight the time windows, which were used for statistical calculations.

3.2.3 Partial reversal of instructed threat/safety

Partial reversal instructions readily changed the previously instructed threat and safety associations. To focus on the impact of these add-on instructions, the additional factor Contingency (maintain vs. reversed) was introduced for the reversal block. The P1 component did not vary as a function of Instruction, Contingency, Laterality, or any higher order interaction ($Fs < 2.31$, $ps > .11$, $\eta_p^2 < .07$). However, the P1 differed for Facial Expression ($F(2,64) = 11.53$, $p < .001$, $\eta_p^2 = .27$), with larger amplitudes for angry and happy relative to neutral faces ($Fs(1,32) = 23.69$ and 4.59, $p < .001$ and $< .05$, $\eta_p^2 = .43$ and 13), and angry compared to happy expressions ($F(1,32) = 7.23$, $p < .05$, $\eta_p^2 = .18$).

Contingency main effects emerged for the N170, EPN, and P3 components ($Fs(1,32) = 19.20$, 7.48 and 15.38, $ps < .001$, $< .05$, and $< .001$, $\eta_p^2 = .38$, $.19$, and $.33$). Specifically, face identities whose threat/safety associations were reversed (e.g., new threat or safety cues) led to more negative N170 and EPN, as well as more positive P3 amplitudes compared to those face identities who maintained cueing threat or safety (as they did in the instantiation block). Interestingly, these effects were qualified by higher-order interactions Contingency by Facial Expression for the N170, EPN, and P3 components ($Fs(2,64) = 11.01$, 8.76 and 3.41, $ps < .001$, $< .01$ and $< .05$, $\eta_p^2 = .26$, $.22$ and $.10$). Reversal instructions were particularly effective in changing the processing of angry faces ($Fs(1,32) = 33.94$, 31.71 and 20.46, $ps < .001$, $\eta_p^2 = .52$, $.50$ and $.39$; Figure 5), but not that of happy expressions $Fs(1,32) = .74$, $.05$ and $.47$, $ps = .40$, $.82$ and $.50$, $\eta_p^2 < .02$). For neutral faces, the N170 component showed more negative amplitudes when these were displayed by faces whose meaning were reversed relative to maintained ($F(1,32) = 4.77$, $p < .05$, $\eta_p^2 = .13$). EPN and fronto-central P3 amplitudes did not show Contingency effects for neutral faces ($Fs(1,32) = .07$ and 2.06, $ps = .79$ and $.16$, $\eta_p^2 < .06$). Moreover, an interaction Contingency by Laterality emerged for the N170 ($F(1,32) = 5.22$, $p < .05$, $\eta_p^2 = .14$) showing more pronounced negativity for reversed face identities over the right hemisphere. The only other Laterality effect emerged for the EPN with more negative amplitudes over the left hemisphere ($F(1,32) = 6.82$, $p < .05$, $\eta_p^2 = .18$). No significant effects were found regarding the LPP component as a function of Contingency ($F(1,32) = 3.0$, $p = .093$, $\eta_p^2 = .09$) or Contingency by Facial Expression interaction ($F(2,64) = 1.57$ $p = .22$, $\eta_p^2 = .05$).

Finally, selective processing of instructed threat compared to safe identities was observed for the EPN, P3 and LPP components ($Fs(1,32) = 12.16$, 5.70 and 6.14, $ps < .05$, $\eta_p^2 = .28$, .15 and .20), but not for the P1 and N170 ($Fs(1,32) = .33$ and .01, $ps = .57$ and .94, $\eta_p^2 < .01$). No further main or interaction effects were observed ($Fs < 2.39$, $ps > .11$, $\eta_p^2 < .07$).

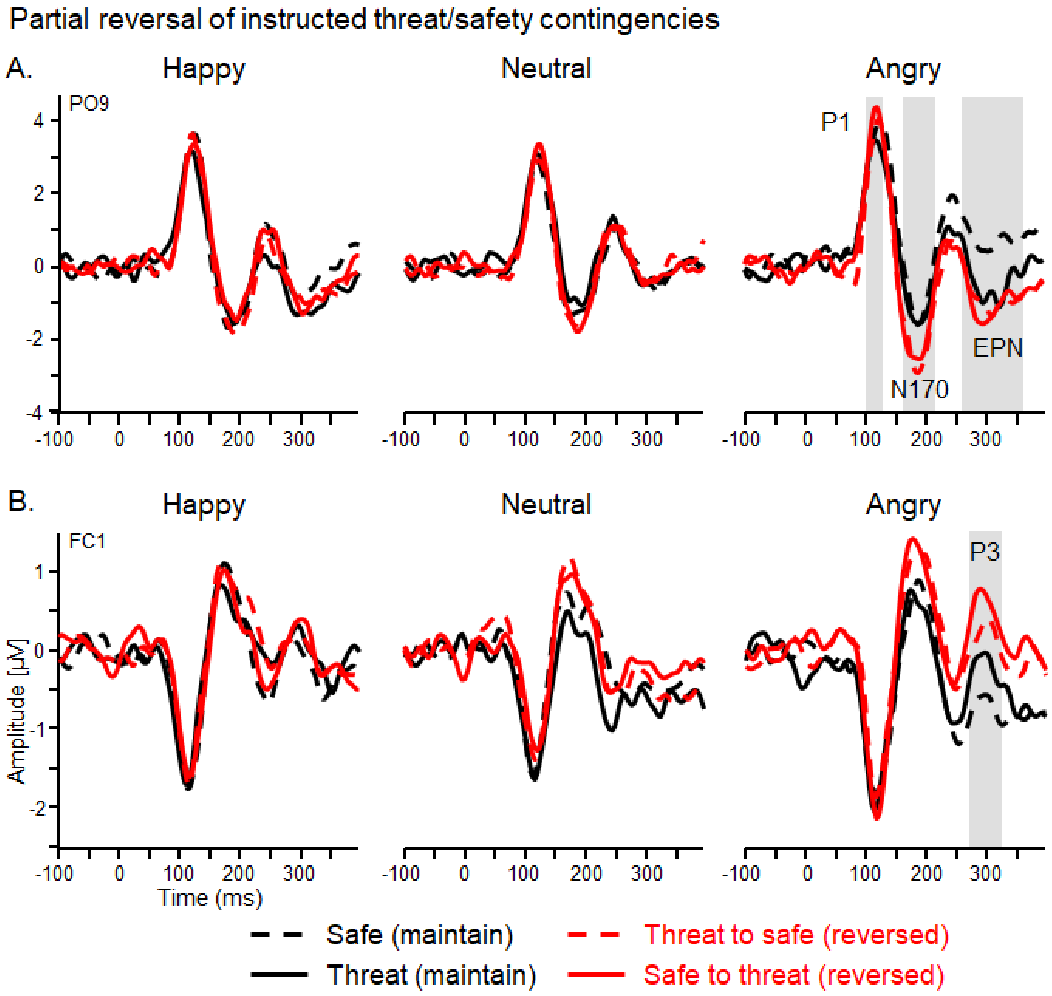


Figure 5. Event-related brain potential waveforms as a function of facial expression and reversal instructions. (A) Instructed threat/safety reversal modulates the N170 and EPN amplitudes specifically for angry threat identities over parieto-occipital sensor sites. (B) Partial reversal instructions modulate fronto-central P3 amplitudes specifically for angry faces. Grey-shaded area markers highlight the time windows, which were used for statistical calculations.

