Motivated attention and task relevance in the processing of cross-modally associated faces: Behavioral and electrophysiological evidence

Annika Ziereis[1] · **Anne Schacht**[1]

Abstract

It has repeatedly been shown that visually presented stimuli can gain additional relevance by their association with affective stimuli. Studies have shown effects of associated affect in event-related potentials (ERP) like the early posterior negativity (EPN), late positive complex (LPC), and even earlier components as the P1 or N170. However, findings are mixed as to the extent associated affect requires directed attention to the emotional quality of a stimulus and which ERP components are sensitive to task instructions during retrieval. In this preregistered study (https://osf.io/ts4pb), we tested cross-modal associations of vocal affect-bursts (positive, negative, neutral) to faces displaying neutral expressions in a flash-card-like learning task, in which participants studied face-voice pairs and learned to correctly assign them to each other. In the subsequent EEG test session, we applied both an implicit ("old-new") and explicit ("valence-classification") task to investigate whether the behavior at retrieval and neurophysiological activation of the affect-based associations were dependent on the type of motivated attention. We collected behavioral and neurophysiological data from 40 participants who reached the preregistered learning criterium. Results showed EPN effects of associated negative valence after learning and independent of the task. In contrast, modulations of later stages (LPC) by positive and negative associated valence were restricted to the explicit, i.e., valence-classification, task. These findings highlight the importance of the task at different processing stages and show that cross-modal affect can successfully be associated to faces.

Keywords Associative learning · Faces · ERPs · Attention · Cross-modal

The human brain navigates the complexities of our everyday social lives very efficiently, for example, by quickly extracting various information from other people's faces (Haxby et al., 2000). Research has repeatedly shown that what we know about a person and what is relevant to us impacts how we perceive that person (Bublatzky et al., 2014; Davis et al., 2009; Heisz & Shedden, 2009; Wieser & Brosch, 2012). This includes, but is not limited to, biographical information and relevant experiences with that person. In the laboratory, relevance often is manipulated through associations with

valence-laden stimuli and actions, ranging from receiving monetary (Hammerschmidt et al., 2017, 2018a, b) or social reward and punishment (Aguado et al., 2012; Wieser et al., 2014) to highly aversive stimuli such as loud noise bursts (Watters et al., 2018) or electric shocks (Rehbein et al., 2014). It has been repeatedly shown that various types of affective stimuli impact face processing promptly (Wieser & Brosch, 2012) and through learned associations (Miskovic & Keil, 2012).

Although the term *attention* is not clearly defined in the literature, there is consensus that certain stimuli are processed preferentially over others, because they are physically salient, they resemble targets that match our current goals, or because we have learned their relevance through past experience. Especially experience-driven attention (Anderson et al., 2021) aims to explain phenomena, such as impaired performance in the presence of learned aversive distractors (Öhman et al., 2001; Vuilleumier, 2005) or self-referential cues as described in the cocktail-party effect

✉ Annika Ziereis
annika.ziereis@uni-goettingen.de

Anne Schacht
schacht@psych.uni-goettingen.de

1   Department for Cognition, Emotion and Behavior,
Affective Neuroscience and Psychophysiology Laboratory,
Georg-August-University of Göttingen, Goßlerstraße 14,
37073 Göttingen, Germany

(Röer & Cowan, 2021). To date, there is more evidence for conditioning effects with threat-related stimuli. However, also appetitive cues have been shown to be associated to different types of stimuli, e.g., faces, objects, or abstract stimuli, such as meaningless words (Aguado et al., 2012; Blechert et al., 2016; Davis et al., 2009; Hammerschmidt et al., 2017, 2018a, b; Rossi et al., 2017; Steinberg et al., 2013; Ventura-Bort et al., 2016). Although recognizing and reacting to both appetitive and aversive environmental cues appears adaptive, the need to detect and respond quickly is greater in a threatening environment (Öhman et al., 2001). Avoiding predictable and unpleasant situations also may be preferred over detecting potentially pleasant ones (Gottfried et al., 2002).

