Received: 28 October 2022 | Revised: 1 May 2023 | Accepted: 17 June 2023

DOI: 10.1111/psyp.14380

ORIGINAL ARTICLE

PSYCHOPHYSIOLOGY SPR WILEY

Gender congruence and emotion effects in cross-modal associative learning: Insights from ERPs and pupillary responses

Annika Ziereis | **Anne Schacht**

Department for Cognition, Emotion and Behavior, Affective Neuroscience and Psychophysiology Laboratory, Institute of Psychology, Georg-August-University of Göttingen, Göttingen, Germany

Correspondence
Annika Ziereis, Department for Cognition, Emotion and Behavior, Affective Neuroscience and Psychophysiology Laboratory, Institute of Psychology, Georg-August-University of Göttingen, Göttingen, Germany.
Email: annika.ziereis@uni-goettingen.de

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: 254142454 / GRK 2070 and 454648639/SFB1528; Leibniz-Gemeinschaft, Grant/Award Number: W45/2019

Abstract

Social and emotional cues from faces and voices are highly relevant and have been reliably demonstrated to attract attention involuntarily. However, there are mixed findings as to which degree associating emotional valence to faces occurs automatically. In the present study, we tested whether inherently neutral faces gain additional relevance by being conditioned with either positive, negative, or neutral vocal affect bursts. During learning, participants performed a gender-matching task on face-voice pairs without explicit emotion judgments of the voices. In the test session on a subsequent day, only the previously associated faces were presented and had to be categorized regarding gender. We analyzed event-related potentials (ERPs), pupil diameter, and response times (RTs) of $N=32$ subjects. Emotion effects were found in auditory ERPs and RTs during the learning session, suggesting that task-irrelevant emotion was automatically processed. However, ERPs time-locked to the conditioned faces were mainly modulated by the task-relevant information, that is, the gender congruence of the face and voice, but not by emotion. Importantly, these ERP and RT effects of learned congruence were not limited to learning but extended to the test session, that is, after removing the auditory stimuli. These findings indicate successful associative learning in our paradigm, but it did not extend to the task-irrelevant dimension of emotional relevance. Therefore, cross-modal associations of emotional relevance may not be completely automatic, even though the emotion was processed in the voice.

KEYWORDS
associative learning, conditioning, cross-modal, emotion, ERPs, faces, pupil size

1 | INTRODUCTION

In the most everyday encounter with another person, we automatically extract a variety of information from their face (Haxby et al., 2000). To effectively behave in a social situation, it is equally important to recognize the other's emotions and intentions and to consider the current context and previous experience with this person. It has been demonstrated that the affective context of the same and other modalities modulates face perception (Aviezer et al., 2011; Hassin et al., 2013; McCrackin & Itier, 2018; for a review, see Wieser & Brosch, 2012). However, there are open questions about the boundary conditions, under which context is integrated with and associated with faces, for example, how automatically humans gain knowledge about others and how this generalizes to other situations. To fill this research gap, we investigated whether the perception of novel and neutral faces changes when associated with task-irrelevant emotional context and whether these associations transfer to a different test setting. To this aim, we implemented a cross-modal associative learning paradigm and recorded emotion-sensitive event-related brain potentials (ERPs), pupil size changes, and behavioral measures.

Research on Pavlovian aversive conditioning has repeatedly shown that faces as conditioned stimuli (CS^+) can acquire negative valence when being paired with biologically aversive unconditioned stimuli (US), such as aversive odors (Steinberg et al., 2012), electric shocks (Rehbein et al., 2014), or loud noise bursts (Watters et al., 2018) as US (for a review: Miskovic & Keil, 2012). Neutral faces have also been shown to acquire positive valence, for example, when associated with monetary reward (Hammerschmidt et al., 2017; Hammerschmidt, Kagan, et al., 2018; Hammerschmidt, Kulke, et al., 2018). The CS^+ faces can then evoke physiological reactions (CR), such as changes in skin conductance, heart rate, pupil size, and enhanced neural processing, for example, in evoked steady-state potentials (ssVEP, e.g., Wieser, Miskovic, et al., 2014), neural oscillations (e.g., Chen et al., 2021), and ERPs: To better understand the mechanisms underlying associative learning, several studies compared ERP modulations of conditioned faces with typical effects of inherent emotional facial expressions, ranging from early sensory processing to higher cognitive evaluations.

The *P1* reflects early attention and typically peaks around 100 ms after stimulus onset with a bilateral occipital positivity generated by the extrastriate cortex (Hillyard & Anllo-Vento, 1998; Russo, 2003). It was found to be enhanced for emotional compared to neutral facial expressions (Bublatzky et al., 2014; Foti et al., 2010; Hammerschmidt et al., 2017; Rellecke et al., 2011), although other studies reported a lack of modulation of the P1 by emotional expressions (for a review, see Schindler & Bublatzky, 2020). The P1 is followed by the face-sensitive *N170* component, a negative deflection over occipito-temporal regions, which peaks around 170 ms and is typically enhanced for faces relative to other objects (Bentin et al., 1996; Rossion et al., 2000). Similar to P1 modulations, emotion effects on the N170 have been reported inconsistently, potentially due to the use of different stimuli across studies, and thus variations in low-level visual factors such as contrast (Bobak et al., 1987), size (Kornmeier et al., 2011; Yiannikas & Walsh, 1983), or luminance (Bieniek et al., 2013). The early posterior negativity (*EPN*), a relative negativity most pronounced around 200–300 ms over occipito-temporal regions, was reported robustly in several ERP studies for emotional versus neutral stimuli across several stimulus domains (e.g., Bayer & Schacht, 2014; Schacht & Sommer, 2009), and has been assumed to reflect the facilitation of sensory encoding and selective attention mechanisms (Schupp et al., 2006). Later ERP modulations like the *LPP/LPC* seem to be more strongly affected by specific task requirements (e.g., Rellecke et al., 2012a) but have reliably been demonstrated to be augmented by particularly facial expressions of aversive emotions (Schindler & Bublatzky, 2020; Schupp et al., 2004). Compared to studying facial expressions, where disentangling relevant and confounding low-level features can be challenging, studying faces associated with emotional information reduces the amount of confounding visual stimulus features by randomly assigning the CS–US pairing, thereby contrasting effects that are not intrinsic to the facial stimulus.

1.1 | ERP findings on faces with associated relevance

There is a long tradition of conditioning research, using faces as CS^+ and different types of (mostly aversive) US stimuli, which showed that face perception changed at different processing stages (for a review: Miskovic & Keil, 2012). Aside from the described ERP components, a few studies reported effects on early processing stages, observable already before 100 ms after the CS^+ onset (Morel et al., 2012; Mueller & Pizzagalli, 2015; Sperl et al., 2021; Steinberg et al., 2012; Steinberg, Bröckelmann, Dobel, et al., 2013; Steinberg, Bröckelmann, Rehbein, et al., 2013), and, remarkably, for different types of unconditioned stimuli (US), like odors, auditory startle, and electric shocks (Steinberg, Bröckelmann, Rehbein, et al., 2013). Typical ERP modulations for classical or evaluative conditioning and instrumental learning paradigms were reported at latencies from 100 ms on, during short- (P1, N170), mid- (EPN), and long- (LPC) latencies. Enhanced amplitudes for faces with associated relevance have been reported for the *P1* (monetary reward: Hammerschmidt et al., 2017; facial expressions of emotion: Aguado et al., 2012), *N170* (fear-conditioning: Camfield et al., 2016; Schellhaas et al., 2020; Sperl et al., 2021; aversive screams: Bruchmann et al., 2021; person knowledge: Luo et al., 2016; Schindler et al., 2021; facial expressions of emotion: Aguado et al., 2012), *EPN* (vocal emotional expressions: Ziereis & Schacht, 2023; fear-conditioning: Bruchmann et al., 2021; de Sá et al., 2018; Schellhaas et al., 2020; person knowledge: Abdel Rahman, 2011; Luo et al., 2016; Suess et al., 2014; Xu et al., 2016; affective communication: Wieser, Gerdes, et al., 2014), and the *LPC* (fear-conditioning: de Sá et al., 2018; Panitz et al., 2015; Rehbein et al., 2018; Sperl et al., 2021; Wiemer et al., 2021; aversive screams: Bruchmann et al., 2021; monetary reward: Hammerschmidt, Kulke, et al., 2018; person knowledge: Abdel Rahman, 2011; Baum et al., 2020; Kissler & Strehlow, 2017; Schindler et al., 2021; Xu et al., 2016). Despite this evidence, there is still uncertainty about the boundary conditions under which associated effects occur, for example, regarding the need for explicit awareness of a CS–US contingency for stable associations (Mertens & Engelhard, 2020). Implicit conditioning usually refers to the subliminal presentation of the CS^+ and not the US. Furthermore, many studies use salient aversive stimuli like electric shocks as US or explicit instructions to draw attention to the contingency between CS^+ and US, for example, by picturing actions or encounters of the CS^+ face (Aguado et al., 2012; Verosky et al., 2018). However, very few studies investigated whether and how robustly emotional contextual information is (automatically) associated, even when this information is not task-relevant. Task-irrelevant emotional stimuli have been supposed to capture attention (Armony, 2002; Morris et al., 1998; Öhman et al., 2001), in bottom-up or top-down manner, respectively, which is a prerequisite for automatic associations. In this line, amygdala activations have been reported for emotional visual stimuli, even when emotion was not task-relevant, but only when the task load was not too high (Pessoa, Kastner, & Ungerleider, 2002; Pessoa, McKenna, et al., 2002). In contrast, emotional auditory stimuli appear to be more robust against distractors as long as the attentional focus stays within the auditory modality (e.g., Ethofer et al., 2006; Quadflieg et al., 2008; Sander et al., 2005; cf. Bach et al., 2008). However, little is known about the cross-modal transfer of emotional information, particularly when features other than the emotional content of the US are relevant for the task during associative learning.

1.2 | Aim of the study

The present study aimed to fill this gap and specifically tested whether inherently neutral faces gain additional relevance when being associated with either positive or negative compared with neutral vocal affect bursts when the emotion of the burst is not task-relevant. We recorded ERPs and pupil size during a learning and a delayed test phase to investigate the temporal dynamics of the acquisition and extinction of the associated reactions. Based on previous research (Hammerschmidt et al., 2017; Hammerschmidt, Kagan, et al., 2018; Schacht et al., 2012; Sperl et al., 2021), we set an approximate 24-hour interval between learning and test to allow for memory consolidation (Menz et al., 2016; Pace-Schott et al., 2015; Sopp et al., 2017). We chose to associate neutral faces with vocal affect bursts because faces and voices are considered socially and biologically relevant, naturally co-occurring, and usually integrated into a holistic percept (Freeman & Ambady, 2011). Neutral facial expressions are more likely to be perceived as ambiguous (Schwarz et al., 2012; Wieser, Gerdes, et al., 2014; Yoon & Zinbarg, 2008) and suit well as CS^+ stimuli (Bublatzky et al., 2020). We presented vocal affect bursts as US, as they do not have the segmental structure of speech or pseudo speech, are relatively short in length, and unfold the emotional information rapidly. To ensure active processing of the auditory US, we presented gender-matching or gender-mismatching face-voice pairs during learning, where participants performed gender-congruence decisions. Since the face-voice pairing, and thus the gender congruence, was fixed for every trial, we randomly interspersed no-go trials (beep sound instead of the US) to counteract cross-modal inhibition (Johnson & Zatorre, 2006) and responses based solely on the target face after its repeated presentation.

1.3 | Hypotheses

Our overall hypothesis was that inherently neutral faces acquire emotional relevance through learned associations with affective (CS^+: CS^+_{pos}, CS^+_{neg}) but not neutral (CS_{neu}) vocal bursts. Emotional relevance was operationalized in terms of differential neural responses to the faces as a function of learning and extinction, and different speed and accuracy measures when performing a gender-matching (learning session) or gender (test session) decision.

During *learning*, we expected slower responses and lower accuracy for emotional (CS^+: CS^+_{pos}, CS^+_{neg}) and particularly for threatening (CS^+_{neg}; Öhman et al., 2001) face-voice pairs compared to neutral (CS_{neu}) face-voice pairs, in line with previous findings on the attentional binding of emotional information (Anderson, 2005; Gutiérrez-Cobo et al., 2019): Increased attention to emotional stimuli has been shown to interfere with tasks requiring processing of other, non-emotional, information (Schacht & Sommer, 2009; Zhang et al., 2019; cf: Roesch et al., 2010) Moreover, the emotional incongruence of the neutral CS^+ faces and emotional voices should lead to less efficient processing of the face-voice pairs and lower the behavioral performance (Föcker et al., 2011). Increased arousal and attention to the emotional compared to the neutral voices should increase the pupil size (e.g., Cosme et al., 2021; Kret et al., 2013), and this response may be elicited by the predicting CS^+ faces as a function of learning (for a review of pupil dilation in conditioning, see Finke et al., 2021). At the neural level, acquired valence associations of the CS^+ faces should modulate early processing (P1) and subsequent processing stages (i.e., N170, EPN, LPC). We expected enhanced P1 amplitudes for CS^+ faces, (similar to Aguado et al., 2012; Hammerschmidt et al., 2017), whereas for the N170, EPN, and LPC,[1] we expected a difference between CS^+ and CS_{neu} faces, but were uncertain about the direction due to mixed findings in the literature. Regarding the effects of gender congruence, we expected lower behavioral performance for gender-mismatching compared with matching face-voice pairs, as the conflicting input signals between the two modalities would interfere with an automatic and integrated perception of gender (Freeman & Ambady, 2011). We had no specific a priori pupil size and ERP-related hypotheses regarding differences between the gender-matching and mismatching conditions.

In the *test session* following overnight consolidation, we investigated the extinction of the associated effects by presenting solely the conditioned faces. Replicating findings of a behavioral pilot study ($N=40$), we predicted that behavioral effects for CS^+ faces would be reversed compared with learning: Gender decisions for CS^+ faces would be faster and more accurate than for CS_{neu} faces, possibly due to the increased consolidation and more robust memory traces of emotional information (e.g., Sharot & Phelps, 2004). Furthermore, we predicted happy associations (CS^+_{pos}) to enhance task performance more than angry associations (CS^+_{neg}), based on reward biases and potential commonalities between social and monetary rewards vs. punishments (Hammerschmidt et al., 2017; Rossi et al., 2017; cf. Öhman et al., 2001). We included a likability rating at the end of the test session and predicted that emotion associations should manifest in $CS^+_{pos} > CS_{neu} > CS^+_{neg}$ ratings (similar to Suess et al., 2014). According to Hammerschmidt, Kagan, et al. (2018), effects of associated emotional relevance should not become evident in pupil size, but in the same ERPs that would be modulated during learning, even if we assume that they would partly extinguish over the course of the test session. We did not have specific behavioral, pupil size, and ERP-related hypotheses with regard to the previously associated gender congruence on the test task.

[1]For the expected LPC effects, ongoing face processing and the overlapping sound onset might be hard to disentangle and thus require additional caution.

2 | MATERIALS AND METHODS

The study was pre-registered prior to data collection (https://osf.io/b3fh2).

2.1 | Stimuli

We selected 16 frontal portrait photographs of faces with neutral expressions from the Göttingen Faces Database (Kulke et al., 2017). The faces were presented in their natural color on a light gray background, edited, and combined with a transparent mask covering the hairline, ears, and neck. They had a visual angle of approximately 3.16×5.14 degrees and a 200×300 pixels resolution. Images were controlled for luminance (HSV: $M=0.47$, $SD=0.01$, $\chi^2(225)=240$, $p=.235$; Dal Ben, 2019). Vocal stimuli were taken from the Montreal Affective Voices database (Belin et al., 2008). Based on the findings by Lausen and Schacht (2018), we selected 12 sounds with the highest recognition of emotion (angry, happy, neutral) and gender (female, male). The duration of the selected sounds ranged from 511 to 1831 ms (for details see Supplementary Information). Non-parametric independence tests showed no significant difference between speaker's gender on duration ($Z=-0.98$; $p=.393$) but between emotion levels ($maxT=2.16$; $p=.031$), with a significant difference between neutral and happy stimuli ($Z=1.89$; $p=.025$). Their maximum intensity was digitally equalized (Praat; Boersma & Weenink, 2018), resulting in a mean peak sound level of $M=47$ dB ($SD=1.8$ dB) at the participants' head position. The "beep" tone for the no-go trials was a 630 Hz, 300 ms sinusoidal tone with an initial amplitude ramp of 30 ms. We presented 12 unique face-voice pairs whereby one face stimulus was contingently paired with one voice stimulus for each participant. Half of the face-voice pairs were gender-congruent (i.e., female face and voice or male face and voice), and the other half were gender-incongruent pairs (e.g., female face and male voice).