4. Discussion

Learning that a particular person might be dangerous changes that person's perception within fractions of a second. This is particularly true when the allegedly threatening person is looking angry or even smiles. To examine visual attention processes involved in this effect, the present study combined the well-established picture viewing and threat-of-shock paradigms. Key results from both manipulations were confirmed. Differential ERP waveforms emerged for emotional (happy and/or angry) compared to neutral facial expressions. This pattern of selective emotion processing was observable for several ERP components throughout the duration of picture presentation (i.e., control block: P1, N170, EPN, P3, and LPP). In addition, verbal information about threat-of-shock or safety effectively changed face perception (i.e., instantiation block: N170, EPN, and LPP). Differential processing of instructed threat compared to safe identities was indicated by EPN and LPP components. Importantly, such threat-related biases in face and person perception varied as a joint function of visual (face identity, facial expression) and verbally instructed information. Specifically, the N170 component was pronounced for threat identities smiling at the observer and suggest shared neural mechanisms involved in identity and emotion processing. Moreover, verbal instructions readily reversed previously learned threat/safety-associations (i.e., reversal block: N170, EPN, and P3), and this reversal learning was most evident for implicit and explicit threatening facial information (i.e., the angry ‘aggressor’). Thus, early face encoding (identity and facial expression) selectively varies according to the mere verbal instruction about whether a person is potentially dangerous or safe. This attentional selection process flexibly readjusts when new information is learned (i.e. reversal instructions).

Recent research suggested prioritized processing of emotional over neutral facial expressions. Reflecting the mere impact of facial expressions without concurrent threat-of-shock, the present passive viewing control condition (Block 1) provides clear support for this notion. As early as one hundred milliseconds after picture onset, the P1 component differentiated angry from neutral face processing. While literature on P1 emotion effects is rather inconsistent (Schindler & Bublatzky, under revision), this component has been suggested as an indicator of enhanced early vigilance and spatial attention towards biologically relevant stimuli such as threatening faces, especially in anxious participants (Brosch et al., 2008; Bublatzky et al., 2014b; Pourtois et al., 2004; Wieser & Keil, 2020). A similar pattern emerged for the N170, which presumably reflects structural face encoding within the temporal cortex (Eimer & Holmes, 2007; Itier & Taylor, 2004). Here, the present findings are in line with several studies showing that the N170 is also sensitive to facial emotions (Hinojosa et al., 2015) and biases in social perception (e.g., own-race/age biases, stereotyping, or social exclusion; Ofan et al., 2011; Bublatzky, Pittig et al., 2017; Wiese et al., 2008). Furthermore, as an indicator of motivated attention to affective picture materials (Schupp et al., 2003), the EPN revealed differential processing for both happy and angry expressions compared to neutral faces. Similar findings were observed for fronto-central positivities (P3) and late positive potentials suggesting in-depth evaluation of affective information (Schupp et al., 2004).

When threat-of-shock was associated with face identity information (Block 2), the EPN and LPP components differentiated the processing of instructed threat from safe identities. This was observed during both the instantiation and reversal blocks, and extends previous research showing persistent threat-selective processing patterns using affective scenes (Bublatzky & Schupp, 2012) to the domain of face and person perception. Such attentional threat biases provide the perceptual base for organizing adaptive behaviors towards signals of threat and harm. For instance, verbally instructed threat linked to facial stimuli has been shown effective to provoke enhanced activity of the somatic and autonomic nervous system (e.g., potentiated startle reflex and enhanced skin conductance responses; Grillon & Charney, 2011; Bublatzky et al., 2018, 2019). This physiological priming of defensive response systems is further associated with overt avoidance behaviors and anxious decisions for costly but safe behavioral choices (Pittig et al., 2018; Bublatzky et al., 2017).

Importantly, the N170 component varied as a joint function of the verbally instructed meaning and the facial expression of a person. Relative to safe identities, pronounced N170 amplitudes emerged for threat identities smiling at the observer. Here, the smiling threat-identity might have boosted attention because of the affective incongruence and conflict between the visual and instructed meaning (Hajcak & Foti, 2008; Rothermund, 2003). Congruency effects have been observed previously as a function of emotional facial expression and their contextual settings (for a review see Wieser & Brosch, 2012), such as affective background scenes (Righart and de Gelder, 2006, 2008) and temporally preceding information (Diéguez-Risco et al., 2015; Hietanen and Astikainen, 2013). Moreover, much research from the cognitive domain observed enhanced perceptual processing as a function of context-incongruent or deviant information (e.g. Näätänen et al., 2007; Kutas & Federmeier, 2000, Woldorff et al., 1998). Such mismatch detection mechanism seem also to be involved in affective violation of expectations as in the present study. For instance, viewing pleasant scenes were associated with selective attention processes when undergoing aversive apprehensions (Bublatzky et al., 2010), and triggered defensive responding when cueing threat-of-shock (Bradley et al., 2005). In a recent MEG study (Bublatzky et al., 2020), we found a similar pattern of early affective incongruence processing at parietal cortex (63-127ms), ventrolateral PFC and temporal pole regions (103–157 ms) when viewing fearful faces during safety and happy expressions during threat.

In the present study, however, viewing angry expressions of safe identities were not associated with enhanced N170, thus pointing to the involvement of further attentional and/or memory-related processes. For instance, smiling while threatening someone might gain even more aversive qualities as it indicates a particular mean or dangerous opponent (Gerdes et al., 2012). Whereas previous research showed the N170 component relates to the processing of face identity (e.g., Schweinberger & Neumann, 2016), and emotional facial expression (e.g., Hinojosa et al., 2015), the present interaction of both sources of information suggest at least shared neural mechanisms involved in identity and emotion processing (Young & Bruce, 2011). Intriguingly, the mere social communication about threatening events (which actually never occurred) triggered threat-related attentional biases towards allegedly threatening persons, with implications for social interactions and behavior (e.g., impression formation, social bonding; Fiske & Neuberg, 1990; Golkar & Olsson, 2017).