Neurophysiological research allows to investigate processes beyond overt behavior and has demonstrated that some acquired associations with affective stimuli elicit differential neural responses, even when a conditioned behavioral or physiological response has extinguished (Antov et al., 2020; Apergis-Schoute et al., 2014). However, in the case of absent effects in behavioral and neural measures, it remains open whether the information was learned at all, or whether it was not apparent under the specific test condition.

The overarching goal of the present study was to investigate how directed, experience-driven attention through task requirements impacts face perception at different levels (i.e., early/automatic vs. later/elaborate processing). More specifically, we tested whether the retrieval of valence-implicit or valence-explicit features moderates the neurophysiological and behavioral response to valence-based associations in faces.

Several affect-sensitive ERPs have been related to different stages of the processing of associated faces: The P1 typically peaks around 100 ms after face onset with an occipital, bilateral positivity. It is generated by the extrastriate cortex (Hillyard & Anllo-Vento, 1998; Russo, 2003) and has been reported to be enhanced for faces associated with affect-laden or valent stimuli, such as monetary reward (Hammerschmidt et al., 2017), emotional expressions of the associated face (Aguado et al., 2012), and threatening stimuli, although some fear-conditioning studies reported even earlier effects (Steinberg et al., 2013). More reliably than for the P1, associated and conditioned effects have been reported for the N170 and subsequent components. The N170 is a face-sensitive neural marker in the form of a negative deflection peaking around 170 ms over occipito-temporal regions, generated to a large extent by the fusiform face area (Gao et al., 2019). N170 effects of conditioned faces have been reported by a number of studies on fear-conditioning (Bruchmann et al., 2021; Camfield et al., 2016; Schellhaas et al., 2020; Sperl et al., 2021) as well as by studies on associated person knowledge (Luo et al., 2016; Schindler et al., 2021) and on conditioned facial expressions (Aguado et al.,

2012). Modulations of the early posterior negativity (EPN), a relative negativity over occipito-temporal regions related to the early detection of emotional relevance and most pronounced around 200-300 ms for face stimuli, also have been reported for conditioned faces with different types of unconditioned stimuli (US), e.g., in conditioned fear (Bruchmann et al., 2021; de Sá et al., 2018; Schellhaas et al., 2020), and verbal descriptions about a person (Luo et al., 2016; Suess et al., 2014; Xu et al., 2016) and produced by a person (Wieser et al., 2014). Sustained motivated attention has been related to the late positive complex (LPC). Effects on the LPC, a centro-parietal positivity, have been reported for faces associated with different contexts in fear-conditioning (Bruchmann et al., 2021; Panitz et al., 2015; Rehbein et al., 2018; de Sá et al., 2018; Sperl et al., 2021; Wiemer et al., 2021), reward (Hammerschmidt et al., 2018b), and person-knowledge studies (Abdel Rahman, 2011; Baum et al., 2020; Kissler & Strehlow, 2017; Xu et al., 2016). Whereas there are several studies on cross-modal perception that include faces and affective voices (de Gelder & Vroomen, 2000; Pell et al., 2022), to our knowledge, the processing of faces, which have previously been associated with both positive and negative affect bursts, has not yet been tested.

The relevance of information to a specific situational task or context has been shown to play an important role in both learning and in retrieval (Shin et al., 2020). In learning, task-relevant (and thus context-congruent) information is supposed to be more easily integrated into a preactivated schema (van Kesteren et al., 2010). However, the relationship between task-relevance and context for retrieval is not straightforward, considering reports of generalized effects across different tasks. Similar to the processing of faces with emotional expressions (Hudson et al., 2021; Rellecke et al., 2012a; Schindler et al., 2020; Valdés-Conroy et al., 2014) and affective stimuli in general (Olofsson & Polich, 2008), task requirements are likely to moderate conditioned effects especially at later stages of processing (Schupp et al., 2006). That early and late effects have been reported in (not necessarily the same) fear-conditioning studies may be primarily caused by the use of intense and highly arousing stimuli. Additionally, most fear-conditioning studies have implemented valence and arousal ratings of the conditioned stimuli (CS faces) before and after the conditioning phase (Panitz et al., 2015; Rehbein et al., 2018; de Sá et al., 2018; Sperl et al., 2021), which might influence the attentional processes in other tasks, i.e., during learning and retrieval. Previously target-defining features of a stimulus have been reported to automatically withdraw processing resources even when they are no longer task-relevant (Kyllingsbæk et al., 2014). Nevertheless, conditioned effects have been reported for valence-unrelated tasks, such as old-new categorizations of faces (e.g., early effects: Hammerschmidt et al., 2017; late effects: Abdel Rahman, 2011; Baum et al.,

2020; Kissler & Strehlow, 2017) and passive-viewing tasks (Xu et al., 2016).