2.2 | Randomization

We created four different versions of the congruence-emotion condition to counterbalance the sound stimuli (based on how well emotion and gender were recognized in Lausen & Schacht, 2018) between congruence conditions across participants. The allocation of the face stimuli to the voices was fully randomized: We pseudo-randomly drew 12 out of 16 face stimuli (assuring six male and six female faces) and randomly assigned them to the 12 voices. The four remaining faces were used as new, that is, not

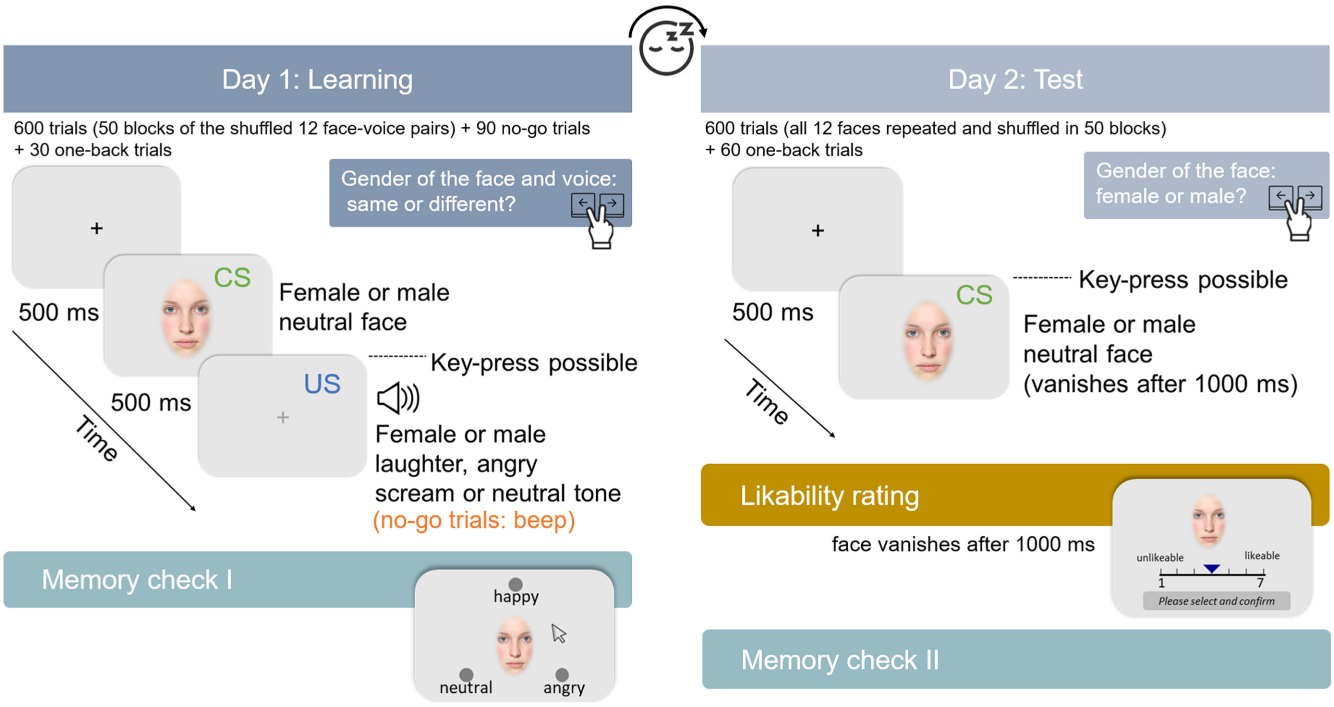

FIGURE 1 Procedure of the learning and test session.

associated, faces only for the likeability rating. Stimuli were presented in 50 blocks, with each block containing a random sequence of the 12 stimuli (learning session: face-sound pairs; test session: faces only). For the memory checks after the learning and the test session, the order of the 12 faces and the emotion category labels' positions were also shuffled.

2.3 | Procedure

Prior to the experimental sessions, participants completed an online questionnaire of the German version of the Social Interaction Anxiety Scale (SIAS, Stangier et al., 1999). The laboratory sessions took place on two subsequent days and lasted approximately two hours each. At the beginning of each session, participants gave written consent to participate voluntarily in the study. At the beginning and the end of each session, we assessed their current mood with the German version of the Positive and Negative Affect Schedule (PANAS, Breyer & Bluemke, 2016, see Supplementary Information). Additionally, we assessed socio-demographic data and handedness (day one) and quality of sleep with a modified version of the Pittsburgh Sleep Quality Index (PSQI, Buysse et al., 1989) (Day 2). Participants were seated in front of a computer screen in a dimly lit, electrically shielded, and sound-attenuated room at a viewing distance of approx. 78 cm from the presented face stimuli. They positioned their chin in a height-adjustable chin rest to avoid head movements. Two loudspeakers were placed to the left and right and at the height of the monitor. For the presentation of the experiment, we used Python (2.7), including the modules PsychoPy (Peirce, 2009), PyGame (Shinners, 2011), and PyGaze (Dalmaijer et al., 2013) among standard modules. At the beginning of both sessions, we presented detailed instructions about the task on screen, and participants completed six practice trials (incl. feedback). After ensuring that the task was understood and calibrating the eye-tracker, the main experimental task began. A visualization of the procedure is shown in Figure 1.

2.3.1 | Learning session

The task was to indicate as fast and accurately as possible via key press whether the gender of the face and voice matched. At the center of the screen, first, a fixation cross and then the neutral face stimulus were presented for 500 ms each, followed by the vocal stimulus (negative, neutral, or happy affect burst). Participants could respond as soon as the voice started.[2] The next trial started automatically after a response and a variable intertrial interval ($M = 1800$ ms; $SD = 200$ ms). At random positions, we included filler trials (90 no-go trials and 30 one-back tasks)

[2]In separate experimental checks, we measured an asynchrony between face offset and voice onset with an audio photo-diode, ranging between −9 and +9 ms and occurring independently from certain stimuli or conditions due to hardware imprecision.

to motivate participants to stay focused. The no-go condition, in which a beep sound followed the face and in which participants were not allowed to press a key, was implemented to ensure that attention was focused on to the auditory stimulus. Thus, it was not sufficient to learn/know the assignment of a face to a response key (which was always consistent within each participant). Each participant completed 50 × 12 = 600 trials (+ filler trials). There was a short break to rest break after every 120 trials. *Memory check*. After the learning session, we assessed whether participants were able to match the faces and the emotional categories of the voices, although they were not instructed at any time to memorize the faces or the face-voice pairs. We presented each face individually, and participants had to click on one of three labeled buttons (happy, angry, neutral) around the face to indicate the emotional category of the associated voice.

2.3.2 | Test session

The following day, participants performed a gender decision task on the previously associated faces but without any voices or sounds present. The faces were displayed for 1000 ms, and the response was indicated by a key press. Again, participants completed 50 blocks of the shuffled 12 faces and 60 additional one-back tasks.

2.3.3 | Likability rating and second memory check

After completing the main part of the test session, all previously associated faces were presented once each, intermixed with four new faces, and participants had to judge the faces concerning their likability on a 7-point Likert scale (1 = "unlikeable," 7 = "likable"). This was followed by the same explicit memory check as in the first session. Finally, participants were debriefed about the aims of the study and were able to clarify any open questions with the experimenter.

2.4 | Sample size and power analysis

The study had a 2 (gender congruence: match/mismatch) × 3 (emotion: angry/happy/neutral) within-subject design. As there is no standardized way to do a power analysis and sample size estimation for linear mixed models, we based the power estimation on a within-factors repeated-measures ANOVA (G*Power 3.1.9.2, Erdfelder et al., 1996), assuming a correlation among the repeated factors of .50, targeting a power of 0.80 with an alpha level of .05 to detect an effect of intermediate size (Cohens $f^2 = 0.04$) with 40 participants. However, it served only as a rough estimate because it did not accurately reflect the analyses that were planned and conducted. With the beginning of the COVID-19 pandemic, we had to stop data collection, that is, after 41 participants had completed both sessions. Unlike pre-registered we did not collect further data after the complete laboratory shutdown, because the data loss for both sessions and measures was very uneven and due to the hard hygiene constraints that existed at the time. For the learning session in particular, we had more ERP trials with artifacts than we had initially anticipated. To make the data from the learning and test sessions comparable, we also excluded the data for the other measures and sessions to ensure the same group of 32 participants in the learning and testing sessions. The (observed) simulated power and effect sizes for the test session based on 32 participants and 39 participants are included in Table A48 in the Supplementary Information. Tables A49 and A50 of the Supplementary Information list the number of trials rejected for each participant and which participants were included in the analysis.

2.5 | Participants

Our final sample consisted of 32 participants (22 female, 10 male, 0 diverse; age: 19–34 years, mean = 23.5 years). All participants were right-handed (according to Oldfield, 1971), fluent in German, and did not self-report any (neuro-)psychiatric disorders. Participants with visual correction of more than plus/minus one diopter or any self-reported hearing difficulties were excluded. Participants were recruited through advertisements on campus and in social network groups in Göttingen; hence, the sample consisted mainly of students (29 out of 32). Participation was reimbursed by a fixed amount of money or course credits. The study was conducted following the Declaration of Helsinki (WMA, 1964), and all participants signed an informed consent form before both experimental sessions. The mean SIAS score of our sample was 23.19 (range: 7 to 44) out of max. 80. Seven[3] participants showed elevated scores (>30, Stangier et al., 1999).

2.6 | EEG recording and pre-processing

We recorded EEG from 64 (+6 external) electrodes during the learning and test session. Participants wore an electrode cap (Easy-Cap, BioSemi, Amsterdam, Netherlands)

[3]IDs: 6, 7, 21, 28, 23, 36, 40.

according to the extended 10–20 system (Pivik et al., 1993). External electrodes were positioned at the left and right mastoids, at the outer canthi of and below both eyes to record electro-oculograms. Common Mode Sense (CMS; active) and Driven Right Leg (DRL; passive) are special "ground" electrodes serving as an online reference during recording (see www.biosemi.com/faq/cms&drl.htm). Continuous EEG was recorded with a sampling rate of 512 Hz and a bandwidth of 102.4 Hz. Offline, the raw data were pre-processed in MATLAB (2018) with EEGLAB (v2019.0, Delorme & Makeig, 2004). We shifted all event markers by a constant of 24.3 ms to account for the monitor's systematic delay in stimulus appearance. Data were re-referenced to average (whole head) reference excluding external electrodes and filtered with a 0.01 Hz high-pass filter. The plugin "CleanLine" (v1.04, Mullen, 2012) was used to remove 50 Hz line noise. Data were epoched from −500 to 2000 ms and corrected to a 200 ms pre-stimulus baseline. We performed Independent Component Analysis (ICA) on a 1 Hz high-pass filtered copy of the data set and subsequently transferred the resulting ICA weights to the original 0.01 Hz filtered data set. ICA components were used to detect eye and muscle-related activity in the data. Data were corrected by removing components with a high probability of being labeled as such (muscle >80%, eye-related >90%, or channel noise >90%) using "IClabel" (v1.2.4, Pion-Tonachini et al., 2017). Consequently, channels were interpolated if classified as bad. We trimmed epochs to −200 to 1000 ms and performed trial-wise rejections: amplitudes exceeding −100/100 µV (Learning: avg. 7.7%; Test: 7.4%) during face presentation, steep amplitude changes (5000 µV within epoch; Learning: avg. 3.9%; Test: 2.8%) or improbable activation (>5 deviation of mean distribution for every time point; Learning: avg. 11.9%; Test: 12.8%) were excluded. Overall, there was a mean rejection rate of 17.3% (range: 9.7%–36.8%) of trials for the learning session and 17.2% (range: 6.5%–36.2%) for the test session due to these artifacts.[4] As eyeblinks were corrected with ICA, we extracted blink information from the pupil data to reject trials in which participants blinked during face presentation. We defined the time windows and regions of interest (ROIs) electrodes for the ERP components of interest based on previous research with similar stimuli and settings (face-locked ERPs: Hammerschmidt et al., 2017; voice-locked ERPs: Paquette et al., 2020; Pell et al., 2015) as follows: For the visual (face-locked) components: (a) P1: mean and peak amplitudes, 80–120 ms; occipital electrode cluster: O1, O2, and Oz; (b) N170: mean and peak amplitudes, 130 and 200 ms; occipito-temporal electrode cluster: P10, P9, PO8, and PO7[5]; (c) EPN: mean amplitudes, 250–300 ms; occipito-temporal cluster: O1, O2, P9, P10, PO7, and PO8; (d) LPC: mean amplitudes, 400 and 600 ms; occipito-parietal electrode cluster: Pz, POz, PO3, and PO4. In addition to the pre-registered face-locked components, we analyzed voice-locked[6] ERPs with the following ROIs taken from Paquette et al. (2020) as they also used stimuli of the Montreal Affective Voices database (MAV, Belin et al., 2008): N1-P2 complex with N1 (90–145 ms) and P2 (165–300 ms), both with the identical fronto-central electrode cluster: F3, F1, Fz, F2, F4, FC1, FC3, FC2, FC4, C3, C1, Cz, C2, C4, CP1, CP3, CPz, CP2, and CP4.

2.7 | Pupil recording and pre-processing

Pupil size was recorded binocularly in arbitrary units (AUs) at a sampling rate of 500 Hz using the EyeLink 1000 desktop mount eye-tracker (SR Research, Mississauga, ON, Canada). Before the start of the experiment, a 9-point eye-tracking calibration- and validation procedure was performed. We based artifact detection in pupil samples on the guidelines proposed by Kret and Sjak-Shie (2018). The pupil time series were time-locked to the onset of the face stimulus, and all artifact detection was performed sequentially. Samples were classified as blinks or invalid when both eyes were lost. We marked invalid samples and specific trial windows to indicate trials in which the participant missed the stimulus (−onset), for example, during baseline, face presentation, or later. Median absolute deviation (MAD) speed was estimated, and samples with a speed higher than 16 times MAD were marked as artifacts. A smoothed trend line was calculated, and clusters of samples with a strong deviation from the trend line were flagged (in four iterations). Isolated samples within longer periods of missing data (separated clusters) were dismissed. Samples at the border of a gap were trimmed (50 samples pre- and post-gap, i.e., "extended blinks") and interpolated, but only for gaps with a maximum of 125 samples of missing data. The eye with fewer invalid trials was selected, and a baseline adjustment was performed (samples were subtracted by the mean of the baseline period from −200 ms to face onset). The time course of a trial was segmented into 60 bins, and outlier samples within a bin were flagged (>3 standard deviations from the bin mean). A trial was rejected if >75% of

[4]Referring to the reported 32 data sets.

[5]PO9 and PO10 were replaced by PO8 and P07 as they were not part of the used EEG recording system.

[6]Our setting allowed measuring the exact voice onset with an audio-photo-diode only without the use of the speakers and not during the actual experiment. For the target event to epoch the data, we used the offset of the face stimulus. Due to the slight jitter, the timing of the the auditory ERPs is less precise compared to the visual ERPs.

its samples or pre-specified time windows of the trial were invalid. A smoothing 4 Hz filter was applied to the data before averaging by participant and condition.

2.8 | Statistical analysis

Tables with statistical models (incl. estimates, confidence intervals, stability measures, and likelihood ratio tests) are in the Supplementary Information. We used linear mixed models to analyze RTs and neurophysiological data, aggregated by participant and condition, using the function "lmer" of the package "lme4" (Bates et al., 2015). All statistical analysis was conducted in R (v 4.0, R Core Team, 2020). For parameter estimation, we chose the maximum likelihood (ML) method. The model predictors are the *emotion* of the associated sounds ("neutral", "angry", "happy"), *congruence* of the face-voice pairs ("match", "mismatch"), and a random intercept (participant) to consider the dependency in the data due to the repeated-measures design and variability between participants. We compared models (full model including the interaction, additive models, and leaving out each predictor) with goodness of fit tests, known as likelihood ratio (LR) chi-square difference tests, to identify which predictors add significantly to explaining variance in the data (Snijders & Bosker, 2012). For likelihood ratio tests (LRT), we used the "mixed" function of the package "afex" (Singmann et al., 2020). Regression coefficients (β), standard errors (*SE*), 95% confidence intervals (CI), and stability of the coefficients are reported. To obtain CIs, we used a parametric bootstrap ($N=1000$). We estimated the stability of model coefficients by fitting the same model on subsets of the data (dropping one random effect at a time). Residuals of the models were inspected visually, and potential collinearity among predictors was determined with Variance Inflation Factors (VIF), which will be reported if model assumptions appeared violated. Reference levels in all models were "match" for congruence and "neutral" for emotion, respectively. The following linear mixed models are sum-contrast-coded, reflecting main effects rather than marginal effects. Here, the intercept corresponds to the (unweighted) grand mean, and lower-level effects are estimated at the level of the grand mean. This coding implies that the reference factor level ("match" or "neutral," respectively) receives a value of −1 on all contrast variables, whereas all other factor levels are mapped onto exactly one contrast variable with a value of 1. This implies that for every factor with k levels, k-1 parameters are estimated and cannot be directly mapped to the factor levels. Post hoc Šidák adjusted tests were used to test the difference between levels of factors with more than two levels, using "emmeans" (Lenth, 2020).

For the ERPs (P1, N170, EPN, and LPC) and pupil size, in addition to overall effects, we planned to explore the *dynamics* of acquisition during learning and a possible extinction of the association during the test session (only for ERPs). To this end, we applied a flexible regression approach using Generalized Additive Models of Location Shape and Scale (GAMLSS; Rigby & Stasinopoulos, 2005), including a smoothing function (Eilers & Marx, 2010) to model changes over the course of the learning and test session. The model results and visualizations of the model predictions can be found in the Supplementary Information.