Given the relevance of accurate threat perception for social behavior, tracking the link between a particular person and potential harmful consequences is crucial to realign interpersonal relations (Kringelbach & Rolls, 2003). In the present study, explicit reversal instructions turned the instructed shock threat from one person to another. Conceptually, this reversal learning process implicates both the inhibition of old (from threat to safe) and the acquisition of new threat-associations (from safe to threat; Schiller & Delgado, 2010). As in previous research, our rating data demonstrate that instructions readily instantiated and reversed threat/safety-associations. On the other hand, reversal instructions have been shown to effectively modulate the activity of the autonomic and somatic nervous system. For instance, threat-enhanced skin conductance responses and startle reflex activity indicated reversed fear responses (Atlas & Phelps, 2018; Bublatzky et al, 2018, 2019; Costa et al., 2015; Mertens et al., 2018). Extending these findings, the present study revealed reversal-related brain activity specifically to newly learned threat and safe identities over visual processing areas at around 200 ms after the onset of a face picture (N170 and EPN). This presumably reflects the early attentional tagging of motivationally salient information (Kissler et al., 2007; Schupp et al., 2003).

Importantly, on the perceptual-attentional level, reversal learning interacted with the facial expression of the threatening person. Specifically, viewing the new threat identity (safe reversed to threat) led to pronounced N170 and EPN when they displayed angry, but not happy emotions. Mirroring these early posterior reversal effects, the fronto-central P3 showed increased amplitudes to identities who were newly learned compared to those who maintained their meaning. Thus, reversing the knowledge about an angry looking person was associated with significantly larger P3 amplitudes compared to faces that maintained their threatening or safe value. These findings add to previous research that reported modulation of P3 amplitudes when participants were cued to switch associations between face identity and facial expressions (Willis et al., 2010). Interestingly, these reversal effects were particularly marked for angry compared to happy facial expressions, and were associated with enhanced BOLD-responses in orbito-frontal and ACC cortex (Kringelbach & Rolls, 2003). Taken together, these findings relate to memory update processes, which presumably originate from distributed network activity associated with attention and subsequent memory processing (Kok, 2001; Polich, 2007). Here, the present N170-EPN-P3 findings might reflect similar attention-memory update processes in face and person perception. That such processes vary as a function of social learning about allegedly threatening or safe persons or situations provide a new and promising avenue to investigate the neural underpinnings of acquisition, reversal, and extinction of social fears and anxieties (Debiec & Olsson, 2017).

Several aspects of the present study need to be noted and may be considered for future research. First, while we focused on the electrocortical correlates of face identity and facial expression processing as a function of instructed and reversed threat/safety associations, rating data were obtained only for the combined identity/expression stimuli (i.e. identities were presented with all expressions at once). Thus, no interaction effects could be tested for self-reported threat, valence, and arousal. Moreover, questionnaire data revealed multiple significant covariation effects among anxiety scores and ERP components. However, as we had no a priori hypotheses regarding our healthy student sample, these effects did not survive correction for multiple comparisons. Moreover, the sample consisted of mostly female participants, which furthermore reduces generalizability of the results. Future research may clarify the role of interindividual differences in threat and safety learning as a function of potential vulnerability factors (e.g. sex, intolerance of uncertainty).

From a clinical perspective, the inclusion of selected participants with deficits in social interaction (e.g. social anxiety disorder) or interpersonal trauma experiences (Schellhaas et al., in revision) would be very interesting to track the persistence of threat-related attentional biases (e.g. P1 modulations as an index of anxious hypervigilance; Wieser & Keil, 2020). Here, the threat-of-shock paradigm provides a laboratory analog for testing persisting fears and anxieties that can be very resistant against extinction learning, even in the complete absence of the anticipated aversive events (Bublatzky et al., 2012, 2013, 2014). Moreover, threat instructions have been shown very effective even after reversal instructions, future research may follow up on multiple reversals with social stimuli and/or situations (e.g., Atlas, 2019; Bublatzky & Schupp, 2012). A particular focus on the behavioral consequences would be helpful to describe the social functions of reversal learning (e.g., decision-making or trust behavior; Kringelbach & Rolls, 2003; Paret & Bublatzky, 2020; Willis et al., 2010).

In summary, this study shows that social learning through hearsay is very effective in directing selective attention to a specific person. The mere verbal statement that an individual indicates shock threat led to prioritized processing of this allegedly threatening person (relative to safe identities; EPN and LPP). Intriguingly, viewing smiling threat identities (but not angry safe identities) revealed particularly pronounced N170 effects, reflecting enhanced incongruence and/or emotional amplification of perceptual processing. Moreover, verbal instructions readily reversed previously acquired threat/safety-associations. This reversal learning specifically modulated the processing of combined implicit and explicit threatening facial information (i.e., threat-persons displaying angry facial expressions; N170, EPN, and P3). As shock threat was not substantiated by any aversive experiences (i.e., no shocks during the experiment), these findings demonstrate the impact of mere anticipatory processes on perceptual biases which are relevant to social stereotyping and/or maladaptive extinction learning in anxiety disorders.

Author contributions

All authors designed the study, F.B. and G.A. supervised the conduct of the study, F.B. analyzed data and drafted the manuscript. All authors revised the manuscript.

Acknowledgements

We are grateful to S. Schellhaas, A-L. Tebbe, and M. Berckemann for their assistance in data collection. This research was supported by the German Research Foundation (DFG) grant to F. Bublatzky (BU 3255/1-1 and -2).

References

1. Amodio, D. M. (2014). The neuroscience of prejudice and stereotyping. *Nature Reviews Neuroscience, 15*(10), 670. https://doi.org/10.1038/nrn3800. Epub 2014 Sep 4.
2. Aguado, L., Valdés-Conroy, B., Rodríguez, S., Román, F. J., Diéguez-Risco, T., & Fernández-Cahill, M. (2012). Modulation of early perceptual processing by emotional expression and acquired valence of faces. *Journal of Psychophysiology. 26*, 29-41. DOI: 10.1027/0269-8803/a000065
3. Atlas, L. Y. (2019). How instructions shape aversive learning: higher order knowledge, reversal learning, and the role of the amygdala. *Current Opinion in Behavioral Sciences, 26*, 121-129. https://doi.org/10.1016/j.cobeha.2018.12.008
4. Atlas, L. Y., & Phelps, E. A. (2018). Prepared stimuli enhance aversive learning without weakening the impact of verbal instructions. *Learning & Memory, 25*(2), 100-104. https://doi.org/10.1101/lm.046359.117
5. Askew, C., & Field, A. P. (2008). The vicarious learning pathway to fear 40 years on. *Clinical psychology review, 28*(7), 1249-1265. https://doi.org/10.1016/j.cpr.2008.05.003
6. Baas, J. M., Milstein, J., Donlevy, M., & Grillon, C. (2006). Brainstem correlates of defensive states in humans. *Biological Psychiatry, 59*(7), 588-593. https://doi.org/10.1016/j.biopsych.2005.09.009
7. Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience, 8*(6), 551-565. https://doi.org/10.1162/jocn.1996.8.6.551
8. Blechert, J., Sheppes, G., Di Tella, C., Williams, H., & Gross, J. J. (2012). See what you think: Reappraisal modulates behavioral and neural responses to social stimuli. *Psychological Science, 23*(4), 346-353. https://doi.org/10.1177/0956797612438559

9. Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry, 25*(1), 49-59. https://doi.org/10.1016/0005-7916(94)90063-9

10. Bradley, M. M., Moulder, B., & Lang, P. J. (2005). When good things go bad: The reflex physiology of defense. *Psychological Science, 16*(6), 468-473. https://doi.org/10.1111/j.0956-7976.2005.01558.x

11. Brosch, T., Sander, D., Pourtois, G., & Scherer, K. R. (2008). Beyond fear: Rapid spatial orienting toward positive emotional stimuli. *Psychological Science, 19*(4), 362-370. https://doi.org/10.1111/j.1467-9280.2008.02094.x

12. Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology, 77*, 305–327.