Only a few studies systematically have investigated the role of attention on the perception of faces associated with context information. Three recent studies tested the effects of feature- and memory-based attention with different tasks, which included a) discrimination of lines that overlayed the faces, b) the faces' gender (Bruchmann et al., 2021; Schindler et al., 2022) or age (Schindler et al., 2021), and c) the associated CS category. In their threat-conditioning study, Schindler et al. (2022) reported interactions between task and conditioning for the P1 and the EPN, but not for the N170 and LPC components and hence show no clear distinction between task influences on early and later processing. In contrast, associated verbal descriptions of crime-related actions differentially moderated early and late processing in Schindler et al. (2021), of which the N170 was enhanced in all tasks for negatively associated faces, whereas both associated effects on the EPN and LPC were reported only for the valence-focused condition. In these studies, associated context information was also presented during the test, either interspersed (in 33% of trials in Bruchmann et al., 2021; Schindler et al., 2022) or at the beginning of each task (Schindler et al., 2021). In experimental studies, researchers have often used intense and highly aversive stimuli to maximize the differentiability between conditions. While this is valid and important to demonstrate a general potential of associating context with faces, it neglects the real and true diversity of affective contexts, especially positive associations. Moreover, it has not been particularly well researched whether associating affective stimuli of lower intensity and whose contextual relevance has not been made explicit for learning also elicit robust effects in face perception. Expressions in faces and voices naturally co-occur and construct a perception of the whole person (Freeman & Ambady, 2011). Emotional expressions in both modalities are situation-dependent and naturally vary within individuals; conversely, facial and vocal expressions share inherent social and biological relevance (Straube et al., 2011). Both factors might impact the effectiveness of using these stimuli in associative learning, making them a compelling research topic.

Goal of the study

To address this gap, we investigated the role of attentional focus in the retrieval of faces associated with cross-modal affect. To do so, we used a valence-implicit (old-new) and valence-explicit (valence-classification of the associated voices) task in a delayed test session to investigate associated valence under different attention conditions while recording face-sensitive ERPs. For learning, we paired faces displaying neutral expressions with short auditory affect bursts of positive (elation and amusement), negative (anger and disgust), and neutral (yawning and throat-clearing) valence, because they rapidly unfold emotional information and do not have the segmental structure of speech. In our newly developed Internet-based learning phase, unlike in classical (Pavlovian) or instrumental learning paradigms, our participants studied the face-voice pairs to correctly assign them to each other (similar to learning with flashcards). In addition, we did not provide any further information about the task requirements of the test session to not prompt participants to pay attention to specific stimulus features.

Hypotheses

Our global hypothesis was that task requirements during the test would activate (goal-directed) memory-based attention to the associated face-voice pairs, which in turn would modulate the processing of the faces. More specifically, we expected the differential effects of task on different processing stages of valence-based associations, with early processing being less impacted than later, more elaborate processing (according to Rellecke et al., 2012a). In this sense, goal-directed attention through the task and experience-based attention through the relevance association would produce additive effects on visually-evoked potentials.

Learning

In our online learning hub, participants could study the face-voice pairs flexibly and according to their own schedule. As a result, we expected high variability in the individual learning styles and in the time it took to reach the predefined learning criterion (95% correct in 24 subsequent test trials), and analyzed the learning data only in an exploratory way.