2.9 | Exclusion criteria and re-coding

We detected unexpected systematic errors in some of the participant's behavioral data. To preserve as much power as possible, we decided to apply additional criteria for exclusion or re-coding in the following cases: (a) Same errors occurring consistently (i.e., in at least two-thirds of trials) for a given stimulus pair (face + voice): As it was not possible to distinguish in the learning session whether the participants had difficulties identifying the gender of the face or the gender of the voice, we only re-coded cases in which the same faces were consistently misclassified in both the learning and the test session. This happened in four cases: VP 35 (l10_neu_m.png), VP 27 (l5_neu_f.png), VP 7 (l5_neu_f.png) and VP 36 (l2_neu_f.png). In these cases, previously incorrect responses were re-coded as correct, and the congruence of the stimulus pair was changed (if originally incongruent to congruent and vice versa). (b) All trials with (systematic) errors which occurred only in the learning session but not in the test session were excluded and not re-coded. (c) One participant (VP 25) initially confused the key assignment and answered all trials incorrectly. After re-instructing the participant, all subsequent trials were answered correctly. We did not recode the answers of this participant but excluded the trials before the re-instruction. Unlike pre-registered and due to the unexpected systematic error patterns, we only report descriptive statistics of the accuracy data in this study.

2.10 | Outlier removal and model robustness

Despite the EEG and eye-tracking clean-up, there was still pronounced variability and outlier observations in some components and measures. Instead of excluding a participant if they lost more than 25% of the trials as pre-registered, we set a lower limit of 30 valid trials per condition. To increase model stability and robustness, we consistently excluded observations for all pupil size and ERP models in both sessions for which their cook's distance was larger than 0.5 in any model. These exclusion procedures resulted in a final sample size of 32 participants for all ERP measures.

3 | RESULTS

3.1 | Implicitness of the association learning

Participants differed in their ability to recall the emotion category of the voices when the faces were presented separately from the voices. Some participants performed significantly above chance, that is, ≥8 out of 12 faces correct, corresponding to a 2-sided Exact Binomial Test with $p \leq .05$. After the learning session, 5 out of 32 participants and after the test session, 3 out of 32 participants met this criterion. However, only one participant performed above chance in both session checks and would therefore not be considered an implicit learner.

3.2 | Learning session

Table 1 contains all means and standard deviations of the visual and auditory ERPs, pupil size, and behavioral measures.

3.2.1 | Face-locked ERPs

P1: P1 mean amplitudes were not significantly modulated by emotion ($\chi^2(2)=3.91$, $p=.141$). There was a trend toward an effect of gender congruence ($\chi^2(1)=3.41$, $p=.065$) with larger mean amplitudes ($d_{\text{mismatch-match}}=0.19\,\mu\text{V}$) in gender-mismatching than matching trials. The interaction between emotion and congruence was not significant ($\chi^2(2)=3.6$, $p=.165$). There were no significant effects on P1 peak amplitudes, neither for emotion ($\chi^2(2)=2.86$, $p=.239$), congruence ($\chi^2(1)=0.68$, $p=.409$), nor their interaction ($\chi^2(2)=4.05$, $p=.132$). *N170*: There was a trend for N170 mean amplitudes to be modulated by emotion ($\chi^2(2)=5.75$, $p=.057$), with happy ($-5.39\,\mu\text{V}$) and angry ($-5.40\,\mu\text{V}$) descriptively being less negative compared with neutral-associated faces ($-5.66\,\mu\text{V}$) when averaged across gender-congruence conditions. However, post hoc contrasts between emotion levels were not significant (all $ps \geq .05$). There was no significant modulation by congruence ($\chi^2(1)=2.47$, $p=.116$), and no significant interaction between congruence and emotion ($\chi^2(2)=2.6$, $p=.273$).

TABLE 1 Learning session: Mean and SD reaction times in ms, ERP mean amplitudes in μV and Pupil size [AUs] for congruence and emotion.

	Match						Mismatch					
	Neutral		Happy		Angry		Neutral		Happy		Angry	
Measure	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
P1	3.95	3.16	3.58	3.03	3.50	3.07	3.87	3.05	3.85	2.97	3.87	3.25
P1.peak	6.30	3.59	5.90	3.29	5.76	3.32	6.08	3.34	6.01	3.27	6.17	3.66
N170	−5.80	3.44	−5.52	3.23	−5.37	3.33	−5.51	3.29	−5.26	3.35	−5.44	3.02
N170.peak	−10.27	4.77	−10.07	4.55	−10.02	4.65	−10.04	4.61	−9.77	4.60	−10.00	4.44
EPN	0.27	3.75	0.23	3.95	0.38	3.40	0.84	3.78	0.61	3.62	0.55	3.61
LPC	3.09	2.35	3.23	2.20	3.14	2.46	2.99	2.30	2.95	2.20	2.78	2.37
N1(auditory)	−0.83	0.77	−0.69	0.81	−0.59	0.79	−0.72	0.79	−0.69	0.72	−0.72	0.77
P2(auditory)	1.12	1.35	1.38	1.29	1.29	1.26	0.79	1.16	0.88	1.19	0.97	1.28
Pupil.early	−20.20	40.13	−18.26	40.30	−21.34	48.05	−20.22	38.52	−25.69	40.72	−28.59	45.55
Pupil.late	83.83	75.15	91.02	80.75	81.21	80.91	100.38	88.49	90.33	86.34	79.17	84.82
RT	691.94	165.66	734.95	165.24	713.70	162.32	785.11	164.11	833.31	192.28	791.97	173.74
Accuracy	94.34	11.78	96.41	3.93	93.62	10.76	93.06	13.60	96.06	8.86	88.66	17.48

N170 peak amplitudes were not significantly modulated by emotion ($\chi^2(2)=2.24$, $p=.327$), nor congruence ($\chi^2(1)=2.03$, $p=.155$) nor their interaction ($\chi^2(2)=0.82$, $p=.664$). *EPN*: The mean amplitudes of the EPN component were not significantly modulated by emotion ($\chi^2(2)=0.73$, $p=.693$) but by congruence ($\chi^2(1)=7.49$, $p=.006$) with less negative amplitudes for gender-mismatching (0.67 μV; $\beta_{\text{mismatch}}=0.19$, $SE=0.07$, $t=2.72$) compared to matching trials (0.29 μV). No interaction between congruence and emotion was present ($\chi^2(2)=1.47$, $p=.480$), see also Figure 2.

LPC: LPC mean amplitudes were not significantly modulated by emotion ($\chi^2(2)=0.84$, $p=.658$), but there was a main effect of congruence ($\chi^2(1)=4.21$, $p=.040$). Gender-matching trials (3.15 μV) had more positive amplitudes compared with mismatching trials (2.91 μV; $\beta_{\text{mismatch}}=-0.12$, $SE=0.06$, $t=-2.03$). The interaction between congruence and emotion ($\chi^2(2)=0.84$, $p=.657$) was not significant, see also Figure 3.

3.2.2 | Voice-locked ERPs

N1: For the N1 component, neither emotion ($\chi^2(2)=3.35$, $p=.187$) nor congruence ($\chi^2(1)=0.01$, $p=.918$) nor their interaction ($\chi^2(2)=2.98$, $p=.226$) were significant. *P2*: The P2 component was modulated both by emotion ($\chi^2(2)=6.3$, $p=.043$) and congruence ($\chi^2(1)=29.51$, $p<.001$). However, there was no interaction between congruence and emotion ($\chi^2(2)=1.44$, $p=.486$). Gender-mismatching trials (0.88 μV; $\beta_{\text{mismatch}}=-0.19$, $SE=0.03$, $t=-5.60$) had a smaller P2 amplitude compared to matching trials (1.26 μV). Emotional voice stimuli (angry: 1.13 μV; $\beta_{\text{angry}}=0.06$, $SE=0.05$, $t=1.22$; happy: 1.13 μV; $\beta_{\text{happy}}=0.06$, $SE=0.05$, $t=1.27$) elicited more positive P2 amplitudes compared to the neutral voice stimuli (0.95 μV). When adjusting for multiple comparisons, pairwise comparisons failed significance. Descriptively, the largest difference was between neutral and happy ($d_{\text{hap-neu}}=0.18$, $p=.090$) bursts followed by neutral and angry bursts ($d_{\text{ang-neu}}=0.18$, $p=.097$). Both, happy and angry bursts showed a similar pattern ($d_{\text{hap-ang}}=0.00$, $p=1.000$). A visualization of the auditory ERP results is presented in Figure 4.

3.2.3 | Pupil size modulations

For the early time window (0–1000 ms after face onset), which mainly reflects a modulation of the pupil constriction, there was no significant modulation of the pupil size by emotion ($\chi^2(2)=3.69$, $p=.158$), but there was a main effect of congruence ($\chi^2(1)=5.7$, $p=.017$), with a stronger constriction for mismatching stimuli ($d_{\text{mismatch-match}}=-4.90$, $p=.019$). No interaction between emotion and congruence was present ($\chi^2(2)=2.86$, $p=.240$). In a later time window (1000–2000 ms after face onset), the pupil size was significantly modulated by emotion ($\chi^2(2)=8.07$, $p=.018$). Pairwise comparisons revealed significant differences in pupil size between the angry and neutral condition ($d_{\text{ang-neu}}=-11.92$, $p=.031$), but no significant differences between happy and neutral trials ($d_{\text{hap-neu}}=-1.43$, $p=.986$) and happy and angry trials ($d_{\text{hap-ang}}=10.49$, $p=.070$). There was no main effect of congruence in the later time window ($\chi^2(1)=1.55$, $p=.214$). Although there was only a trend for interaction between emotion and congruence ($\chi^2(2)=5.16$, $p=.076$), the interpretation of the main effect shall be taken with caution. A visualization of the pupil results can be found in Figure 5.

3.2.4 | Behavioral measures

Accuracy

Prior to any re-coding or rejection, the overall accuracy in the gender-matching task was 94% ($N=32$), with descriptively higher accuracy for match (95%) than for mismatch trials (93%). The accuracies for each emotion category were 96% for happy, 91% for angry, and 94% for neutral trials. None of the subjects fell into the exclusion criteria (>25% incorrect trials). After re-coding systematic error patterns (see Section 2.9), the overall accuracy was 94%. Again, match trials had a higher accuracy (95%) compared to mismatch trials (93%), and concerning emotion categories, happy trials (97%) had a higher accuracy compared with neutral (94%) and angry (92%) trials.

Response time

RT data were analyzed only for correctly answered trials. First, we trimmed RTs using a maximum cutoff of 5000 ms. Then, we applied a skewness-adjusted boxplot method to exclude extreme values separately for every subject, using the function "adjbox" of the package "robustbase" (Maechler et al., 2021; based on: Hubert & Vandervieren, 2008). The overall (non-aggregated) mean RT for the gender-matching task of the learning session was 758 ms ($SD=369$ ms). We based the RT model estimation on aggregated data, taking the mean for each condition (emotion, congruence) and subject.[7] Results showed a modulation by both, congruence ($\chi^2(1)=71.84$, $p<.001$) and emotion ($\chi^2(2)=15.49$, $p<.001$), but there was no interaction ($\chi^2(2)=0.81$, $p=.667$). Matching trials (714 ms) were answered faster compared to mismatching trials

[7] Taking the median instead of the mean did change parameters slightly but not the direction or significance of the effects.

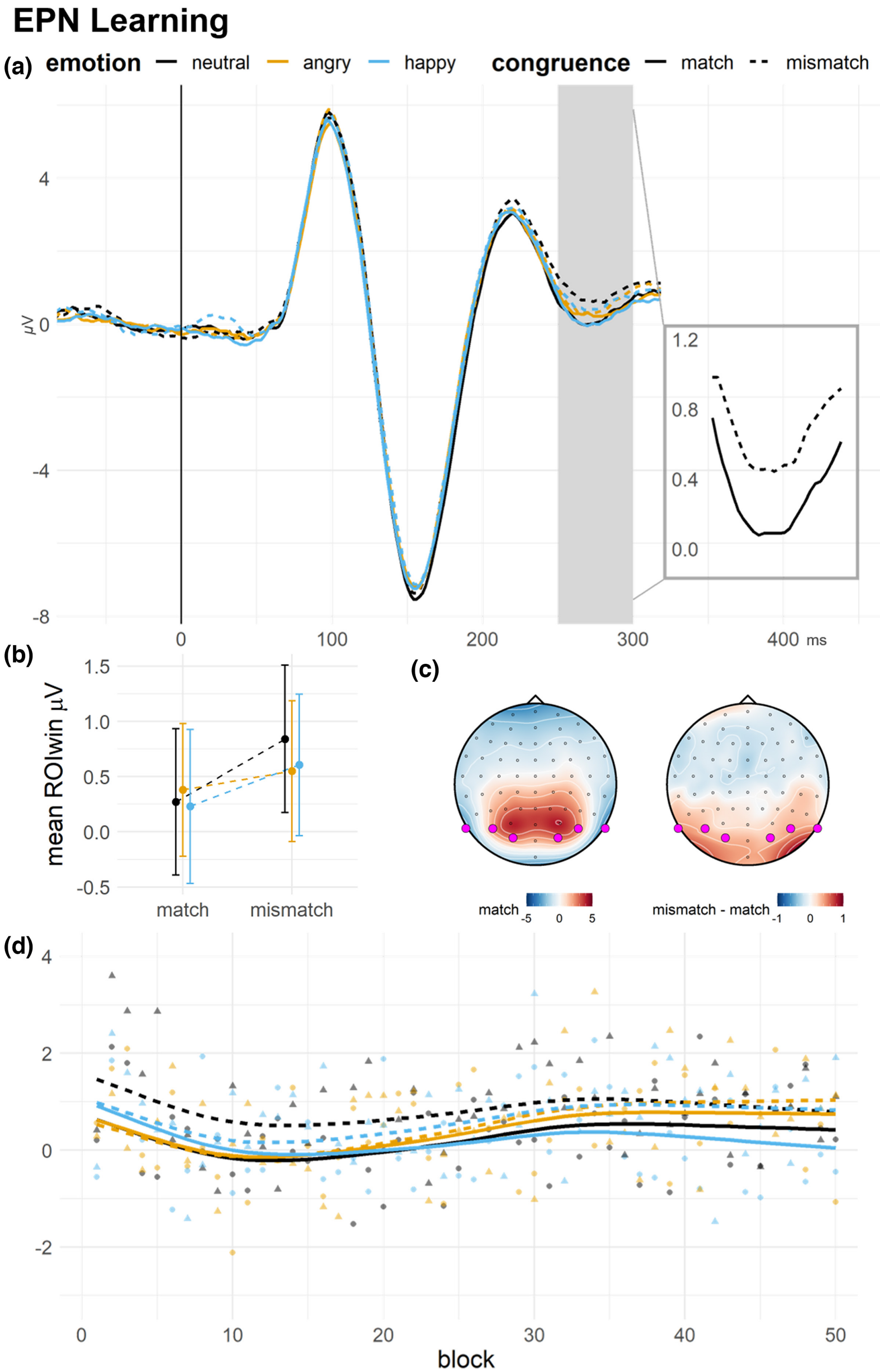

FIGURE 2 Face-locked EPN in the learning session. (a) Grand average ERP of the averaged ROI channels. The highlighted area displays the ROI time window. The zoomed-in window shows the main difference of gender congruence, averaged over all emotion conditions. (b) Grand averaged ERP amplitudes of the ROI, contrasted for all conditions. Error bars indicate ±1 SE of the mean. (c) Topographies of the ERP distribution for gender-congruent faces and the difference between gender-incongruent and congruent faces. ROI channels are highlighted in pink. (d) Mean EPN amplitudes over the course of the learning session. Dots represent the grand averages per block and condition. The curves represent the fitted values of the GAMLSS model (see Supplementary Information).