13. Bublatzky, F., & Alpers, G. W. (2017). Facing two faces: Defense activation varies as a function of personal relevance. *Biological Psychology, 125*, 64-69. https://doi.org/10.1016/j.biopsycho.2017.03.001

14. Bublatzky, F., Alpers, G. W., & Pittig, A. (2017). From avoidance to approach: The influence of threat-of-shock on reward-based decision making. *Behaviour Research and Therapy, 96*, 47-56. https://doi.org/10.1016/j.brat.2017.01.003

15. Bublatzky, F., Flaisch, T., Stockburger, J., Schmälzle, R., & Schupp, H. T. (2010). The interaction of anticipatory anxiety and emotional picture processing: An event-related brain potential study. *Psychophysiology*, *47*(4), 687-696. https://doi.org/10.1111/j.1469-8986.2010.00966.x

16. Bublatzky, F., Gerdes, A., & Alpers, G. W. (2014a). The persistence of socially instructed threat: Two threat-of-shock studies. *Psychophysiology*, *51*(10), 1005-1014. https://doi.org/10.1111/psyp.12251

17. Bublatzky, F., Gerdes, A. B. M., White, A. J., Riemer, M., & Alpers, G. W. (2014b). Social and emotional relevance in face processing: Happy faces of future interaction partners

enhance the late positive potential. *Frontiers in Human Neuroscience, 8*, 493. https://doi.org/10.3389/fnhum.2014.00493

18. Bublatzky, F., Guerra, P., & Alpers, G. W. (2018). Verbal instructions override the meaning of facial expressions. *Scientific Reports, 8*(1), 1-11. https://doi.org/10.1038/s41598-018-33269-2
19. Bublatzky, F., Guerra, P. M., Pastor, M. C., Schupp, H. T., & Vila, J. (2013). Additive effects of threat-of-shock and picture valence on startle reflex modulation. *PloS one, 8*(1). https://doi.org/10.1371/journal.pone.0054003
20. Bublatzky, F., Kavcıoğlu, F., Guerra, P., Doll, S., & Junghöfer, M. (2020). Contextual information resolves uncertainty about ambiguous facial emotions: behavioral and magnetoencephalographic correlates. *NeuroImage*, 116814.
21. Bublatzky, F., Pittig, A., Schupp, H. T., & Alpers, G. W. (2017). Face-to-face: Visual attention to emotional facial expressions depend on face orientation. *Social Cognitive and Affective Neuroscience, 12*(5), 811-822. https://doi.org/10.1093/scan/nsx001
22. Bublatzky, F., Riemer, M., & Guerra, P. (2019). Reversing threat to safety: incongruence of facial emotions and instructed threat modulates conscious perception but not physiological responding. *Frontiers in Psychology, 10*, 2091. https://doi.org/10.3389/fpsyg.2019.02091
23. Bublatzky, F., & Schupp, H. T. (2012). Pictures cueing threat: brain dynamics in viewing explicitly instructed danger cues. *Social Cognitive and Affective Neuroscience*, *7*, 611-622. https://dx.doi.org/10.1093/scan/nsr032
24. Costa, V. D., Bradley, M. M., & Lang, P. J. (2015). From threat to safety: Instructed reversal of defensive reactions. *Psychophysiology, 52*(3), 325-332. https://doi.org/10.1111/psyp.12359

25. Debiec, J., & Olsson, A. (2017). Social fear learning: from animal models to human function. *Trends in Cognitive Sciences, 21*(7), 546-555. https://doi.org/10.1016/j.tics.2017.04.010

26. Diéguez-Risco, T., Aguado, L., Albert, J., & Hinojosa, J. A. (2015). Judging emotional congruency: Explicit attention to situational context modulates processing of facial expressions of emotion. *Biological Psychology, 112*, 27-38. https://doi.org/10.1016/j.biopsycho.2015.09.012

27. Eimer, M., & Holmes, A. (2007). Event-related brain potential correlates of emotional face processing. *Neuropsychologia, 45*(1), 15-31.

28. Etkin, A., & Wager, T. D. (2007). Functional neuroimaging of anxiety: a meta-analysis of emotional processing in PTSD, social anxiety disorder, and specific phobia. *American Journal of Psychiatry, 164*(10), 1476-1488. https://doi.org/10.1176/appi.ajp.2007.07030504

29. 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. https://doi.org/10.3758/BF03193146

30. Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In Advances in experimental social psychology (Vol. 23, pp. 1–74). Academic Press. https://doi.org/10.1016/S0065-2601(08)60317-2.

31. Flaisch, T., Junghöfer, M., Bradley, M. M., Schupp, H. T., & Lang, P. J. (2008a). Rapid picture processing: affective primes and targets. *Psychophysiology, 45*(1), 1-10. https://doi.org/10.1111/j.1469-8986.2007.00600.x

32. Flaisch, T., Stockburger, J., & Schupp, H. T. (2008b). Affective prime and target picture processing: an ERP analysis of early and late interference effects. *Brain Topography, 20*(4), 183-191.

33. Gerdes, A. B., Wieser, M. J., Alpers, G. W., Strack, F., & Pauli, P. (2012). Why do you smile at me while I'm in pain?—Pain selectively modulates voluntary facial muscle responses to happy faces. *International Journal of Psychophysiology, 85*(2), 161-167. https://doi.org/10.1016/j.ijpsycho.2012.06.002

34. Golkar, A. & Olsson, A. (2017). The interplay of social group biases in social threat learning. *Scientific Reports 7*(1), 7685, https://doi.org/10.1038/s41598-017-07522

35. Grillon, C., Ameli, R., Woods, S. W., Merikangas, K., & Davis, M. (1991). Fear-potentiated startle in humans: Effects of anticipatory anxiety on the acoustic blink reflex. *Psychophysiology, 28*(5), 588-595. https://doi.org/10.1111/j.1469-8986.1991.tb01999.x

36. Grillon, C., & Charney, D. R. (2011). In the face of fear: anxiety sensitizes defensive responses to fearful faces. *Psychophysiology*, *48*(12), 1745-1752. . https://doi.org/10.1111/j.1469-8986.2011.01268.x

37. Hajcak, G., & Foti, D. (2008). Errors are aversive: Defensive motivation and the error-related negativity. *Psychological Science, 19*(2), 103-108. https://doi.org/10.1111/j.1467-9280.2008.02053.x

38. Haxby, J. V., & Gobbini, M. I. (2011). Distributed neural systems for face perception (pp. 93-110). The Oxford Handbook of Face Perception.