Test: Behavioral hypotheses

We expected an effect of task difficulty with slower responses and lower accuracy for the valence-classification task (3-choice responses) compared to the old-new-task (2-choice responses). Furthermore, we expected an interaction effect of task × valence with larger RT and accuracy differences between the affectively and neutrally associated faces for the valence-classification task compared with the old-new task, as the latter required only superficial recognition of the faces. Regarding the valence effect in the valence-classification task, we expected higher accuracy for affectively compared to neutrally associated faces. Furthermore, faces previously associated with voices expressing an emotion of positive valence (i.e., elation and amusement) should be rated as more likable than faces associated with neutral bursts, and analogously, bursts of negative emotions

(i.e., anger and disgust) should be rated as less likable compared to neutral bursts (similar to but not as pronounced as in Suess et al., 2014).

Test: ERP hypotheses

We expected visually evoked potentials related to face perception to be modulated by the emotional valence of the associated voices, such that individual expressions of emotion would cluster according to their valence (e.g., amusement and elation for positive valence). Additionally, we expected modulatory effects of task, and interaction effects of task × valence, especially on the mid- (EPN) and long- (LPC) latency ERPs. However, for comparability with other studies and due to inconsistent findings on the influence of goal-directed attention (via task demands), we tested the interaction (task × valence) in all of our models. We predicted larger (mean and peak) amplitudes of the P1 for affectively (i.e., positively or negatively) associated faces than for neutrally associated faces and similar effects in both tasks. Although we expected valence-based modulations of the N170 and EPN, we did not specify the direction of the effects, as effects of associated valence have been inconsistently reported for these components. In contrast to the P1 and N170, we expected that EPN differences between affectively and neutrally associated faces would be more pronounced in the valence-classification task. The EPN is suggested to reflect enhanced sensory encoding of valence-laden stimuli (independent of the task). However, it is unknown whether recently conditioned faces would also produce an EPN component modulated by the associated valence in a superficial task like the old-new task (e.g., Rellecke et al., 2012a for reduced emotion effects on the EPN for facial expressions in superficial tasks). Finally, we did not expect effects of associated valence on the LPC due to unreported effects in similar studies. However, in the case of effects of associated valence, we expected them to be exclusive to the valence-classification task.

Method and materials

This study was preregistered on https://osf.io/ts4pb.

Participants

Of the 61 participants, who signed up for the study, 54 started learning, and 43 completed the EEG test session, of which our target sample size of 40 participants met the required number of trials (min. 30 valid EEG trials per valence condition and task after artifact rejection). Our sample consisted mainly of students (38 of 40; 30 females, 10 males, 0 diverse; age: 18-32, $M = 21.62$ years) who

reported normal or corrected-to-normal vision (max. ±1 diopter), normal hearing, and no neurological or psychiatric disorders. All participants were right-handed, according to Oldfield (1971), and proficient in German. We recruited via campus advertisements and postings on social media, the university's job portal, the website of the Institute for Psychology, and the department's recruitment database. Participants were reimbursed with a fixed amount of money for completing the online learning phase and an additional hourly rate for the test session in the laboratory, or an equivalent amount of course credits.

Stimuli

Twenty-four faces were selected from the Goettingen Faces Database (Kulke et al., 2017) and presented in their natural color on a light gray background. The face stimuli were edited and combined with a transparency mask that covered the hairline, ears, and neck. In the test session, they had a visual angle of approx. 3.2×4.8 degrees and a resolution of 200×300 pixels. The mean luminance (HSV) of the images ranged from 0.45 to 0.48 ($M = .47$; $\chi^2(528) = 550$, $p = .246$) (Dal Ben, 2019). Affect bursts (happy, elated, angry, and disgusted) were selected from a validated database (Cowen et al., 2019) and supplemented with neutral vocalizations (clearing throat, yawning) from the social media platform "youtube." All silent periods at the beginning and the end of the sound files were manually trimmed and normalized to −23 LUCS (Loudness Units Full Scale) using the open-source software Audacity® (v.2.4.2, Audacity Team, 2021). The perceived loudness of the audio files was normalized based on an algorithm following the EBU R 128 recommendation (https://tech.ebu.ch/docs/r/r128.pdf) for limiting the loudness of audio signals. Compared to other normalization methods (peak and RMS normalization), this method resulted in the smallest range of estimated loudness across stimuli using the R package Soundgen (Anikin, 2019). There were two separate stimulus sets of faces. One set included the 12 CS$^+$ faces used in the learning and test phases, and the other set contained 12 new faces, which were used in the old-new task and the rating task of the test session. The assignment of the CS$^+$ faces to the US voices was counterbalanced and matched for gender. Each emotion category (positive: amusement and elation, neutral: yawning and throat-clearing, negative: disgust and anger) of the voices entailed two stimuli in our experiment (one female and one male). Hence, each valence category contained four stimuli (four positive, four negative, and four neutral), resulting in a total of 12 face-voice pairs included in the learning phase. There were six different versions of the learning set for the face-voice pairs. Participants were pseudo-randomly assigned to one of the six versions to ensure a balanced distribution of stimulus-set versions.