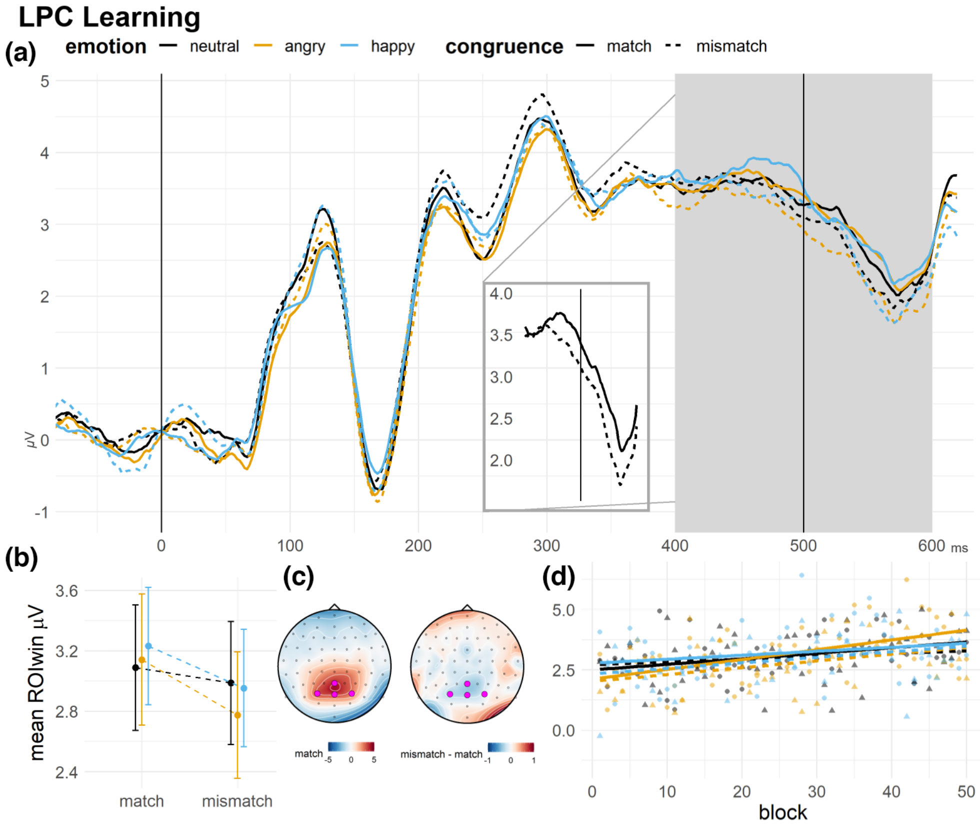

FIGURE 3 Face-locked LPC in the learning session. (a) Grand average ERP of the averaged ROI channels. The highlighted area displays the ROI time window. The zoomed-in window shows the main difference of gender congruence, averaged over all emotion conditions. (b) Grand averaged ERP amplitudes of the ROI, contrasted for all conditions. Error bars indicate ±1 SE of the mean. (c) Topographies of the ERP distribution for gender-congruent faces and the difference between gender-incongruent and congruent faces. ROI channels are highlighted in pink. (d) Mean LPC amplitudes over the course of the learning session. Dots represent the grand averages per block and condition. The curves represent the fitted values of the GAMLSS model (see Supplementary Information).

(803 ms; $\beta_{mismatch} = 44.97$, $SE = 4.80$, $t = 9.37$). Neutral trials were answered fastest (739 ms), followed by angry (753 ms; $\beta_{angry} = -5.66$, $SE = 6.79$, $t = -0.83$) and happy trials (784 ms; $\beta_{happy} = 25.63$, $SE = 6.79$, $t = 3.78$). Pairwise comparisons revealed that estimated RT differed significantly between neutral (fastest) and happy (slowest) trials ($d_{hap\text{-}neu} = 46$ ms, $p < .001$) and between happy and angry trials ($d_{hap\text{-}ang} = 31$ ms, $p = .025$).

3.3 | Test session

Table 2 contains all means and standard deviations of the visual ERPs, pupil size, and behavioral measures.

3.3.1 | Face-locked ERPs

P1: P1 mean amplitudes were not significantly modulated by emotion ($\chi^2(2) = 2.73$, $p = .255$), but by congruence ($\chi^2(1) = 6.3$, $p = .012$). No interaction was found ($\chi^2(2) = 0.08$, $p = .963$). Similar to the learning session, gender-mismatching trials (5.27 µV; $\beta_{mismatch} = 0.14$, $SE = 0.06$, $t = 2.49$) had a descriptively higher P1 amplitude than matching trials (5.00 µV), although the sound stimuli were not presented anymore. A similar pattern was found for P1 peak amplitudes, which were not modulated by emotion ($\chi^2(2) = 3.52$, $p = .172$), but by congruence ($\chi^2(1) = 5.52$, $p = .019$). Again, no interaction between emotion and congruence was found ($\chi^2(2) = 0.19$, $p = .910$), see also Figure 6.

N170: N170 mean amplitudes were not significantly modulated by emotion ($\chi^2(2) = 1.7$, $p = .428$) or congruence ($\chi^2(1) = 2.66$, $p = .103$),[8] but there was a significant interaction between emotion and congruence ($\chi^2(2) = 6.16$, $p = .046$), with the difference between previously matching and mismatching trials was larger for neutrally associated faces (0.38 µV, $t = 2.68$, $p = .008$) compared with faces previously associated with affective sounds ($\beta_{happy} = 0.13$ µV, $t = 0.92$, $p = .361$; $\beta_{angry} = -0.12$ µV, $t = -0.81$,

[8] For $N = 39$, the congruence effect of the N170 was significant ($\chi^2(1) = 4.93$, $p = .026$).

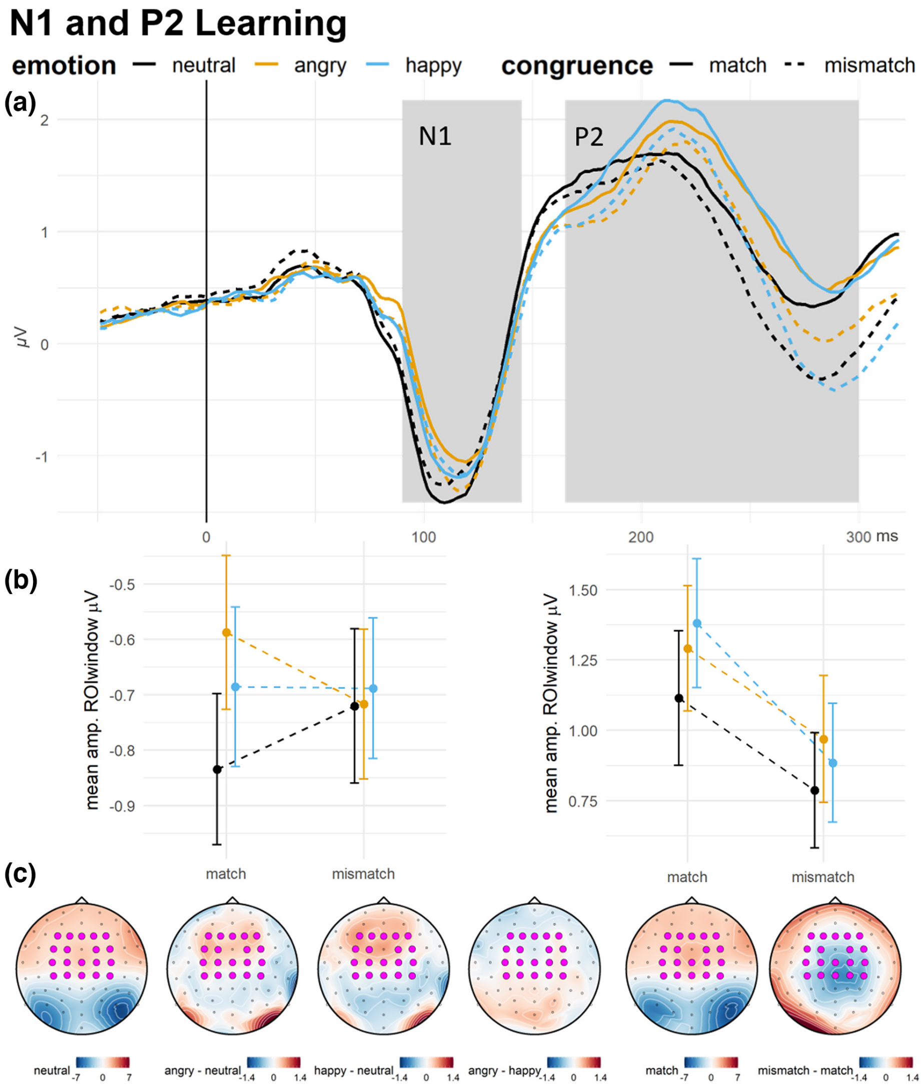

FIGURE 4 Voice-locked N1 and P2 in the learning session. ROI channels are highlighted in pink and identical for the N1 and P2. (a) Grand average ERP of the averaged ROI channels. The highlighted areas display the ROI time windows of the N1 (left) and P2 (right). (b) Grand averages of the N1 ROI (left panel) and P2 ROI (right panel), contrasted for all conditions. Errorbars indicate ±1 SE of the mean. (c) Topographies of the ERP distribution of the P2, depicting the main effects of congruence and emotion: Neutral bursts (collapsed across gender congruence levels), pairwise differences of the emotion levels, and gender-congruent bursts (collapsed across emotion levels) compared with gender-incongruent bursts. ROI channels are highlighted in pink.

$p = .421$). For neutrally associated faces in previously mismatching trials ($-5.47\,\mu V$), the N170 showed a less negative deflection compared with matching trials ($-5.86\,\mu V$). In contrast to N170 mean amplitudes, peak amplitudes were neither significantly modulated by emotion ($\chi^2(2) = 0.18$, $p = .914$), nor congruence ($\chi^2(1) = 0.98$, $p = .323$) nor their interaction ($\chi^2(2) = 3.25$, $p = .197$), see also Figure 7.

EPN: For EPN mean amplitudes, no significant modulation by emotion was present ($\chi^2(2) = 1.59$, $p = .451$). There was a main effect for congruence ($\chi^2(1) = 8.91$, $p = .003$), with smaller amplitudes for faces associated with gender-mismatching compared with matching sounds, similar to the learning phase. The interaction between congruence and emotion ($\chi^2(2) = 2.8$, $p = .247$) was not significant, see also Figure 8.

LPC: LPC mean amplitudes were neither significantly modulated by emotion ($\chi^2(2) = 0.8$, $p = .669$), nor by congruence ($\chi^2(1) = 0.04$, $p = .843$), nor their interaction ($\chi^2(2) = 3.33$, $p = .189$).[9]

[9] For $N = 38$, the interaction of emotion and congruence of the LPC was significant ($\chi^2(1) = 4.93$, $p = .026$), however, none of the post hoc contrasts were (all $p \geq .1$) with the largest difference between associated gender-congruent anger and neutral faces ($0.31\,\mu V$; $t = 1.81$, $p = .199$). Note that one participant was excluded from this model due to influential observations (Cook's distance >0.5).

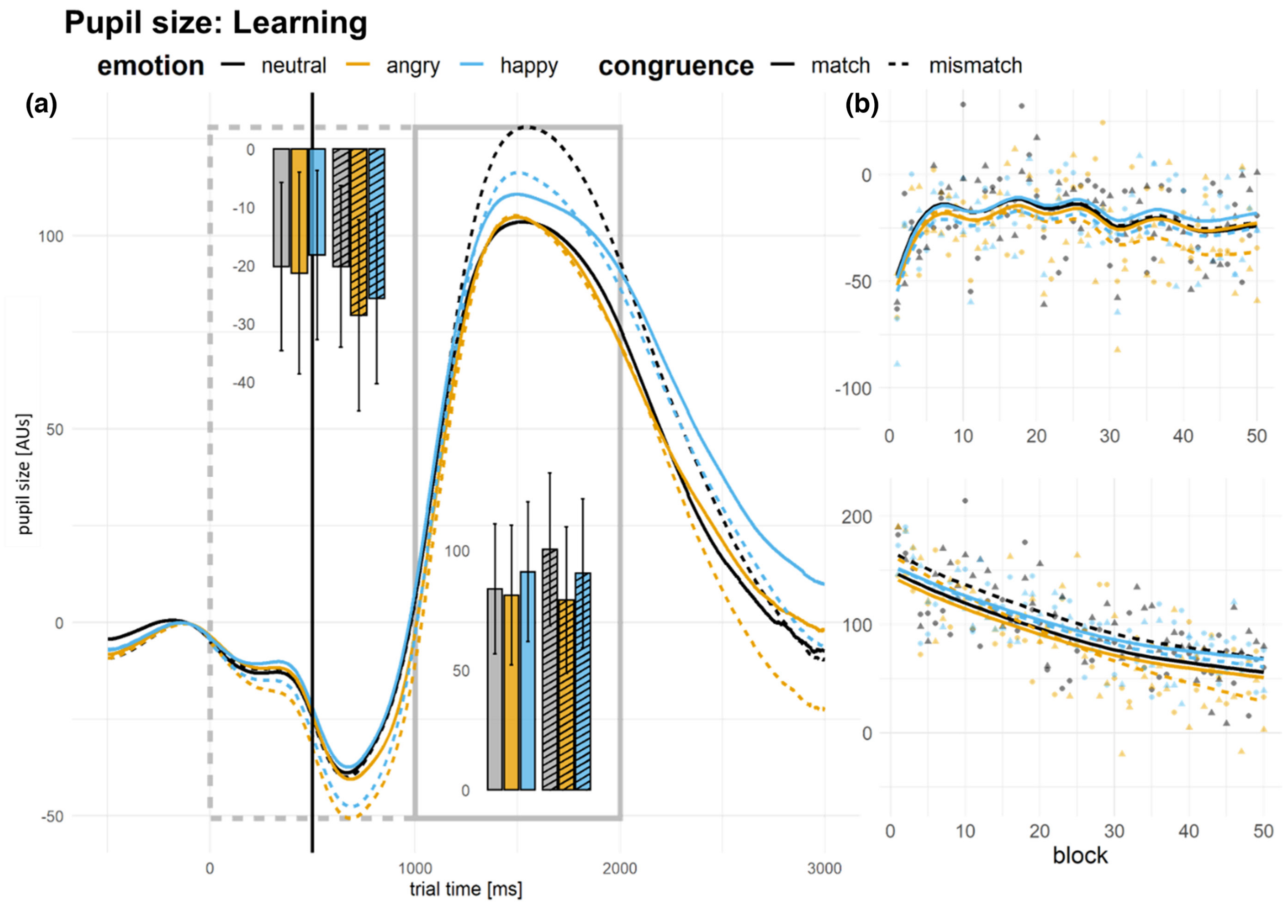

FIGURE 5 Pupil size results of the learning session. (a) Grand average pupil size time series. Bar plots refer to the pupil size in each condition for the respective marked time windows. (b) Pupil size time series across the experiment for the early time window (0–1000 ms; upper panel) and the later time window (1000–2000 ms; lower panel; see Supplementary Information).

3.3.2 | Pupil size modulations

For the pupil early time window (0–1000 ms) there were no significant effects of emotion ($\chi^2(2) = 1.05$, $p = .591$) nor congruence ($\chi^2(1) = 0.01$, $p = .911$) nor their interaction ($\chi^2(2) = 1.23$, $p = .540$). Similarly, pupil size was not modulated in a later time window (1000–2000 ms) by emotion ($\chi^2(2) = 0.37$, $p = .832$) or congruence ($\chi^2(1) = 0.07$, $p = .791$) nor their interaction ($\chi^2(2) = 3.37$, $p = .185$).

3.3.3 | Behavioral measures

Accuracy

Prior to any re-coding or rejection, the overall accuracy for the gender decision task was 97% ($N = 32$), thereby higher than for the learning task. Accuracy was similar between match (98%) and mismatch trials (97%). The accuracies for each emotion category were 98% for happy, 97% for angry, and 97% for neutral trials. None of the subjects fell into the exclusion criteria (>25% incorrect trials). After the re-coding of certain systematic error patterns, the overall accuracy was 97% for both faces, independent of whether they were presented in match (97%) or mismatch trials (97%), happy (98%), angry (96%), or neutral (98%) trials during learning.

Response time

The overall mean (non-aggregated) RT for the test session was 598 ms ($SD = 119$ ms). The RT model estimation on aggregated data was conducted analogously to the learning session. RTs showed a modulation only by congruence ($\chi^2(1) = 5.06$, $p = .024$) but not by emotion ($\chi^2(2) = 2.7$, $p = .259$), nor by the interaction between both factors ($\chi^2(2) = 0.73$, $p = .694$). Gender decisions to faces previously presented with gender-matching bursts (596 ms) were answered faster than those presented with mismatching sounds (601 ms; $\beta_{mismatch} = 2.51$, $SE = 1.12$, $t = 2.23$).

Likeability rating

We ran two cumulative linked mixed models to account for the ordinal scale of the likability ratings; one model only included the associated faces, comparing gender-congruence and emotion levels, and a second for the previously

TABLE 2 Test session: Mean and SD reaction times in ms, ERP mean amplitudes in μV and Pupil size [AUs] for congruence and emotion.

	Match						Mismatch					
	Neutral		Happy		Angry		Neutral		Happy		Angry	
Measure	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
P1	5.05	3.54	5.09	3.29	4.85	3.37	5.30	3.29	5.35	3.30	5.17	3.50
P1.peak	7.26	3.91	7.46	3.53	7.18	3.56	7.64	3.51	7.73	3.58	7.43	3.82
N170	−5.86	3.95	−5.82	3.98	−5.73	3.76	−5.47	3.80	−5.69	3.96	−5.85	3.85
N170.peak	−10.65	5.32	−10.56	5.33	−10.47	5.09	−10.31	5.10	−10.49	5.24	−10.59	5.22
EPN	−0.17	3.92	−0.26	4.27	−0.11	4.16	0.33	4.21	0.13	4.31	−0.03	3.96
LPC	4.09	2.95	4.00	2.64	3.76	2.81	3.88	2.53	3.90	2.71	4.00	2.88
Pupil. early	5.04	58.53	3.94	52.73	5.82	54.67	6.94	54.33	4.04	56.83	3.28	57.30
Pupil.late	32.52	73.50	35.18	65.22	37.63	68.88	37.25	67.76	36.26	74.92	29.48	63.64
RT	596.42	56.33	593.67	58.20	598.61	56.85	603.77	62.27	598.65	56.80	601.33	59.00
Accuracy	98.41	1.34	97.80	3.94	96.71	7.71	96.29	9.70	98.32	1.85	96.85	7.78

associated compared to novel faces. In both models, random intercepts for participant ID and face stimulus were included. First, when comparing the full with reduced models via likelihood ratio tests, a model including only congruence was significant ($\chi^2(1)=4.371$, $p=.037$). However, when allowing for the congruence × emotion interaction, there were no statistically significant differences in likability ratings (all CIs of the OR included 1). Second, there was a significant difference between associated and novel faces ($\chi^2(1)=140.32$, $p\leq.001$), with novel being rated as less likeable (OR = 0.45, CI = [0.31; 0.65]). The predicted probabilities of both models for the likability ratings are shown in Figure 9. Both models' odds ratios and 95% CI can be found in Table A22 of the Supplementary Information.