39. Hietanen, J. K., & Astikainen, P. (2013). N170 response to facial expressions is modulated by the affective congruency between the emotional expression and preceding affective picture. Biological Psychology, 92(2), 114-124. https://doi.org/10.1016/j.biopsycho.2012.10.005

40. Hinojosa, J. A., Mercado, F., & Carretié, L. (2015). N170 sensitivity to facial expression: A meta-analysis. *Neuroscience & Biobehavioral Reviews, 55*, 498-509. https://doi.org/10.1016/j.neubiorev.2015.06.002

41. Itier, R. J., & Taylor, M. J. (2004). N170 or N1? Spatiotemporal differences between object and face processing using ERPs. *Cerebral Cortex, 14*(2), 132-142. https://doi.org/10.1093/cercor/bhg111

42. Kaufmann, J. M., Schweinberger, S. R., & Burton, A. M. (2009). N250 ERP correlates of the acquisition of face representations across different images. *Journal of Cognitive Neuroscience, 21*(4), 625-641. https://doi.org/10.1162/jocn.2009.21080

43. Kim, H., Somerville, L. H., Johnstone, T., Polis, S., Alexander, A. L., Shin, L. M., & Whalen, P. J. (2004). Contextual modulation of amygdala responsivity to surprised faces. *Journal of Cognitive Neuroscience, 16*(10), 1730-1745. https://doi.org/10.1162/0898929042947865

44. Kissler, J., Herbert, C., Peyk, P., & Junghofer, M. (2007). Buzzwords: early cortical responses to emotional words during reading. *Psychological Science, 18*(6), 475-480. https://doi.org/10.1111/j.1467-9280.2007.01924.x

45. Klein, F., Iffland, B., Schindler, S., Wabnitz, P., and Neuner, F. (2015). This person is saying bad things about you: the influence of physically and socially threatening context information on the processing of inherently neutral faces. *Cognitive Affectective Behavioral Neuroscience 15*, 736–748. doi: 10.3758/s13415-015-0361-8

46. Koban, L., Jepma, M., Geuter, S., & Wager, T. D. (2017). What's in a word? How instructions, suggestions, and social information change pain and emotion. *Neuroscience & Biobehavioral Reviews, 81*, 29-42. https://doi.org/10.1016/j.neubiorev.2017.02.014

47. Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology, 38*(3), 557-577. https://doi.org/10.1017/S0048577201990559

48. Kringelbach, M. L., & Rolls, E. T. (2003). Neural correlates of rapid reversal learning in a simple model of human social interaction. *NeuroImage, 20*(2), 1371-1383. https://doi.org/10.1016/S1053-8119(03)00393-8

49. Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*(12), 463-470. https://doi.org/10.1016/S1364-6613(00)01560-6

50. Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska directed emotional faces (KDEF). CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, (1998).

51. McCrackin, S. D., & Itier, R. J. (2018). Is it about me? Time-course of self-relevance and valence effects on the perception of neutral faces with direct and averted gaze. *Biological psychology, 135*, 47-64. https://doi.org/10.1016/j.biopsycho.2018.03.003

52. Mechias, M. L., Etkin, A., & Kalisch, R. (2010). A meta-analysis of instructed fear studies: implications for conscious appraisal of threat. *NeuroImage, 49*(2), 1760-1768. https://doi.org/10.1016/j.neuroimage.2009.09.040

53. Meeren, H. K., van Heijnsbergen, C. C., & de Gelder, B. (2005). Rapid perceptual integration of facial expression and emotional body language. *Proceedings of the National Academy of Sciences, 102*(45), 16518-16523. https://doi.org/10.1073/pnas.0507650102

54. Mertens, G., Boddez, Y., Sevenster, D., Engelhard, I. M., & De Houwer, J. (2018). A review on the effects of verbal instructions in human fear conditioning: Empirical findings, theoretical considerations, and future directions. *Biological Psychology, 137*, 49-64. https://doi.org/10.1016/j.biopsycho.2018.07.002

55. Mertens, G., & De Houwer, J. (2016). Potentiation of the startle reflex is in line with contingency reversal instructions rather than the conditioning history. *Biological Psychology, 113*, 91-99. https://doi.org/10.1016/j.biopsycho.2015.11.014

56. Milad, M. R., & Quirk, G. J. (2012). Fear extinction as a model for translational neuroscience: ten years of progress. *Annual Review of Psychology, 63*, 129-151. https://doi.org/10.1146/annurev.psych.121208.131631

57. Muench, H. M., Westermann, S., Pizzagalli, D. A., Hofmann, S. G., & Mueller, E. M. (2016). Self-relevant threat contexts enhance early processing of fear-conditioned faces. *Biological Psychology, 121*, 194-202. https://doi.org/10.1016/j.biopsycho.2016.07.017

58. Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clinical Neurophysiology, 118*(12), 2544-2590. https://doi.org/10.1016/j.clinph.2007.04.026

59. Ofan, R. H., Rubin, N., & Amodio, D. M. (2011). Seeing race: N170 responses to race and their relation to automatic racial attitudes and controlled processing. *Journal of Cognitive Neuroscience, 23*(10), 3153-3161. https://doi.org/10.1162/jocn_a_00014

60. Olsson, A., & Phelps, E. A. (2007). Social learning of fear. *Nature Neuroscience, 10*(9), 1095. https://doi.org/10.1038/nn1968

61. Paret, C., & Bublatzky, F. (2020). Threat rapidly disrupts reward reversal learning. *Behaviour Research and Therapy*, 103636. https://doi.org/10.1016/j.brat.2020.103636

62. Pittig, A., Treanor, M., LeBeau, R. T., & Craske, M. G. (2018). The role of associative fear and avoidance learning in anxiety disorders: gaps and directions for future research. *Neuroscience & Biobehavioral Reviews*. https://doi.org/10.1016/j.neubiorev.2018.03.015

63. Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology, 118*(10), 2128-2148. https://doi.org/10.1016/j.clinph.2007.04.019

64. Pourtois, G., Grandjean, D., Sander, D., & Vuilleumier, P. (2004). Electrophysiological correlates of rapid spatial orienting towards fearful faces. *Cerebral Cortex, 14*(6), 619-633. https://doi.org/10.1093/cercor/bhh023

65. Righart, R., & De Gelder, B. (2006). Context influences early perceptual analysis of faces—an electrophysiological study. *Cerebral Cortex, 16*(9), 1249-1257. https://doi.org/10.1093/cercor/bhj066

66. Righart, R., & De Gelder, B. (2008). Rapid influence of emotional scenes on encoding of facial expressions: an ERP study. *Social Cognitive and Affective Neuroscience, 3*(3), 270-278. https://doi.org/10.1093/scan/nsn021

67. Robinson, O. J., Vytal, K., Cornwell, B. R., & Grillon, C. (2013). The impact of anxiety upon cognition: perspectives from human threat of shock studies. *Frontiers in Human Neuroscience, 7*, 203. https://doi.org/10.3389/fnhum.2013.00203

68. Rothermund, K. (2003). Motivation and attention: Incongruent effects of feedback on the processing of valence. *Emotion, 3*(3), 223. http://dx.doi.org/10.1037/1528-3542.3.3.223

69. Schellhaas, S., Arnold, N., Schmahl, C., & Bublatzky, F. (in revision). Contextual source information modulates neural face processing in the absence of conscious recognition: A threat-of-shock study.