Procedure

The study was conducted in accordance with the Declaration of Helsinki and approved by the local ethics committee of the Institute of Psychology at the University of Göttingen. Before participation, interested participants visited a website that informed them of the complete procedure, inclusion criteria, data policy, corona regulations, and remuneration for the online learning phase and the EEG session. They were redirected to a form to provide contact and socio-demographic information if they consented. We scheduled EEG sessions with eligible participants and created personalized links and participant codes for the learning platform (learning hub). The link was activated six days before the scheduled EEG session. To participate in the EEG testing in the laboratory, participants had to achieve a learning criterion (95% correct out of 24 test trials) during one of the learning sessions and, independently of that, complete obligatory learning checks during the first four days. Participants were free to choose the length and number of learning sessions, repetitions, and learning checks within the learning phase. An overview of the learning and test procedure is shown in Fig. 1.

Learning phase (online)

The online experiment was programmed in JavaScript with self-written and existing functions from the open-source library jsPsych (v6.3.1, de Leeuw, 2014). The experiment was integrated into JATOS (v3.5.8, Lange et al., 2015) on a local server installation at the University of Goettingen for data management. Participants could start the learning sessions with a personalized link and a participation code. Instructions were given compulsorily at the beginning and optionally displayed for later sessions. When participants logged in for the first time, they rated the valence of the auditory stimuli: On two sliders (without initial thumb), they were asked to rate 1) how positive vs. negative the mood of the speaker expressing the vocalization was, and 2) how pleasant vs. unpleasant they found the auditory stimulus. Then, regardless of their ratings, we individually presented all auditory stimuli with emotion labels to set an anchor for stimuli that may have been ambiguous for the participant. Subsequently, participants were redirected to the learning hub, where they could start their first learning session.

During the learning (association) phase, participants aimed to learn the pairing of the 12 face-voice stimuli (i.e., which face belonged to which voice) within 4 days by using the learning hub. The main page consisted of 12 preview cards showing blurred versions of the 12 faces. Clicking on one of the blurred cards started a conditioning trial with a central fixation cross, followed by the unblurred CS^+ face and the auditory US starting with the face offset. The

number of conditioning trials per CS-US pair was recorded for each session and in total. To assess whether they were able to allocate the face-voice pairs, participants were required to complete at least one obligatory learning check (including 24 test trials, approx. 2 minutes) per day. A learning check trial consisted of a pseudo-randomly selected US voice, played while participants had to select the correct face out of five gender-matching faces. Immediate feedback on the correctness of their response was provided. The learning criterion was met if 95% of the last 24 trials within a session were answered correctly. To prevent early and late learners from having different time delays between learning and test session, we required daily learning checks independent of the learning criterium. On top of the learning deck, information about the number of repetitions (conditioning trials) per session and in total, as well as the accuracy in learning checks for the session and the last 24 learning checks, was displayed. The order of the preview cards was shuffled at the beginning of each new session. For learning checks, a list of all face-voice pairs was shuffled, and the number of test trials was sampled from this list without replacement. The order of faces to choose from in the learning checks was also randomized. Participants could cancel their participation in the online study at any time and request the deletion of their data. Once participation was canceled, it was impossible to resume or restore the data or to participate in the EEG test session.

Questionnaire about learning strategies: One day before the test session, participants completed a questionnaire about their strategies for learning the face-voice pairs. Participants who reached the learning criterium were informed in more detail about the procedure and the safety regulations and were asked for their confirmation of the test session.