4 | DISCUSSION

Previous work has documented that faces can gain additional relevance when associated with affective context information. Not only highly aversive stimuli such as loud noise bursts or electric shocks modulated face processing (for a review, see: Miskovic & Keil, 2012) but also verbal descriptions of behavior (Abdel Rahman, 2011; Baum et al., 2020; Kissler & Strehlow, 2017; Luo et al., 2016; Schindler et al., 2021; Suess et al., 2014; Xu et al., 2016), social and affective signals (Aguado et al., 2012; Bruchmann et al., 2021; Wieser, Gerdes, et al., 2014), and abstract forms of context such as monetary reward and loss (Hammerschmidt et al., 2017; Hammerschmidt, Kulke, et al., 2018). In the present ERP and pupillometry study, we applied an emotion-implicit cross-modal association paradigm, with separate learning and test sessions, to investigate whether and how robustly neutral faces might acquire additional relevance, even when the emotional quality of the US, that is, the affect burst, is not task-relevant.

Although we observed differences in the neural and behavioral responses to emotional compared to neutral face-voice pairs, the emotional relevance of the voices appears to be only partially transferred to the faces, if at all. During learning, we found emotion effects ranging from auditory processing (auditory P2) to pupillary changes and behavioral responses, presumably triggered by the affect bursts. Thus, emotional sounds were automatically processed and elicited typical responses at the neural level. However, at variance with our hypotheses, these emotion-based effects were not transferred to the *conditioned faces* (CS^+_{pos} and CS^+_{neg}), and emotion did not modulate face-locked ERPs as expected. Remarkably, faces acquired a different quality depending on the task-relevant congruent or incongruent gender information. This was present in behavior, auditory ERPs, and mid- and long-latency visual ERPs during learning. Most astonishingly, congruence effects were also

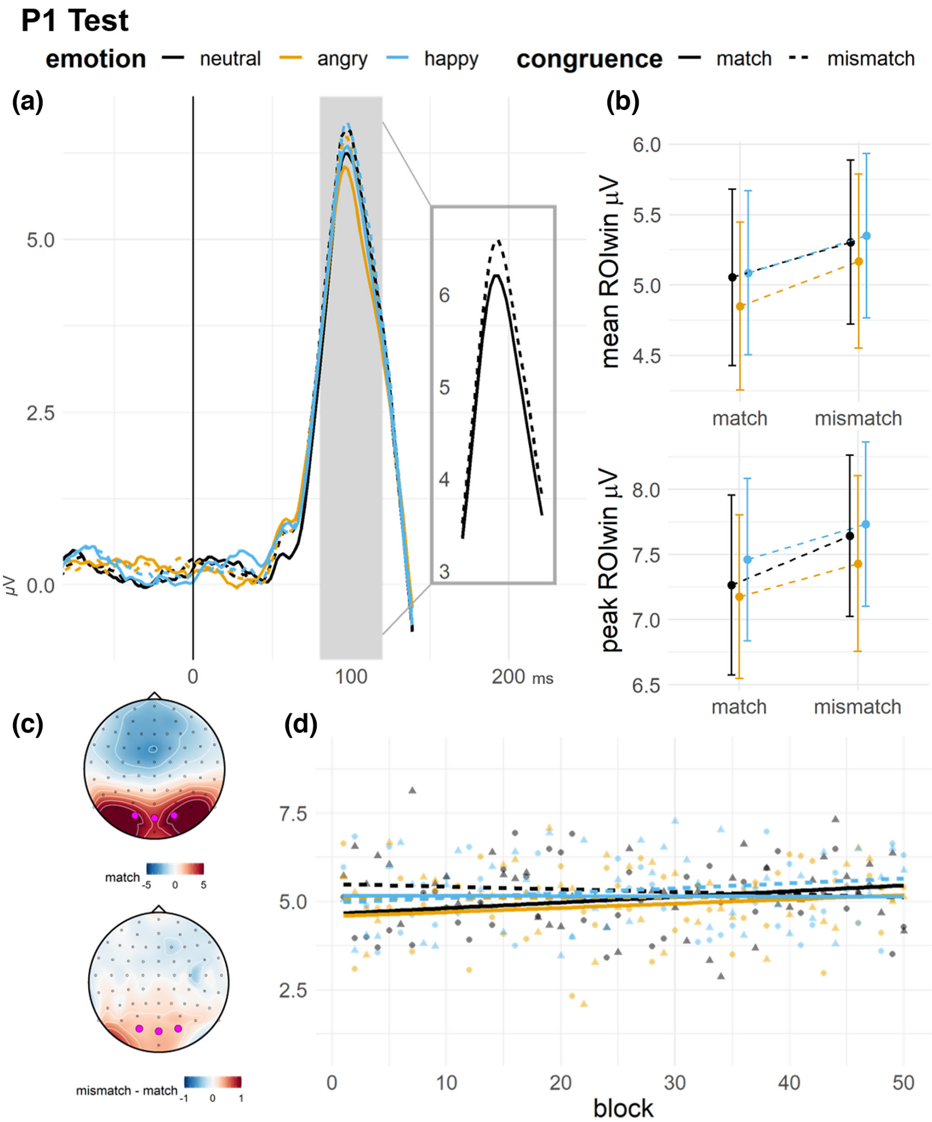

FIGURE 6 Face-locked P1 in the test session. (a) Grand average ERP of the averaged ROI channels. The highlighted area displays the ROI time window. The zoomed-in window shows the main difference in gender congruence, averaged over all emotion conditions. (b) Grand averages of the ROI mean amplitudes (upper panel) and peak amplitudes (lower panel), contrasted for all conditions. Error bars indicate ±1 SE of the mean. (c) Topographies of the ERP distribution for gender-congruent faces and the difference between gender-incongruent and congruent faces. ROI channels are highlighted in pink. (d) Mean P1 amplitudes over the course of the test session. Dots represent the grand averages per block and condition. The lines represent the fitted values of the GAMLSS model (see Supplementary Information).

observed in behavior, early and mid-latency visual ERP components during the delayed test session, indicating successful associative learning of the task-relevant stimulus features. Thus, the comparatively small modulations of the associated relevance of the emotional voices cannot be attributed to a general disregard of the (emotional) voice stimuli or to an unsuccessful face-voice integration since gender congruence was transferred from learning to testing. In the following, we will discuss the emotion-, and congruence-related effects occurring during the learning phase and the conditioned effects visible both in the learning and test session. We will use the term "face-voice pairs" instead of CS^+ and CS^- faces for pupil size, auditory ERPs, and RTs of the learning session, in which it is impossible to disentangle immediate processing (e.g., reacting to the affective voices) from effects that might have occurred because the faces gained predictive value.

4.1 | Emotion and congruence effects of face-voice pairs

As hypothesized, during acquisition, gender-matching decisions were slower for emotional compared to neutral face-voice pairs. This finding corroborates studies demonstrating that task-irrelevant emotional stimuli are more

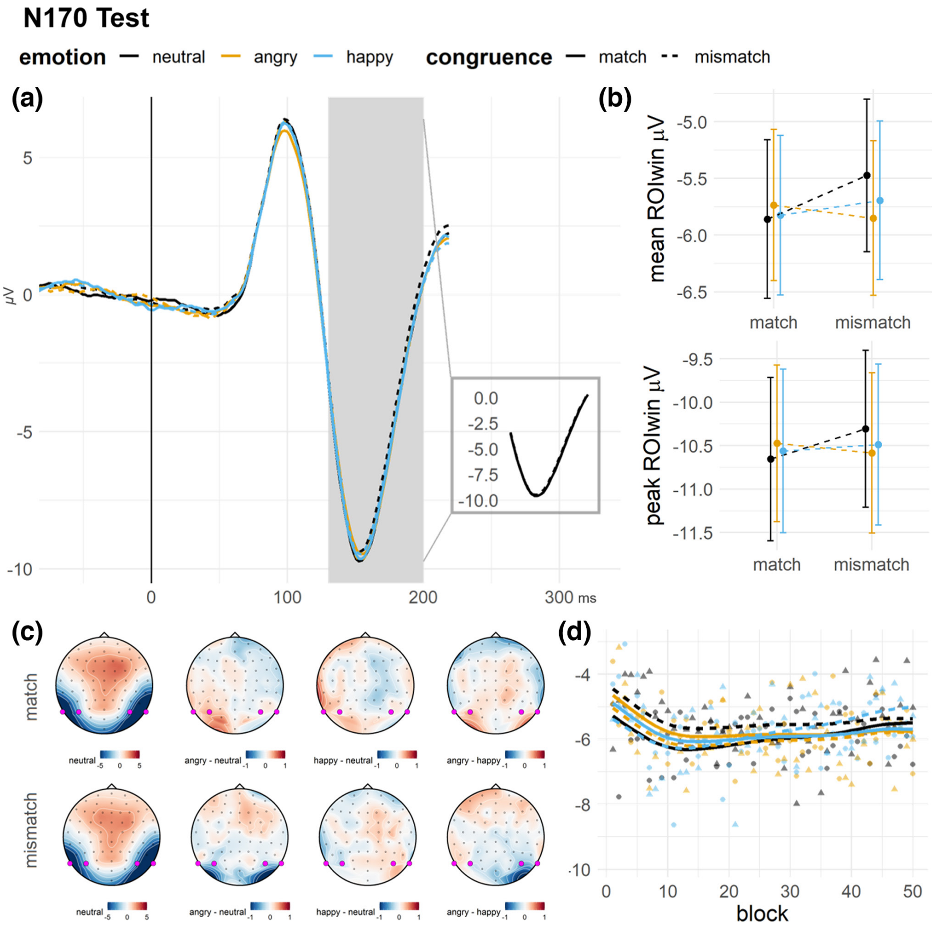

FIGURE 7 Face-locked N170 in the test session. (a) Grand average ERP time series of the averaged ROI channels. The highlighted area displays the ROI time window. The zoomed-in window shows the main difference of gender congruence, averaged over all emotion conditions. (b) Grand averages of the ROI mean amplitudes (upper panel) and peak amplitudes (lower panel), contrasted for all conditions. Error bars indicate ±1 SE of the mean. (c) Topographies of the ERP distribution for faces associated with neutral compared to emotional bursts, separately for gender-congruent and incongruent faces. ROI channels are highlighted in pink. (d) Mean N170 amplitudes over the course of the test session. Dots represent the grand averages per block and condition. The lines represent the fitted values of the GAMLSS model (see Supplementary Information).

difficult to disengage and withdraw attentional resources from the actual task (Carretié, 2014; Dresler et al., 2008; Hur, Iordan, Dolcos, & Berenbaum, 2016; Kotz & Paulmann, 2007; Schimmack, 2005; Vuilleumier, 2005). In addition, the conflicting emotional information of neutral faces and positive/negative affect bursts may have counteracted integrated person perception (de Gelder & Vroomen, 2000; Föcker et al., 2011). Presumably, this type of conflicting information is more detrimental for naturally co-occurring stimuli such as faces and voices than for abstract stimuli. Independent of the emotional category of the voice, responses were slower for gender-mismatching than -matching face-voice pairs, replicating previous findings (Huestegge et al., 2019; Latinus et al., 2010). Gender-mismatching[10] information might be more difficult to integrate due to the long-term and repeated strengthening of the associations of gender-congruent faces and voices.

To investigate at which point the emotion and congruence differences were reflected at the neural level and additionally to the pre-registered visual ERPs, we explored auditory ERPs (N1 and P2), time-locked to the voice onset. We did not find a modulation of the N1 by emotion or congruence. There was an enhancement of the auditory P2 component for happy and angry compared with neutral face-voice pairs, similar to T. Liu et al. (2012). This difference might reflect the prediction effect of the *emotional* congruence, that is, pairs of a neutral face and a neutral voice, as shown in several studies: Typically, auditory suppression effects have been found when a preceding visual stimulus predicted the occurrence of a sound (Vroomen & Stekelenburg, 2010). These suppression effects have also been reported in the context of dynamic faces and

[10]We are only referring to trials in which the participant's response was correct; hence, subjective gender mismatching corresponded to objective gender mismatching according to our stimulus database.

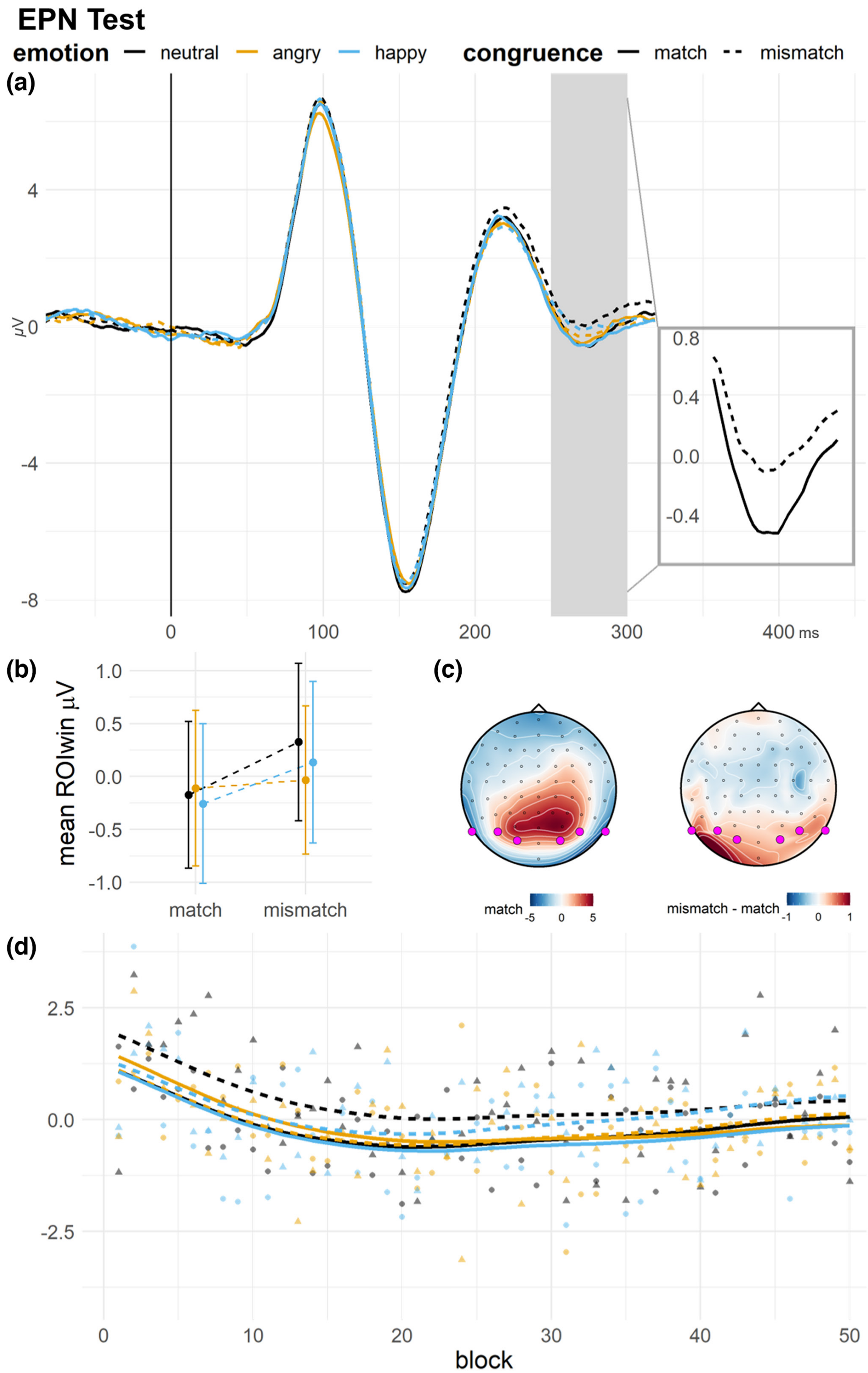

FIGURE 8 Face-locked EPN in the test session. (a) Grand average ERP of the averaged ROI channels. The highlighted area displays the ROI time window. The zoomed-in window shows the main difference of gender congruence, averaged over all emotion conditions. (b) Grand averages of the ROI amplitudes, contrasted for all conditions. Error bars indicate ±1 SE of the mean. (c) Topographies of the ERP distribution for gender-congruent faces and the difference between gender-incongruent and congruent faces. ROI channels are highlighted in pink. (d) Mean EPN amplitudes over the course of the test session. Dots represent the grand averages per block and condition. The lines represent the fitted values of the GAMLSS model (see Supplementary Information).

spoken utterances (Ho et al., 2015; van Wassenhove et al., 2005) or emotionally congruent vocalizations (Jessen & Kotz, 2013; Kokinous et al., 2014). In addition to emotion effects, also gender-congruence affected the auditory P2 component with larger amplitudes for gender-matching compared with mismatching voices.