70. Schiller, D., & Delgado, M. R. (2010). Overlapping neural systems mediating extinction, reversal and regulation of fear. *Trends in Cognitive Sciences, 14*(6), 268-276. https://doi.org/10.1016/j.tics.2010.04.002

71. Schindler, S., & Bublatzky, F. (under review). Attention and emotion: An integrative review of emotional face processing as a function of attention.

72. Schupp, H. T., Flaisch, T., Stockburger, J., & Junghöfer, M. (2006). Emotion and attention: event-related brain potential studies. *Progress in Brain Research, 156*, 31-51. https://doi.org/10.1016/S0079-6123(06)56002-9

73. Schupp, H. T., Junghöfer, M., Weike, A. I., & Hamm, A. O. (2003). Emotional facilitation of sensory processing in the visual cortex. *Psychological Science, 14*(1), 7-13. https://doi.org/10.1111/1467-9280.01411

74. Schupp, H. T., Öhman, A., Junghöfer, M., Weike, A. I., Stockburger, J., & Hamm, A. O. (2004). The facilitated processing of threatening faces: an ERP analysis. *Emotion, 4*(2), 189. https://doi.org/10.1037/1528-3542.4.2.189

75. Schweinberger, S. R., & Neumann, M. F. (2016). Repetition effects in human ERPs to faces. *Cortex, 80*, 141-153. https://doi.org/10.1016/j.cortex.2015.11.001

76. Wiese, H., Schweinberger, S. R., & Neumann, M. F. (2008). Perceiving age and gender in unfamiliar faces: Brain potential evidence for implicit and explicit person categorization. *Psychophysiology, 45*(6), 957-969. https://doi.org/10.1111/j.1469-8986.2008.00707.x

77. Wieser, M. J., & Brosch, T. (2012). Faces in context: a review and systematization of contextual influences on affective face processing. *Frontiers in Psychology, 3*, 471. https://doi.org/10.3389/fpsyg.2012.00471

78. Wieser, M. J., Gerdes, A. B., Büngel, I., Schwarz, K. A., Mühlberger, A., & Pauli, P. (2014). Not so harmless anymore: How context impacts the perception and electrocortical processing of neutral faces. *NeuroImage, 92*, 74-82. https://doi.org/10.1016/ j.neuroimage.2014.01.022

79. Wieser, M. J., & Keil, A. (2020). Attentional threat biases and their role in anxiety: A neurophysiological perspective. *International Journal of Psychophysiology*, 153, 148-158. https://doi.org/10.1016/j.ijpsycho.2020.05.004

80. Wieser, M. J., Pauli, P., Reicherts, P., & Mühlberger, A. (2010). Don't look at me in anger! Enhanced processing of angry faces in anticipation of public speaking. *Psychophysiology, 47*(2), 271-280. https://doi.org/10.1111/j.1469-8986.2009.00938.x

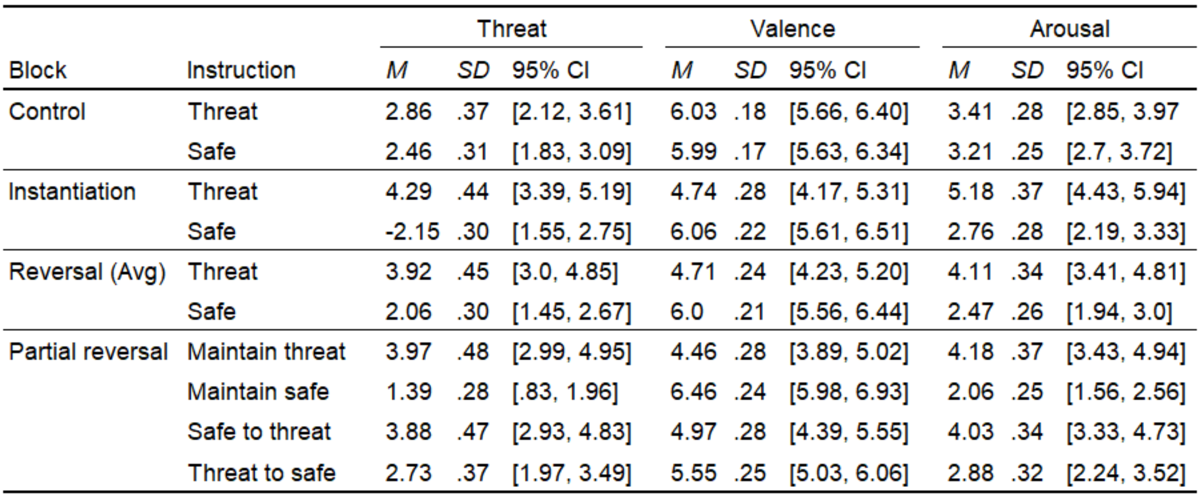
81. Willis, M. L., Palermo, R., Burke, D., Atkinson, C. M., & McArthur, G. (2010). Switching associations between facial identity and emotional expression: A behavioural and ERP study. *NeuroImage, 50*(1), 329-339. https://doi.org/10.1016/j.neuroimage.2009.11.071

82. Woldorff, M. G., Hillyard, S. A., Gallen, C. C., Hampson, S. R., & Bloom, F. E. (1998). Magnetoencephalographic recordings demonstrate attentional modulation of mismatch-related neural activity in human auditory cortex. *Psychophysiology, 35*(3), 283-292.