Test session

After giving written consent, participants were prepared for the EEG session and seated in a dimly lit, electrically shielded room in front of a computer screen at a distance of approximately 78 cm. Two loudspeakers were placed to the left and right of the monitor. Participants positioned their chins in a height-adjustable chin rest. For the presentation of the laboratory experiment, along with standard Python (2.7) libraries, such as numpy and scipy, we used functions of PsychoPy (Peirce, 2009) for the presentation of the faces, PyGame (Shinners, 2011) as the audio library, and PyGaze (Dalmaijer et al., 2013) for the communication with the eyetracker. After a welcoming message, the eye tracker was calibrated with a 9-point calibration. For all participants, the test session had the same order: Refresher trials I (5×12 trials), Old-new task (25×18 trials), refresher trials II (5×12 trials), valence-classification task (25×12 trials), and a likability rating (24 trials). To not reveal that any emotion- or
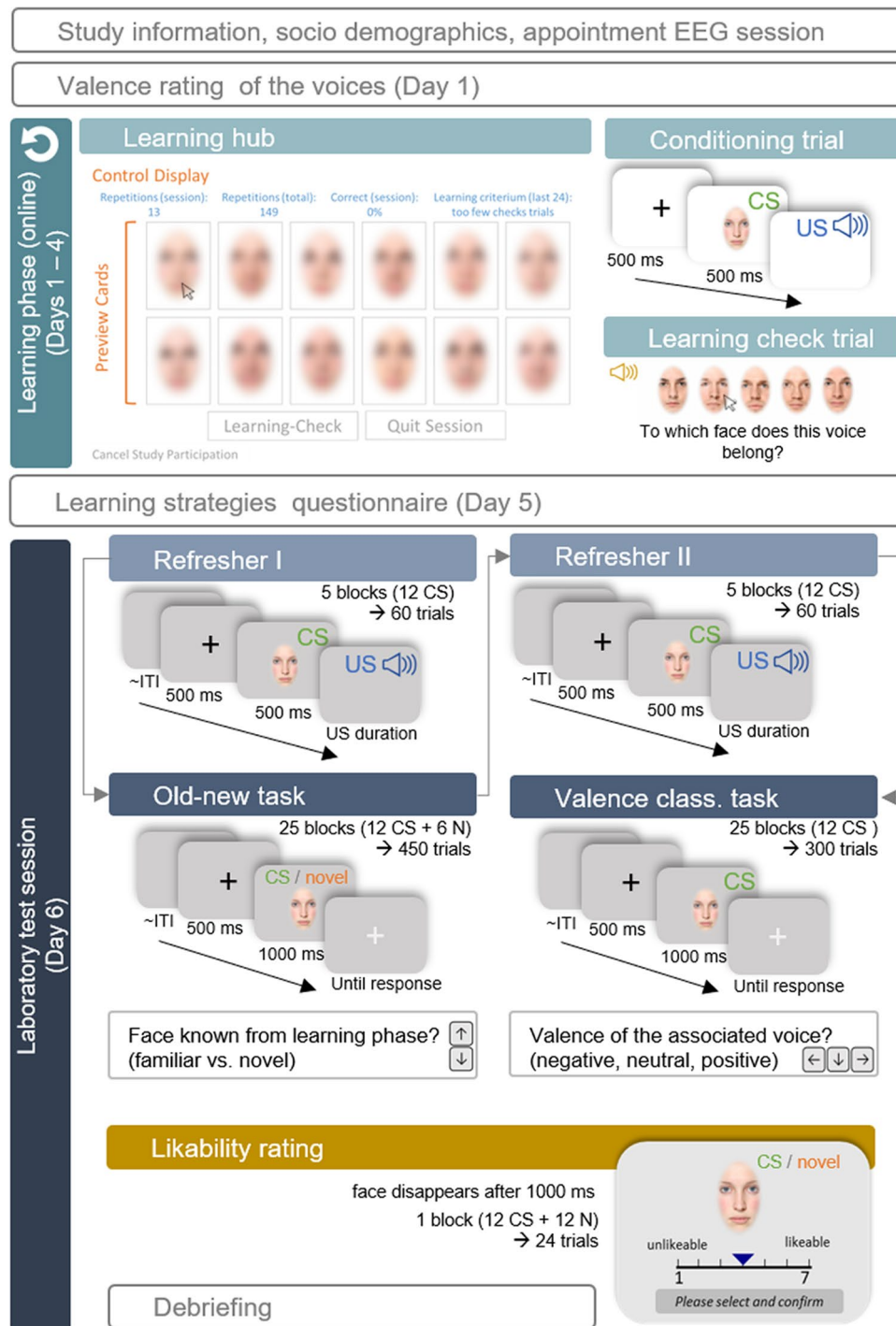
Fig. 1 Procedure of the online learning phase and test session