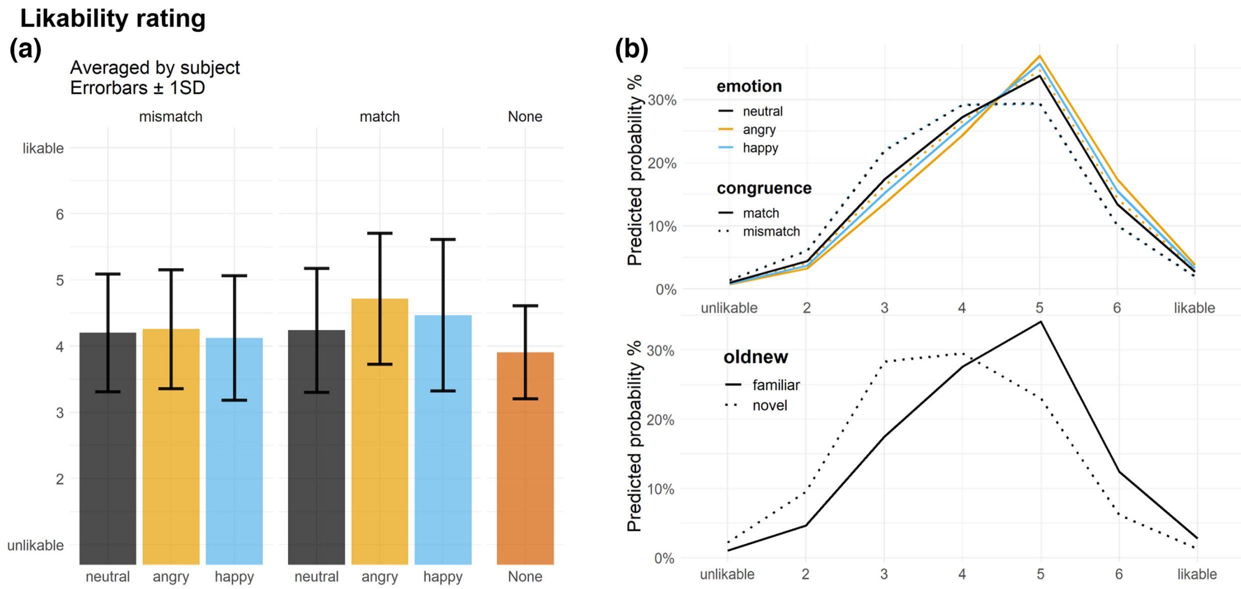

FIGURE 9 Likability rating. (a) Bar plots represent the likability ratings per condition, averaged within and across subjects. (b) Fitted values as predicted probabilities of the ordinal models. The upper panel shows the model including emotion and congruence. Please note that within the gender-mismatching condition, the dotted line for happy (blue) is mostly overlapping with the neutral (black) and hence difficult to see. The lower panel shows the collapsed familiar faces versus the novel faces.

These increased amplitudes might reflect "attention allocation costs," as described by the predictive coding framework (Feldman & Friston, 2010). With the specific task demands of *gender-congruence* decisions, feature-based attention toward *gender-congruent* voices might have led to prioritized processing across learning, also reflected by shorter RTs. The dissociation between emotional congruence and gender congruence at the level of the auditory P2 amplitudes provides novel evidence for its sensitivity to attention and predictive processing. This is consistent with studies demonstrating a sensitivity of the P2 component to various types of incongruence (Stekelenburg & Vroomen, 2007; Vroomen & Stekelenburg, 2010) and to interactive processes of prediction and attention (Schröger et al., 2015).

In our study, we expected that face-voice pairs containing affective compared with neutral bursts would elicit higher arousal and hence, larger pupil dilation, and that this response would be shifted toward the predictive face stimulus. In the time window from face onset to 1000 ms, gender-mismatching trials elicited an overall stronger pupil constriction compared with matching trials, which was most pronounced for face-voice pairs containing affective bursts. However, in the time window from 1000 to 2000 ms after face onset, covering both pupil dilation and voice presentation, pupil size differed between emotion categories. Since the US varied in duration, pupil size in the later time window might have been affected by the different auditory offsets. However, trials with neutral bursts elicited an overall larger dilation compared to affective bursts (both angry and happy, although angry bursts did not differ significantly from neutral burst regarding their duration), which was most pronounced in gender-mismatching trials. This contradicts previous findings that motivationally relevant and affective stimuli elicit larger pupil dilations (e.g., Burley et al., 2017; Finke et al., 2021; Hammerschmidt, Kagan, et al., 2018; Schindler et al., 2022). However, pupil size can also enlarge with increased cognitive load (Oliva & Anikin, 2018; for a review on pupil size correlations, see Zekveld et al., 2018). While our data suggest that both the predictability of emotional and gender congruence interacted, it remains unclear what functional interaction occurred between emotion and congruence. Notably, there was no direct correspondence between pupil size and RTs during learning. To better understand these results, we examined the habituation effects of the pupil during the learning phase across conditions. We found a typical decrease in pupil responsiveness (especially reduced dilation magnitudes) from the beginning to the end of the experiment (see Supplementary Information). However, no systematic pattern was observed with respect to emotion or congruence (e.g., no interaction between conditions and block/repetition).

4.2 | Associated effects

Neither RTs, pupil size measures, nor auditory ERPs can be treated independently of learning. Our conditioning paradigm allowed us to distinctively map the face-locked ERP modulations to different learning processes according to our experimental conditions both for the learning and the test session. Our main hypothesis was that positive and negative CS^+ faces would be processed differently than CS_{neu}. However, virtually none of the pre-registered ERP components indicated a modulation by emotion. Instead, several ERP components, pupil size, and RTs were modulated by the (previously conditioned) *gender congruence* of the face-voice pairs.

4.2.1 | Learning session

Already in the learning session, we found a difference between the gender-matching and mismatching conditions on the EPN, with an enhanced negative amplitude for matching trials. The EPN is typically associated with the attentive encoding of emotional or motivational relevance (Schupp et al., 2006). Evidence from associative learning studies is mixed, with some reports of enhanced ERP negativities for CS^+ compared to CS^- (Bruchmann et al., 2021; de Sá et al., 2018; Schellhaas et al., 2020) and others demonstrating differences between EPN effects of acquired and inherent relevance (Aguado et al., 2012; Hammerschmidt et al., 2017). Considering the functional link between the EPN and other attention-related ERP components (Schupp et al., 2006, 2007), it remains open whether the gender congruence manipulation in our study induced some kind of emotional relevance or whether the EPN effects reflect more general (emotion-independent) attentional processes.

4.2.2 | Test session

In the test session, in line with our predictions, pupil size did not show modulations by emotion or congruence. Physiological measures have been shown to extinguish fast, even when differentiation in neural measures is still persistent (e.g., Hammerschmidt, Kagan, et al., 2018; Pastor et al., 2015). Similar to the learning session, but different from what we expected, virtually all ERP modulations were related to gender congruence. These findings suggest a generalization of some processing differences to a different task (i.e., gender matching vs. gender decision). The most unexpected finding was the modulation at early processing stages, reflected in enhanced P1 amplitudes for the faces of the previously gender-incongruent condition. The P1 has mostly been associated with the processing of lower-level stimulus properties, as well as with the rapid detection of conditioned motivated or emotional salience (Aguado et al., 2012; e.g., Hammerschmidt et al., 2017; but cf. Bruchmann et al., 2020; Müller-Bardorff et al., 2016). In general, early visual processing is influenced by task demands and perceptual load (Handy et al., 2001; Pratt et al., 2011). However, the CS–US pairing was completely randomized for each participant. Therefore, even though some voices or faces might have been more difficult to extract gender information from, physical stimulus characteristics of the face cannot explain this effect. Since participants could not anticipate whether a congruent or incongruent face would be presented due to the randomized order of presentation, P1 differences are necessarily related to the processing of the stimuli. In contrast to other fear-conditioning studies with faces, which reported an enhancement of the CS^+, Sperl et al. (2021) found an increased response for CS^- faces in occipito-temporal channels at the typical P1 time window. The authors suggested that smaller P1 amplitudes for CS^+ could be caused by prolonged attention during learning and thus a smaller prediction error for CS^+. This illustrates an interesting dissociation between early (P1) and subsequent processing stages. Liu et al. (2011) analyzed P1 amplitudes as a function of stimulus repetition and found larger P1 amplitudes for CS^+ at the beginning, with a switch in favor of CS^- stimuli toward the end of the experiment. In our study, gender-incongruent face-voice pairs seemed to be processed less elaborately, as indicated by a larger prediction error with increased amplitudes of the P1 and decreased mid-latency and late ERPs, as well as slower RTs during learning. This prediction error appeared to persist even during extinction. The only pre-registered visual ERP component that was modulated by an interaction of congruence and emotion was the *N170* component during the test session, which was characterized by a smaller mean (but not peak) amplitude for the conditioned incongruent CS_{neu} faces compared to all others. However, we expected that, if at all sensitive to our experimental condition, it would show an enhancement for matching and emotionally, especially negatively conditioned faces, as found in other studies (e.g., Bruchmann et al., 2021). We cannot exclude that (neutral) faces with neutral gender-mismatching voices elicited a stronger interference for configural processing of the faces, again implying functional dissociation of the N170 to the other components, although this assumption would need to be tested by future studies. Notably, during test, the interaction effect was only significant for the N170 mean but not peak amplitudes. As the pre-registered N170 component was measured at partly overlapping electrodes as the EPN component, the mean amplitude effect during the N170 time window might already represent a mixture of configural face processing and relevance encoding (for a discussion about distinct N170 and EPN emotion effects, see Rellecke et al., 2012b). However, for the *EPN*, the interaction between conditioned emotion and congruence failed significance, whereas the difference between conditioned gender-matching and mismatching conditions became more pronounced.

The latter finding, albeit unexpected, is interesting in several respects: First, the pattern of the EPN differences between all conditions was very similar in both sessions and seemed to be robust enough over time and across different task demands. Second, the scalp topography of the congruence effect showed high conformity to typical emotion-based effects known from studies with inherent emotional salience, such as facials expression of emotion, emotional words, or complex scenes (e.g., Bayer & Schacht, 2014). The EPN modulation by conditioned gender congruence is corroborated by indicators of overt behavior, that is, faster gender decisions during the test session. Eventually, gender-matching faces acquired positive valence because they were easier to process during learning. In comparison, gender-mismatching faces did not seem to acquire strong negative valence but were probably more treated as an artificial by-product of the task. Alternatively, gender-incongruent voices might have introduced uncertainty about the gender of the face due to a reflexive integration of the face and the voice toward a whole-person concept, which was not overcome by the repetitions of the learning session. Even when the voice was no longer present in the test session, this learned uncertainty made it more difficult to decide on the gender of the face. Finally, the LPC component was neither modulated by emotion nor by congruence in the test session, probably due to the task setting (e.g., Hammerschmidt et al., 2017; Rossi et al., 2017).

4.3 | Emotional implicitness of the task

We expected that emotional vocalizations would capture increased attention due to their social and biological relevance (Johnstone & Scherer, 2000) and that the relevance would be transferred to the faces via associative learning. Looking at the acquisition period, when faces were conditioned with inherent emotional vocalizations, we found that the RTs were affected not only by the gender (in-)congruence but also by the emotion of the vocalization. Thus, emotional relevance indeed affected behavior as expected, but it did not transfer to learning to the same extent as the task-relevant gender congruence. Possible reasons are discussed in the following.

Can faces be conditioned with auditory emotional expressions? Emotional expressions of the face and voice naturally show variations within people and among situations. One could argue that faces naturally show some "resistance" to be conditioned with information that is naturally very variable and not stable. However, Aguado et al. (2012) successfully associated faces with positive and negative emotional expressions of the identically portrayed individual (same modality). Cross-modal associations have also been demonstrated in fear-conditioning studies (Miskovic & Keil, 2012), with some of them using aversive screams (Bruchmann et al., 2021; Glenn et al., 2012; Schindler et al., 2022). Effects of cross-modally associated valence through affect bursts have been shown for a valence-implicit and valence-explicit retrieval task, although participants were not instructed to attend to a specific stimulus feature during learning (Ziereis & Schacht, 2023), indicating that also less intense vocal stimuli can be successfully associated to faces.

How much attention to the emotional valence of a US is required to form stable emotion-based associations? Crucially, in many fear-conditioning paradigms, the unconditioned stimuli were not task-relevant but consisted of highly aversive stimuli. If stimuli of lower intensity were used, most task instructions aimed at making CS–US contingencies explicit or included valence and/or arousal ratings as outcome measures. In contrast, we not only did *not* instruct participants about the CS–US contingency, but we implemented a task in which exclusively other features, that is, gender information, of the stimuli were relevant. In fact, the majority of participants were not aware of the emotional CS–US contingency, as indicated by our memory checks. Still, our task ensured that they attended to both the CS and US. Attention and executive load can modulate conditioning effects also with stronger aversive stimuli such as electric shocks (Hur, Iordan, Berenbaum, & Dolcos, 2016). Despite the participants' high accuracy, the gender-matching task might have prevented emotional conditioning by being more demanding than other, valence-irrelevant task designs. Nevertheless, learning about the face-voice associations did occur, as indicated by the long-lasting and robust gender congruence effects on ERP measures. Therefore, it seems plausible that the attention toward gender information might have suppressed associative learning of emotion, although emotional vocalizations were processed differently from neutral sounds.

4.4 | Limitations and future directions

First, we only tested the explicit CS–US contingency related to emotion. To rule out that the missing effect was caused by the lack of CS–US contingency awareness, future studies should examine whether the gender congruence is explicitly retrievable and whether this explains differences between congruence and emotion effects. Second, based on previous research, we expected intermediate effect sizes for associated valence. However, the overall associated valence effects in our study were rather small (and small to intermediate for the congruence effects). A larger sample size would be needed to test the robustness of the small(er) valence effects if they are considered to be practically relevant. Several aspects can be identified that may have reduced overall learning in our study, independent of specific experimental conditions: Due to a technical imprecision, the onset of the voice stimulus jittered (±9 ms) around the offset of the face stimulus—unsystematically across stimuli or conditions. Thus, some trials could be considered as delay conditioning trials and others as trace conditioning trials, the latter of which may have reduced the overall learning rate. There is also the possibility that the mere presence of gender-mismatched face-voice pairs, which may have been perceived as artificial, could have altered the overall processing of emotion in the voice, that is, for both matching and mismatching face-voice pairs. A modified learning paradigm with no (or fewer) incongruent face-voice pairs may eliminate this possibility. More generally, we used a paradigm in which emotion was neither task-relevant during learning nor during testing. Therefore, the specific role of feature-based attention and task relevance during conditioning and retrieval should be investigated in a systematic cross-over setting, possibly with multiple stimulus features besides gender and emotion, in order to generalize the findings. Investigating the role of feature-based attention in the context of fear-conditioning and extinction may ultimately contribute to the improvement of therapeutic interventions (e.g., exposure therapy) in the context of clinical research.

5 | CONCLUSION

We implemented an associative learning paradigm to investigate whether neutral faces automatically acquire emotional relevance when associated with cross-modal emotion from the voice, while emotion was not task-relevant. Emotion effects were limited to auditory ERPs, pupil size, and RTs in the acquisition period, possibly being immediately elicited by the emotional burst. In contrast, the task-relevant gender congruence of the face-voice pairs impacted virtually all measures during acquisition. More strikingly, however, it modulated neural (P1, N170, and EPN) and behavioral responses to previously conditioned faces during test on the following day and in a different task. Our results suggest that, despite successful face-voice integration and the effects of emotion on the processing of affective voice stimuli, associative learning of emotional relevance is not guaranteed.

AUTHOR CONTRIBUTIONS

Annika Ziereis: Conceptualization; data curation; formal analysis; visualization; writing – original draft; writing – review and editing. **Anne Schacht:** Conceptualization; supervision; writing – review and editing.

ACKNOWLEDGMENTS

This work was supported by Deutsche Forschungsgemeinschaft, Grant/Award Number: 254142454 /GRK 2070 and Project-ID 454648639/SFB 1528, and the Leibniz-Gemeinschaft by supporting the Leibniz ScienceCampus Primate Cognition (W45/2019). We thank Jantje Becker for her help with recruitment and data collection. The code and data of this study are available upon request from the corresponding author, [AZ]. The data sets are not publicly available for privacy reasons (no consent from participants to publish the raw data). Parts of the data of this study were presented at conferences and scientific meetings. Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The conditions of our ethics approval do not permit public archiving of study data. The entire data and stimulus sets will be made available to interested researchers following completion of a data sharing agreement and approval by the local ethics committee.