83. Young, A.W., & Bruce, V. (2011). Understanding person perception. *British Journal of Psychology, 102*, 959-974. https://doi.org/10.1111/j.2044-8295.2011.02045.x

84. Young, A. W., Ellis, H. D., Szulecka, T. K., & De Pauw, K. W. (1990). Face processing impairments and delusional misidentification. *Behavioural Neurology, 3*(3), 153-168. https://doi.org/10.3233/BEN-1990-3303

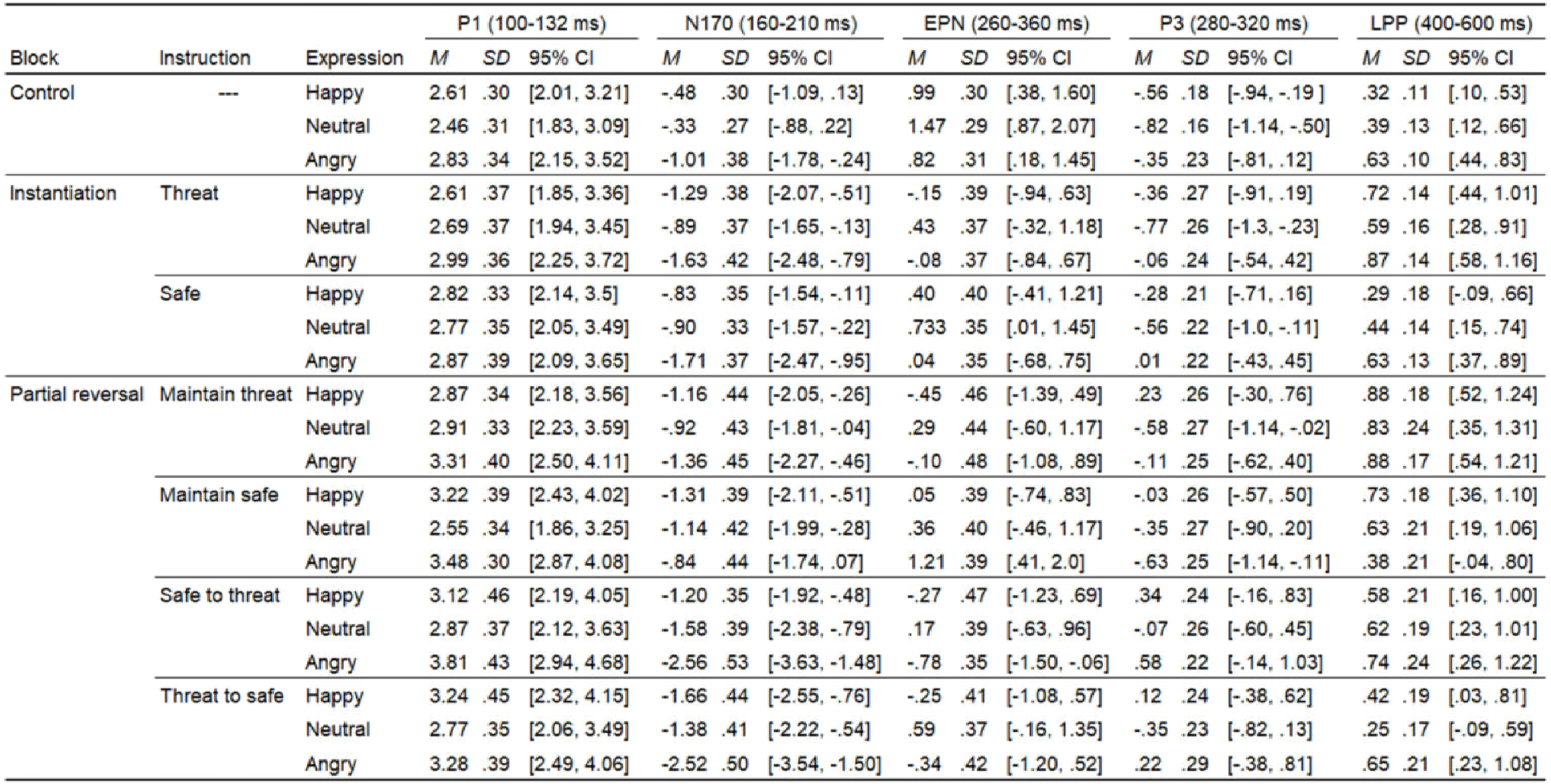
Table 1



		Threat			Valence			Arousal		
Block	Instruction	*M*	*SD*	95% CI	*M*	*SD*	95% CI	*M*	*SD*	95% CI
Control	Threat	2.86	.37	[2.12, 3.61]	6.03	.18	[5.66, 6.40]	3.41	.28	[2.85, 3.97
	Safe	2.46	.31	[1.83, 3.09]	5.99	.17	[5.63, 6.34]	3.21	.25	[2.7, 3.72]
Instantiation	Threat	4.29	.44	[3.39, 5.19]	4.74	.28	[4.17, 5.31]	5.18	.37	[4.43, 5.94]
	Safe	-2.15	.30	[1.55, 2.75]	6.06	.22	[5.61, 6.51]	2.76	.28	[2.19, 3.33]
Reversal (Avg)	Threat	3.92	.45	[3.0, 4.85]	4.71	.24	[4.23, 5.20]	4.11	.34	[3.41, 4.81]
	Safe	2.06	.30	[1.45, 2.67]	6.0	.21	[5.56, 6.44]	2.47	.26	[1.94, 3.0]
Partial reversal	Maintain threat	3.97	.48	[2.99, 4.95]	4.46	.28	[3.89, 5.02]	4.18	.37	[3.43, 4.94]
	Maintain safe	1.39	.28	[.83, 1.96]	6.46	.24	[5.98, 6.93]	2.06	.25	[1.56, 2.56]
	Safe to threat	3.88	.47	[2.93, 4.83]	4.97	.28	[4.39, 5.55]	4.03	.34	[3.33, 4.73]
	Threat to safe	2.73	.37	[1.97, 3.49]	5.55	.25	[5.03, 6.06]	2.88	.32	[2.24, 3.52]

Table 1. Summary of mean, standard deviation and 95% confidence intervals for threat, valence, and arousal ratings as a function of the experimental Block (control vs. instantiation vs. partial reversal) and Instruction (threat vs. safety). For the reversal block, averaged ratings are provided for threat vs. safety (i.e., Reversal Avg) as well as a function of Instruction by Contingency (i.e., Partial reversal: maintain vs. reversed).