ORCID

Annika Ziereis https://orcid.org/0000-0002-3789-2941
Anne Schacht https://orcid.org/0000-0001-9471-3842

REFERENCES

Abdel Rahman, R. (2011). Facing good and evil: Early brain signatures of affective biographical knowledge in face recognition. *Emotion*, *11*(6), 1397–1405. https://doi.org/10.1037/a0024717

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*(1), 29–41. https://doi.org/10.1027/0269-8803/a000065

Anderson, A. K. (2005). Affective influences on the attentional dynamics supporting awareness. *Journal of Experimental Psychology: General*, *134*(2), 258–281. https://doi.org/10.1037/0096-3445.134.2.258

Armony, J. (2002). Modulation of spatial attention by fear-conditioned stimuli: An event-related fMRI study. *Neuropsychologia*, *40*(7), 817–826. https://doi.org/10.1016/s0028-3932(01)00178-6

Aviezer, H., Bentin, S., Dudarev, V., & Hassin, R. R. (2011). The automaticity of emotional face-context integration. *Emotion*, *11*(6), 1406–1414. https://doi.org/10.1037/a0023578

Bach, D. R., Grandjean, D., Sander, D., Herdener, M., Strik, W. K., & Seifritz, E. (2008). The effect of appraisal level on processing of emotional prosody in meaningless speech. *NeuroImage*, *42*(2), 919–927. https://doi.org/10.1016/j.neuroimage.2008.05.034

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Baum, J., Rabovsky, M., Rose, S. B., & Rahman, R. A. (2020). Clear judgments based on unclear evidence: Person evaluation is strongly influenced by untrustworthy gossip. *Emotion*, *20*(2), 248–260. https://doi.org/10.1037/emo0000545

Bayer, M., & Schacht, A. (2014). Event-related brain responses to emotional words, pictures, and faces a cross-domain comparison. *Frontiers in Psychology*, *5*, 1106. https://doi.org/10.3389/fpsyg.2014.01106

Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The Montreal affective voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods*, *40*(2), 531–539. https://doi.org/10.3758/brm.40.2.531

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

Bieniek, M. M., Frei, L. S., & Rousselet, G. A. (2013). Early ERPs to faces: Aging, luminance, and individual differences. *Frontiers in Psychology*, *4*, 268. https://doi.org/10.3389/fpsyg.2013.00268

Bobak, P., Bodis-Wollner, I., & Guillory, S. (1987). The effect of blur and contrast of VEP latency: Comparison between check and sinusoidal grating patterns. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *68*(4), 247–255. https://doi.org/10.1016/0168-5597(87)90045-1

Boersma, P., & Weenink, D. (2018). *Praat: Doing phonetics by computer* (Computer program). Version 6.0.37. http://www.praat.org/

Breyer, B., & Bluemke, M. (2016). Deutsche version der positive and negative affect schedule PANAS (GESIS panel). In *Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS)*. ZIS – GESIS Leibniz Institute for the Social Sciences. https://doi.org/10.6102/ZIS242

Bruchmann, M., Schindler, S., Heinemann, J., Moeck, R., & Straube, T. (2021). Increased early and late neuronal responses to aversively conditioned faces across different attentional conditions. *Cortex*, *142*, 332–341. https://doi.org/10.1016/j.cortex.2021.07.003

Bruchmann, M., Schindler, S., & Straube, T. (2020). The spatial frequency spectrum of fearful faces modulates early and mid-latency ERPs but not the N170. *Psychophysiology*, *57*(9), e13597. https://doi.org/10.1111/psyp.13597

Bublatzky, F., Gerdes, A. B. M., White, A. J., Riemer, M., & Alpers, G. W. (2014). 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

Bublatzky, F., Kavcıoğlu, F., Guerra, P., Doll, S., & Junghoefer, M. (2020). Contextual information resolves uncertainty about ambiguous facial emotions: Behavioral and magnetoencephalographic correlates. *NeuroImage*, *215*, 116814. https://doi.org/10.1016/j.neuroimage.2020.116814

Burley, D. T., Gray, N. S., & Snowden, R. J. (2017). As far as the eye can see: Relationship between psychopathic traits and pupil response to affective stimuli. *PLoS One*, *12*(1), 1–22. https://doi.org/10.1371/journal.pone.0167436

Buysse, D. J., Reynolds, C. F., Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research. *Psychiatry Research*, *28*(2), 193–213. https://doi.org/10.1016/0165-1781(89)90047-4

Camfield, D. A., Mills, J., Kornfeld, E. J., & Croft, R. J. (2016). Modulation of the N170 with classical conditioning: The use of emotional imagery and acoustic startle in healthy and depressed participants. *Frontiers in Human Neuroscience*, *10*, 337. https://doi.org/10.3389/fnhum.2016.00337

Carretié, L. (2014). Exogenous (automatic) attention to emotional stimuli: A review. *Cognitive, Affective, & Behavioral Neuroscience*, *14*(4), 1228–1258. https://doi.org/10.3758/s13415-014-0270-2

Chen, S., Tan, Z., Xia, W., Gomes, C. A., Zhang, X., Zhou, W., Liang, S., Axmacher, N., & Wang, L. (2021). Theta oscillations synchronize human medial prefrontal cortex and amygdala during fear learning. *Science Advances*, *7*(34), eabf4198. https://doi.org/10.1126/sciadv.abf4198

Cosme, G., Rosa, P. J., Lima, C. F., Tavares, V., Scott, S., Chen, S., Wilcockson, T. D. W., Crawford, T. J., & Prata, D. (2021). Pupil dilation reflects the authenticity of received nonverbal vocalizations. *Scientific Reports*, *11*(1), 3733. https://doi.org/10.1038/s41598-021-83070-x

Dal Ben, R. (2019). *SHINE color and Lum_fun: A set of tools to control luminance of colorful images (Version 0.2)* [Computer program]. Open Science Framework. https://doi.org/10.17605/OSF.IO/AUZJY

Dalmaijer, E. S., Mathôt, S., & der Stigchel, S. V. (2013). PyGaze: An open-source, cross-platform toolbox for minimal-effort programming of eyetracking experiments. *Behavior Research Methods*, *46*(4), 913–921. https://doi.org/10.3758/s13428-013-0422-2

de Gelder, B., & Vroomen, J. (2000). The perception of emotion by ear and by eye. *Cognition and Emotion*, *14*(3), 289–311. https://doi.org/10.1080/026999300378824

de Sá, D. S. F., Michael, T., Wilhelm, F. H., & Peyk, P. (2018). Learning to see the threat: Temporal dynamics of ERPs of motivated attention in fear conditioning. *Social Cognitive and Affective Neuroscience*, *14*(2), 189–203. https://doi.org/10.1093/scan/nsy103

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

Dresler, T., Mériau, K., Heekeren, H. R., & Meer, E. (2008). Emotional Stroop task: Effect of word arousal and subject anxiety on emotional interference. *Psychological Research Psychologische Forschung*, *73*(3), 364–371. https://doi.org/10.1007/s00426-008-0154-6

Eilers, P. H. C., & Marx, B. D. (2010). Splines, knots, and penalties. *WIREs Computational Statistics*, *2*(6), 637–653. https://doi.org/10.1002/wics.125

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers*, *28*(1), 1–11. https://doi.org/10.3758/BF03203630

Ethofer, T., Anders, S., Wiethoff, S., Erb, M., Herbert, C., Saur, R., Grodd, W., & Wildgruber, D. (2006). Effects of prosodic emotional intensity on activation of associative auditory cortex. *Neuroreport*, *17*(3), 249–253. https://doi.org/10.1097/01.wnr.0000199466.32036.5d

Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, *4*, 215. https://doi.org/10.3389/fnhum.2010.00215

Finke, J. B., Roesmann, K., Stalder, T., & Klucken, T. (2021). Pupil dilation as an index of Pavlovian conditioning. A systematic review and meta-analysis. *Neuroscience and Biobehavioral Reviews*, *130*, 351–368. https://doi.org/10.1016/j.neubiorev.2021.09.005

Föcker, J., Gondan, M., & Röder, B. (2011). Preattentive processing of audio-visual emotional signals. *Acta Psychologica*, *137*(1), 36–47. https://doi.org/10.1016/j.actpsy.2011.02.004

Foti, D., Olvet, D. M., Klein, D. N., & Hajcak, G. (2010). Reduced electrocortical response to threatening faces in major depressive disorder. *Depression and Anxiety*, *27*(9), 813–820. https://doi.org/10.1002/da.20712

Freeman, J. B., & Ambady, N. (2011). When two become one: Temporally dynamic integration of the face and voice. *Journal of Experimental Social Psychology*, *47*(1), 259–263. https://doi.org/10.1016/j.jesp.2010.08.018

Glenn, C. R., Lieberman, L., & Hajcak, G. (2012). Comparing electric shock and a fearful screaming face as unconditioned stimuli for fear learning. *International Journal of Psychophysiology*, *86*(3), 214–219. https://doi.org/10.1016/j.ijpsycho.2012.09.006

Gutiérrez-Cobo, M. J., Luque, D., Most, S. B., Fernández-Berrocal, P., & Pelley, M. E. L. (2019). Reward and emotion influence attentional bias in rapid serial visual presentation. *Quarterly Journal of Experimental Psychology*, *72*(9), 2155–2167. https://doi.org/10.1177/1747021819840615

Hammerschmidt, W., Kagan, I., Kulke, L., & Schacht, A. (2018). Implicit reward associations impact face processing: Time-resolved evidence from event-related brain potentials and pupil dilations. *NeuroImage*, *179*, 557–569. https://doi.org/10.1016/j.neuroimage.2018.06.055

Hammerschmidt, W., Kulke, L., Broering, C., & Schacht, A. (2018). Money or smiles: Independent ERP effects of associated monetary reward and happy faces. *PLoS One*, *13*(10), e0206142. https://doi.org/10.1371/journal.pone.0206142

Hammerschmidt, W., Sennhenn-Reulen, H., & Schacht, A. (2017). Associated motivational salience impacts early sensory processing of human faces. *NeuroImage*, *156*, 466–474. https://doi.org/10.1016/j.neuroimage.2017.04.032

Handy, T. C., Soltani, M., & Mangun, G. R. (2001). Perceptual load and visuocortical processing: Event-related potentials reveal sensory-level selection. *Psychological Science*, *12*(3), 213–218. https://doi.org/10.1111/1467-9280.00338

Hassin, R. R., Aviezer, H., & Bentin, S. (2013). Inherently ambiguous: Facial expressions of emotions, in context. *Emotion Review*, *5*(1), 60–65. https://doi.org/10.1177/1754073912451331

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*(6), 223–233. https://doi.org/10.1016/s1364-6613(00)01482-0

Hillyard, S. A., & Anllo-Vento, L. (1998). Event-related brain potentials in the study of visual selective attention. *Proceedings of the National Academy of Sciences*, *95*(3), 781–787. https://doi.org/10.1073/pnas.95.3.781

Ho, H. T., Schröger, E., & Kotz, S. A. (2015). Selective attention modulates early human evoked potentials during emotional face-voice processing. *Journal of Cognitive Neuroscience*, *27*(4), 798–818. https://doi.org/10.1162/jocn_a_00734

Hubert, M., & Vandervieren, E. (2008). An adjusted boxplot for skewed distributions. *Computational Statistics and Data Analysis*, *52*(12), 5186–5201. https://doi.org/10.1016/j.csda.2007.11.008

Huestegge, S. M., Raettig, T., & Huestegge, L. (2019). Are face-incongruent voices harder to process? *Experimental Psychology*, *66*(2), 154–164. https://doi.org/10.1027/1618-3169/a000440

Hur, J., Iordan, A. D., Berenbaum, H., & Dolcos, F. (2016). Emotion attention interactions in fear conditioning: Moderation by executive load, neuroticism, and awareness. *Biological Psychology*, *121*, 213–220. https://doi.org/10.1016/j.biopsycho.2015.10.007

Hur, J., Iordan, A. D., Dolcos, F., & Berenbaum, H. (2016). Emotional influences on perception and working memory. *Cognition and Emotion*, *31*(6), 1294–1302. https://doi.org/10.1080/02699931.2016.1213703

Jessen, S., & Kotz, S. A. (2013). On the role of crossmodal prediction in audiovisual emotion perception. *Frontiers in Human Neuroscience*, *7*(7), 1–7. https://doi.org/10.3389/fnhum.2013.00369

Johnson, J. A., & Zatorre, R. J. (2006). Neural substrates for dividing and focusing attention between simultaneous auditory and visual events. *NeuroImage*, *31*(4), 1673–1681. https://doi.org/10.1016/j.neuroimage.2006.02.026

Johnstone, T., & Scherer, K. (2000). In M. Lewis & J. Haviland (Eds.), *The handbook of emotion*. Guilford.

Kissler, J., & Strehlow, J. (2017). Something always sticks? How emotional language modulates neural processes involved in face encoding and recognition memory. *Poznan Studies in Contemporary Linguistics*, *53*(1), 63–93. https://doi.org/10.1515/psicl-2017-0004

Kokinous, J., Kotz, S. A., Tavano, A., & Schröger, E. (2014). The role of emotion in dynamic audiovisual integration of faces and voices. *Social Cognitive and Affective Neuroscience*, *10*(5), 713–720. https://doi.org/10.1093/scan/nsu105

Kornmeier, J., Pfaffle, M., & Bach, M. (2011). Necker cube: Stimulus-related (low-level) and percept-related (high-level) EEG signatures early in occipital cortex. *Journal of Vision*, *11*(9), 12. https://doi.org/10.1167/11.9.12

Kotz, S. A., & Paulmann, S. (2007). When emotional prosody and semantics dance cheek to cheek: ERP evidence. *Brain Research*, *1151*, 107–118. https://doi.org/10.1016/j.brainres.2007.03.015

Kret, M. E., Roelofs, K., Stekelenburg, J. J., & de Gelder, B. (2013). Emotional signals from faces, bodies and scenes influence observers face expressions, fixations and pupil-size. *Frontiers in Human Neuroscience*, *7*, 810. https://doi.org/10.3389/fnhum.2013.00810

Kret, M. E., & Sjak-Shie, E. E. (2018). Preprocessing pupil size data: Guidelines and code. *Behavior Research Methods*, *51*(3), 1336–1342. https://doi.org/10.3758/s13428-018-1075-y

Kulke, L., Janßen, L., Demel, R., & Schacht, A. (2017). *Validating the Goettingen faces database*. Open Science Framework. https://doi.org/10.17605/OSF.IO/4KNPF

Latinus, M., VanRullen, R., & Taylor, M. J. (2010). Top-down and bottom-up modulation in processing bimodal face/voice stimuli. *BMC Neuroscience*, *11*(1), 36. https://doi.org/10.1186/1471-2202-11-36

Lausen, A., & Schacht, A. (2018). Gender differences in the recognition of vocal emotions. *Frontiers in Psychology*, *9*, 882. https://doi.org/10.3389/fpsyg.2018.00882

Lenth, R. (2020). *Emmeans: Estimated marginal means, aka least-squares means*. Version 2.4.2. https://CRAN.R-project.org/package=emmeans

Liu, T., Pinheiro, A. P., Deng, G., Nestor, P. G., McCarley, R. W., & Niznikiewicz, M. A. (2012). Electrophysiological insights into processing nonverbal emotional vocalizations. *Neuroreport*, *23*(2), 108–112. https://doi.org/10.1097/wnr.0b013e32834ea757

Liu, Y., Keil, A., & Ding, M. (2011). Effects of emotional conditioning on early visual processing: Temporal dynamics revealed by ERP single-trial analysis. *Human Brain Mapping*, *33*(4), 909–919. https://doi.org/10.1002/hbm.21259

Luo, Q. L., Wang, H. L., Dzhelyova, M., Huang, P., & Mo, L. (2016). Effect of affective personality information on face processing: Evidence from ERPs. *Frontiers in Psychology*, *7*, 810. https://doi.org/10.3389/fpsyg.2016.00810

Maechler, M., Rousseeuw, P., Croux, C., Todorov, V., Ruckstuhl, A., Salibian-Barrera, M., Verbeke, T., Koller, M., Conceicao, E. L. T., & Anna di Palma, M. (2021). *robustbase: Basic robust statistics*. Version 0.93–8. http://robustbase.r-forge.r-project.org/

MATLAB. (2018). *Version 9.4.0.949201(R2018a)*. The MathWorks Inc.