Table 2



			P1 (100-132 ms)			N170 (160-210 ms)			EPN (260-360 ms)			P3 (280-320 ms)			LPP (400-600 ms)		
Block	Instruction	Expression	*M*	*SD*	95% CI	*M*	*SD*	95% CI	*M*	*SD*	95% CI	*M*	*SD*	95% CI	*M*	*SD*	95% CI
Control	---	Happy	2.61	.30	[2.01, 3.21]	-.48	.30	[-1.09, .13]	.99	.30	[.38, 1.60]	-.56	.18	[-.94, -.19]	.32	.11	[.10, .53]
		Neutral	2.46	.31	[1.83, 3.09]	-.33	.27	[-.88, .22]	1.47	.29	[.87, 2.07]	-.82	.16	[-1.14, -.50]	.39	.13	[.12, .66]
		Angry	2.83	.34	[2.15, 3.52]	-1.01	.38	[-1.78, -.24]	.82	.31	[.18, 1.45]	-.35	.23	[-.81, .12]	.63	.10	[.44, .83]
Instantiation	Threat	Happy	2.61	.37	[1.85, 3.36]	-1.29	.38	[-2.07, -.51]	-.15	.39	[-.94, .63]	-.36	.27	[-.91, .19]	.72	.14	[.44, 1.01]
		Neutral	2.69	.37	[1.94, 3.45]	-.89	.37	[-1.65, -.13]	.43	.37	[-.32, 1.18]	-.77	.26	[-1.3, -.23]	.59	.16	[.28, .91]
		Angry	2.99	.36	[2.25, 3.72]	-1.63	.42	[-2.48, -.79]	-.08	.37	[-.84, .67]	-.06	.24	[-.54, .42]	.87	.14	[.58, 1.16]
	Safe	Happy	2.82	.33	[2.14, 3.5]	-.83	.35	[-1.54, -.11]	.40	.40	[-.41, 1.21]	-.28	.21	[-.71, .16]	.29	.18	[-.09, .66]
		Neutral	2.77	.35	[2.05, 3.49]	-.90	.33	[-1.57, -.22]	.733	.35	[.01, 1.45]	-.56	.22	[-1.0, -.11]	.44	.14	[.15, .74]
		Angry	2.87	.39	[2.09, 3.65]	-1.71	.37	[-2.47, -.95]	.04	.35	[-.68, .75]	.01	.22	[-.43, .45]	.63	.13	[.37, .89]
Partial reversal	Maintain threat	Happy	2.87	.34	[2.18, 3.56]	-1.16	.44	[-2.05, -.26]	-.45	.46	[-1.39, .49]	.23	.26	[-.30, .76]	.88	.18	[.52, 1.24]
		Neutral	2.91	.33	[2.23, 3.59]	-.92	.43	[-1.81, -.04]	.29	.44	[-.60, 1.17]	-.58	.27	[-1.14, -.02]	.83	.24	[.35, 1.31]
		Angry	3.31	.40	[2.50, 4.11]	-1.36	.45	[-2.27, -.46]	-.10	.48	[-1.08, .89]	-.11	.25	[-.62, .40]	.88	.17	[.54, 1.21]
	Maintain safe	Happy	3.22	.39	[2.43, 4.02]	-1.31	.39	[-2.11, -.51]	.05	.39	[-.74, .83]	-.03	.26	[-.57, .50]	.73	.18	[.36, 1.10]
		Neutral	2.55	.34	[1.86, 3.25]	-1.14	.42	[-1.99, -.28]	.36	.40	[-.46, 1.17]	-.35	.27	[-.90, .20]	.63	.21	[.19, 1.06]
		Angry	3.48	.30	[2.87, 4.08]	-.84	.44	[-1.74, .07]	1.21	.39	[.41, 2.0]	-.63	.25	[-1.14, -.11]	.38	.21	[-.04, .80]
	Safe to threat	Happy	3.12	.46	[2.19, 4.05]	-1.20	.35	[-1.92, -.48]	-.27	.47	[-1.23, .69]	.34	.24	[-.16, .83]	.58	.21	[.16, 1.00]
		Neutral	2.87	.37	[2.12, 3.63]	-1.58	.39	[-2.38, -.79]	.17	.39	[-.63, .96]	-.07	.26	[-.60, .45]	.62	.19	[.23, 1.01]
		Angry	3.81	.43	[2.94, 4.68]	-2.56	.53	[-3.63, -1.48]	-.78	.35	[-1.50, -.06]	.58	.22	[-.14, 1.03]	.74	.24	[.26, 1.22]
	Threat to safe	Happy	3.24	.45	[2.32, 4.15]	-1.66	.44	[-2.55, -.76]	-.25	.41	[-1.08, .57]	.12	.24	[-.38, .62]	.42	.19	[.03, .81]
		Neutral	2.77	.35	[2.06, 3.49]	-1.38	.41	[-2.22, -.54]	.59	.37	[-.16, 1.35]	-.35	.23	[-.82, .13]	.25	.17	[-.09, .59]
		Angry	3.28	.39	[2.49, 4.06]	-2.52	.50	[-3.54, -1.50]	-.34	.42	[-1.20, .52]	.22	.29	[-.38, .81]	.65	.21	[.23, 1.08]

Table 2 summarizes means, standard deviation and 95% confidence intervals for the separate ERP components ranging across the visual processing stream (P1, N170, EPN, P3, and LPP). Experimental blocks depict either no threat/safety instruction (control block), instructed threat

and safety (instantiation block), and partial reversal instructions (partial reversal block), leading to maintained threat/safety cues and reversed cues (i.e., from threat to safety or vice versa).

Figure captions

Figure 1. Schematic illustration of the experimental procedure. The first experimental block served as a control condition, in which participants were told to passively view all face pictures. Afterwards verbal instructions were given regarding which face identity (ID) is cueing threat or safety (instantiation block). To this end, two face identities were pointed out as cues for aversive shocks, whereas the other two identities served as instructed safety cues. In the partial reversal block, threat and safety associations were partially changed. Each one identity maintained cueing threat or safety, the associations of the other two identities were reversed (e.g., ID 1 and 3 maintain cueing threat and safety, but ID 2 now newly indicates safety and ID 3 newly cues threat). For each block, all four face identities were presented intermixed displaying happy, neutral and angry expression (360 trials) in a rapid serial picture stream. No shocks were applied throughout the experiment. With permission, face pictures are depicted from the KDEF (http://kdef.se/).

Figure 2. Self-reported threat, valence, and arousal as a function of experimental condition for the control, instantiation, and reversal block. Illustrated are averaged ratings (SEM) for face identities that were instructed as threat or safety cues, no separate facial expression ratings were obtained. For the control conditions (no threat instruction), means are based on an artificial data split. To illustrate the effects of partial reversal instructions, the significant interactions Instruction by Contingency (maintain vs. reversed) are displayed on the right side.

Figure 3. (A) Event-related brain potential waveforms as a function facial expression for the passive viewing control block (left and right sensors). Grey-shaded area markers highlight the time windows, which were used for statistical calculations. (B) Topographical difference plots (happy – neutral and angry – neutral) displaying emotion effects for the P1, N170, EPN, P3 and LPP components. Waveform differences are displayed on the backside or top of a model head.

Figure 4. Event-related brain potential waveforms as a function of Facial Expression (happy, neutral, angry) and Instruction (threat, safety) for the instantiation block. (A) Illustration of the N170 interaction effect and main effect Facial Expression for the N170, EPN and (B) for fronto-central P3, as well as the Instruction main effects for the EPN and LPP (C). Exemplary left sensors are depicted. Grey-shaded area markers highlight the time windows, which were used for statistical calculations.

Figure 5. Event-related brain potential waveforms as a function of facial expression and reversal instructions. (A) Instructed threat/safety reversal modulates the N170 and EPN amplitudes specifically for angry threat identities over parieto-occipital sensor sites. (B) Partial reversal instructions modulate fronto-central P3 amplitudes specifically for angry faces. Grey-shaded area markers highlight the time windows, which were used for statistical calculations.