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

Menz, M. M., Rihm, J. S., & Büchel, C. (2016). REM sleep is causal to successful consolidation of dangerous and safety stimuli and reduces return of fear after extinction. *The Journal of Neuroscience*, *36*(7), 2148–2160. https://doi.org/10.1523/jneurosci.3083-15.2016

Mertens, G., & Engelhard, I. M. (2020). A systematic review and meta-analysis of the evidence for unaware fear conditioning. *Neuroscience and Biobehavioral Reviews*, *108*, 254–268. https://doi.org/10.1016/j.neubiorev.2019.11.012

Miskovic, V., & Keil, A. (2012). Acquired fears reflected in cortical sensory processing: A review of electrophysiological studies of human classical conditioning. *Psychophysiology*, *49*(9), 1230–1241. https://doi.org/10.1111/j.1469-8986.2012.01398.x

Morel, S., Beaucousin, V., Perrin, M., & George, N. (2012). Very early modulation of brain responses to neutral faces by a single prior association with an emotional context: Evidence from MEG. *NeuroImage*, *61*(4), 1461–1470. https://doi.org/10.1016/j.neuroimage.2012.04.016

Morris, J. S., Öhman, A., & Dolan, R. J. (1998). Conscious and unconscious emotional learning in the human amygdala. *Nature*, *393*(6684), 467–470. https://doi.org/10.1038/30976

Mueller, E. M., & Pizzagalli, D. A. (2015). One-year-old fear memories rapidly activate human fusiform gyrus. *Social Cognitive and Affective Neuroscience*, *11*(2), 308–316. https://doi.org/10.1093/scan/nsv122

Mullen, T. (2012). *NITRC: CleanLine: Tool/resource info*. Version 1.04. http://www.nitrc.org/projects/cleanlin

Müller-Bardorff, M., Schulz, C., Peterburs, J., Bruchmann, M., Mothes-Lasch, M., Miltner, W., & Straube, T. (2016). Effects of emotional intensity under perceptual load: An event-related potentials (ERPs) study. *Biological Psychology*, *117*, 141–149. https://doi.org/10.1016/j.biopsycho.2016.03.006

Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: Detecting the snake in the grass. *Journal of Experimental Psychology: General*, *130*(3), 466–478. https://doi.org/10.1037/0096-3445.130.3.466

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Oliva, M., & Anikin, A. (2018). Pupil dilation reflects the time course of emotion recognition in human vocalizations. *Scientific Reports*, *8*(1), 4871. https://doi.org/10.1038/s41598-018-23265-x

Pace-Schott, E. F., Germain, A., & Milad, M. R. (2015). Effects of sleep on memory for conditioned fear and fear extinction. *Psychological Bulletin*, *141*(4), 835–857. https://doi.org/10.1037/bul0000014

Panitz, C., Hermann, C., & Mueller, E. M. (2015). Conditioned and extinguished fear modulate functional corticocardiac coupling in humans. *Psychophysiology*, *52*(10), 1351–1360. https://doi.org/10.1111/psyp.12498

Paquette, S., Rigoulot, S., Grunewald, K., & Lehmann, A. (2020). Temporal decoding of vocal and musical emotions: Same code, different timecourse? *Brain Research*, *1741*, 146887. https://doi.org/10.1016/j.brainres.2020.146887

Pastor, M. C., Rehbein, M. A., Junghöfer, M., Poy, R., López, R., & Moltó, J. (2015). Facing challenges in differential classical conditioning research: Benefits of a hybrid Design for Simultaneous Electrodermal and Electroencephalographic Recording. *Frontiers in Human Neuroscience*, *9*, 336. https://doi.org/10.3389/fnhum.2015.00336

Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, *2*, 10. https://doi.org/10.3389/neuro.11.010.2008

Pell, M. D., Rothermich, K., Liu, P., Paulmann, S., Sethi, S., & Rigoulot, S. (2015). Preferential decoding of emotion from human non-linguistic vocalizations versus speech prosody. *Biological Psychology*, *111*, 14–25. https://doi.org/10.1016/j.biopsycho.2015.08.008

Pessoa, L., Kastner, S., & Ungerleider, L. G. (2002). Attentional control of the processing of neutral and emotional stimuli. *Cognitive Brain Research*, *15*(1), 31–45. https://doi.org/10.1016/s0926-6410(02)00214-8

Pessoa, L., McKenna, M., Gutierrez, E., & Ungerleider, L. G. (2002). Neural processing of emotional faces requires attention. *Proceedings of the National Academy of Sciences*, *99*(17), 11458–11463. https://doi.org/10.1073/pnas.172403899

Pion-Tonachini, L., Makeig, S., & Kreutz-Delgado, K. (2017). Crowd labeling latent Dirichlet allocation. *Knowledge and Information Systems*, *53*(3), 749–765. https://doi.org/10.1007/s10115-017-1053-1

Pivik, R. T., Broughton, R. J., Coppola, R., Davidson, R. J., Fox, N., & Nuwer, M. R. (1993). Guidelines for the recording and quantitative analysis of electroencephalographic activity in research contexts. *Psychophysiology*, *30*(6), 547–558. https://doi.org/10.1111/j.1469-8986.1993.tb02081.x

Pratt, N., Willoughby, A., & Swick, D. (2011). Effects of working memory load on visual selective attention: Behavioral and electrophysiological evidence. *Frontiers in Human Neuroscience*, *5*, 57. https://doi.org/10.3389/fnhum.2011.00057

Quadflieg, S., Mohr, A., Mentzel, H.-J., Miltner, W. H. R., & Straube, T. (2008). Modulation of the neural network involved in the processing of anger prosody: The role of task-relevance and social phobia. *Biological Psychology*, *78*(2), 129–137. https://doi.org/10.1016/j.biopsycho.2008.01.014

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing https://www.R-project.org/

Rehbein, M. A., Pastor, M. C., Moltó, J., Poy, R., López-Penadés, R., & Junghöfer, M. (2018). Identity and expression processing during classical conditioning with faces. *Psychophysiology*, *55*(10), e13203. https://doi.org/10.1111/psyp.13203

Rehbein, M. A., Steinberg, C., Wessing, I., Pastor, M. C., Zwitserlood, P., Keuper, K., & Junghöfer, M. (2014). Rapid plasticity in the prefrontal cortex during affective associative learning. *PLoS One*, *9*(10), e110720. https://doi.org/10.1371/journal.pone.0110720

Rellecke, J., Palazova, M., Sommer, W., & Schacht, A. (2011). On the automaticity of emotion processing in words and faces: Event-related brain potentials evidence from a superficial task. *Brain and Cognition*, *77*(1), 23–32. https://doi.org/10.1016/j.bandc.2011.07.001

Rellecke, J., Sommer, W., & Schacht, A. (2012a). Does processing of emotional facial expressions depend on intention? Time-resolved evidence from event-related brain potentials. *Biological Psychology*, *90*(1), 23–32. https://doi.org/10.1016/j.biopsycho.2012.02.002

Rellecke, J., Sommer, W., & Schacht, A. (2012b). Emotion effects on the N170: A question of reference? *Brain Topography*, *26*(1), 62–71. https://doi.org/10.1007/s10548-012-0261-y

Rigby, R. A., & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape (with discussion). *Applied Statistics*, *54*, 507–554. https://doi.org/10.1111/j.1467-9876.2005.00510.x

Roesch, E. B., Sander, D., Mumenthaler, C., Kerzel, D., & Scherer, K. R. (2010). Psychophysics of emotion: The QUEST for emotional attention. *Journal of Vision*, *10*(3), 1–9. https://doi.org/10.1167/10.3.4

Rossi, V., Vanlessen, N., Bayer, M., Grass, A., Pourtois, G., & Schacht, A. (2017). Motivational salience modulates early visual cortex responses across task sets. *Journal of Cognitive Neuroscience*, *29*(6), 968–979. https://doi.org/10.1162/jocn_a_01093

Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., Linotte, S., & Crommelinck, M. (2000). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects an electrophysiological account of face-specific processes in the human brain. *Neuroreport*, *11*(1), 69–72. https://doi.org/10.1097/00001756-200001170-00014

Russo, F. D. (2003). Source analysis of event-related cortical activity during Visuo-spatial attention. *Cerebral Cortex*, *13*(5), 486–499. https://doi.org/10.1093/cercor/13.5.486

Sander, D., Grandjean, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). Emotion and attention interactions in social cognition: Brain regions involved in processing anger prosody. *NeuroImage*, *28*(4), 848–858. https://doi.org/10.1016/j.neuroimage.2005.06.023

Schacht, A., Adler, N., Chen, P., Guo, T., & Sommer, W. (2012). Association with positive outcome induces early effects in event-related brain potentials. *Biological Psychology*, *89*(1), 130–136. https://doi.org/10.1016/j.biopsycho.2011.10.001

Schacht, A., & Sommer, W. (2009). Emotions in word and face processing: Early and late cortical responses. *Brain and Cognition*, *69*(3), 538–550. https://doi.org/10.1016/j.bandc.2008.11.005

Schellhaas, S., Arnold, N., Schmahl, C., & Bublatzky, F. (2020). Contextual source information modulates neural face processing in the absence of conscious recognition: A threat-of-shock study. *Neurobiology of Learning and Memory*, *174*, 107280. https://doi.org/10.1016/j.nlm.2020.107280

Schimmack, U. (2005). Attentional interference effects of emotional pictures: Threat, negativity, or arousal? *Emotion*, *5*(1), 55–66. https://doi.org/10.1037/1528-3542.5.1.55

Schindler, S., Bruchmann, M., Krasowski, C., Moeck, R., & Straube, T. (2021). Charged with a crime: The neuronal signature of processing negatively evaluated faces under different attentional conditions. *Psychological Science*, *32*, 1311–1324. https://doi.org/10.1177/0956797621996667

Schindler, S., & Bublatzky, F. (2020). Attention and emotion: An integrative review of emotional face processing as a function of attention. *Cortex*, *130*, 362–386. https://doi.org/10.1016/j.cortex.2020.06.010

Schindler, S., Heinemann, J., Bruchmann, M., Moeck, R., & Straube, T. (2022). No trait anxiety influences on early and late differential neuronal responses to aversively conditioned faces across three different tasks. *Cognitive, Affective, & Behavioral Neuroscience*, *22*, 1157–1171. https://doi.org/10.3758/s13415-022-00998-x

Schröger, E., Marzecová, A., & SanMiguel, I. (2015). Attention and prediction in human audition: A lesson from cognitive psychophysiology. *European Journal of Neuroscience*, *41*(5), 641–664. https://doi.org/10.1111/ejn.12816

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

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), 189–200. https://doi.org/10.1037/1528-3542.4.2.189

Schupp, H. T., Stockburger, J., Codispoti, M., Junghofer, M., Weike, A. I., & Hamm, A. O. (2007). Selective visual attention to emotion. *Journal of Neuroscience*, *27*(5), 1082–1089. https://doi.org/10.1523/JNEUROSCI.3223-06.2007

Schwarz, K. A., Wieser, M. J., Gerdes, A. B. M., Muehlberger, A., & Pauli, P. (2012). Why are you looking like that? How the context influences evaluation and processing of human faces. *Social Cognitive and Affective Neuroscience*, *8*(4), 438–445. https://doi.org/10.1093/scan/nss013

Sharot, T., & Phelps, E. A. (2004). How arousal modulates memory: Disentangling the effects of attention and retention. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(3), 294–306. https://doi.org/10.3758/CABN.4.3.294

Shinners, P. (2011). *PyGame*. http://pygame.org/

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2020). *Afex: Analysis of factorial experiments*. https://CRAN.R-project.org/package=afex

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Sage.

Sopp, M. R., Michael, T., Weeß, H.-G., & Mecklinger, A. (2017). Remembering specific features of emotional events across time: The role of REM sleep and prefrontal theta oscillations. *Cognitive, Affective, & Behavioral Neuroscience*, *17*(6), 1186–1209. https://doi.org/10.3758/s13415-017-0542-8

Sperl, M. F. J., Wroblewski, A., Mueller, M., Straube, B., & Mueller, E. M. (2021). Learning dynamics of electrophysiological brain signals during human fear conditioning. *NeuroImage*, *226*, 117569. https://doi.org/10.1016/j.neuroimage.2020.117569

Stangier, U., Heidenreich, T., Berardi, A., Golbs, U., & Hoyer, J. (1999). Die Erfassung sozialer Phobie durch die Social Interaction Anxiety Scale (SIAS) und die Social Phobia Scale (SPS). *Zeitschrift für Klinische Psychologie Und Psychotherapie*, *28*(1), 28–36. https://doi.org/10.1026//0084-5345.28.1.28

Steinberg, C., Bröckelmann, A.-K., Dobel, C., Elling, L., Zwanzger, P., Pantev, C., & Junghöfer, M. (2013). Preferential responses to extinguished face stimuli are preserved in frontal and occipito-temporal cortex at initial but not later stages of processing. *Psychophysiology*, *50*(3), 230–239. https://doi.org/10.1111/psyp.12005

Steinberg, C., Bröckelmann, A.-K., Rehbein, M., Dobel, C., & Junghöfer, M. (2013). Rapid and highly resolving associative affective learning: Convergent electro- and magnetoencephalographic evidence from vision and audition. *Biological Psychology*, *92*(3), 526–540. https://doi.org/10.1016/j.biopsycho.2012.02.009

Steinberg, C., Dobel, C., Schupp, H. T., Kissler, J., Elling, L., Pantev, C., & Junghöfer, M. (2012). Rapid and highly resolving: Affective evaluation of Olfactorily conditioned faces. *Journal of Cognitive Neuroscience*, *24*(1), 17–27. https://doi.org/10.1162/jocn_a_00067

Stekelenburg, J. J., & Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *Journal of Cognitive Neuroscience*, *19*(12), 1964–1973. https://doi.org/10.1162/jocn.2007.19.12.1964

Suess, F., Rabovsky, M., & Rahman, R. A. (2014). Perceiving emotions in neutral faces: Expression processing is biased by affective person knowledge. *Social Cognitive and Affective Neuroscience*, *10*(4), 531–536. https://doi.org/10.1093/scan/nsu088

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences*, *102*(4), 1181–1186. https://doi.org/10.1073/pnas.0408949102

Verosky, S. C., Porter, J., Martinez, J. E., & Todorov, A. (2018). Robust effects of affective person learning on evaluation of faces. *Journal of Personality and Social Psychology*, *114*(4), 516–528. https://doi.org/10.1037/pspa0000109

Vroomen, J., & Stekelenburg, J. J. (2010). Visual anticipatory information modulates multisensory interactions of artificial audiovisual stimuli. *Journal of Cognitive Neuroscience*, *22*(7), 1583–1596. https://doi.org/10.1162/jocn.2009.21308

Vuilleumier, P. (2005). How brains beware: Neural mechanisms of emotional attention. *Trends in Cognitive Sciences*, *9*(12), 585–594. https://doi.org/10.1016/j.tics.2005.10.011

Watters, A. J., Rupert, P. E., Wolf, D. H., Calkins, M. E., Gur, R. C., Gur, R. E., & Turetsky, B. I. (2018). Social aversive conditioning in youth at clinical high risk for psychosis and with psychosis: An ERP study. *Schizophrenia Research*, *202*, 291–296. https://doi.org/10.1016/j.schres.2018.06.027

Wiemer, J., Leimeister, F., & Pauli, P. (2021). Subsequent memory effects on event-related potentials in associative fear learning. *Social Cognitive and Affective Neuroscience*, *16*(5), 525–536. https://doi.org/10.1093/scan/nsab015

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

Wieser, M. J., Gerdes, A. B. M., 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

Wieser, M. J., Miskovic, V., Rausch, S., & Keil, A. (2014). Different time course of visuocortical signal changes to fear-conditioned faces with direct or averted gaze: A ssVEP study with single-trial analysis. *Neuropsychologia*, *62*, 101–110. https://doi.org/10.1016/j.neuropsychologia.2014.07.009

World Medical Association. (1964). Human experimentation: Code of ethics of the world medical association. *The British Medical Journal*, *2*(5402), 177. http://www.jstor.org/stable/25399522

Xu, M., Li, Z., Diao, L., Fan, L., & Yang, D. (2016). Contextual valence and sociality jointly influence the early and later stages of neutral face processing. *Frontiers in Psychology*, *7*, 1258. https://doi.org/10.3389/fpsyg.2016.01258

Yiannikas, C., & Walsh, J. C. (1983). The variation of the pattern shift visual evoked response with the size of the stimulus field. *Electroencephalography and Clinical Neurophysiology*, *55*(4), 427–435. https://doi.org/10.1016/0013-4694(83)90131-1

Yoon, K. L., & Zinbarg, R. E. (2008). Interpreting neutral faces as threatening is a default mode for socially anxious individuals. *Journal of Abnormal Psychology*, *117*(3), 680–685. https://doi.org/10.1037/0021-843x.117.3.680

Zekveld, A. A., Koelewijn, T., & Kramer, S. E. (2018). The pupil dilation response to auditory stimuli: Current state of knowledge. *Trends in Hearing*, *22*, 233121651877717. https://doi.org/10.1177/2331216518777174

Zhang, X., Li, Q., Sun, S., & Zuo, B. (2019). Facial expressions can inhibit the activation of gender stereotypes. *Cognition and Emotion*, *33*(7), 1424–1435. https://doi.org/10.1080/02699931.2019.1586648

Ziereis, A., & Schacht, A. (2023). Motivated attention and task relevance in the processing of cross-modally associated faces: Behavioral and electrophysiological evidence. *Cognitive, Affective, and Behavioral Neuroscience*, *23*(3). https://doi.org/10.3758/s13415-023-01112-5

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Appendix S1. Supplementary material

How to cite this article: Ziereis, A., & Schacht, A. (2023). Gender congruence and emotion effects in cross-modal associative learning: Insights from ERPs and pupillary responses. *Psychophysiology*, *00*, e14380. https://doi.org/10.1111/psyp.14380