valence-relevant task would be part of the experiment, the specific task instructions were shown before each task. Refresher trials were passive-viewing trials in which participants did not have to respond. However, we instructed them to focus on the face-voice pairs to "refresh" what they had learned. For the other tasks, specific instructions were followed by four example trials using face-like shapes and the correct answer as a label on top to familiarize with the response keys. Only here did participants receive feedback on whether they were correct and had the possibility to clarify the remaining questions with the experimenter. Breaks for stretching and relaxing were scheduled between tasks,

and there were four additional breaks within the old-new task and three within the valence-classification task. A drift correction of the eye-tracker (1-point calibration) was implemented to resume or start the next task.

In all tasks, the order of faces or face-sound combinations was shuffled at the beginning of each block, with each block consisting of a single set of associated faces (or all 12 associated faces plus six randomly selected faces from the set of novel faces). Assignments of response keys were counterbalanced. Participants were instructed to answer as accurately and fast as possible, and to guess if unsure. Refresher trials started with a black fixation cross in the center of the screen for 500 ms, which was replaced by one of the CS^+ faces displayed for 500 ms. With the offset of the face, the US set in (duration varied between stimuli). After a jittered inter-trial interval ($M = 2,800$ ms, $SD = 200$ ms), the subsequent trial began. In the **old/new task**, the participant's task was to decide whether a face was known from the online learning phase or a novel set of faces. A black fixation cross was displayed for 500 ms in the center of the screen replaced by either a CS^+ face from the learning phase or a CS^- face (novel). All faces were presented individually for 1000 ms and participants could respond as soon as the face stimulus set in. With the offset of the face, a gray fixation cross was displayed if no answer had been registered yet and continued until an answer was given via keypress. After the face-offset and a registered response, the next trial started after an additional jittered inter-trial interval ($M = 1,800$ ms, $SD = 200$ ms). In the **valence-classification task**, participants had to recall the valence category (negative, positive, neutral) of the associated voices, with only the CS^+ faces presented. The presentation duration of the trial elements (fixation cross, face, response fixation cross, inter-trial interval) was identical to the old-new task. At the end of the test session, participants rated the likability of the CS^+ and the novel CS^- faces. The faces were presented individually for 1,000 ms and were rated on a Likert scale in the appearance of a 1-7 slider positioned below the faces. There was no value or choice shown by default. A value was selected by clicking on the slider with the mouse, but had to be confirmed to start the next trial. At the very end of the session, participants were informed about the main aims and background of the study (presented on the computer screen) and could clarify any questions with the experimenter.

Collected data

For each learning session, the number of conditioning trials for each individual face-voice pair and the accuracy of the learning check trials were recorded. In the test session, in addition to performance (RT and accuracy), we recorded EEG and pupil size (see Supplementary Information) during the refresher, old-new, and valence-classification tasks.

No neurophysiological measures were collected during the likability rating.

EEG recording and preprocessing

The continuous EEG was recorded with a sampling rate of 512 Hz (bandwidth: 102.4 Hz) at 64 active electrodes (AgAgCl) mounted in an electrode cap (Easy Cap™). The arrangement was based on the extended 10-20 system (Pivik et al., 1993). Additionally, two external electrodes were used: one each for the left and right mastoids. Reference electrodes were the common mode sense (CMS) active electrode and, as ground electrode, the driven right leg (DLR) passive electrode. The scalp voltage signals were amplified by a BiosemiActiveTwo AD-Box and recorded with the software ActiView. The data were preprocessed offline in MATLAB (2018) with functions of the toolbox EEGLAB (2019.9, Delorme & Makeig, 2004). To account for a systematic delay that was measured with a photodiode, event markers were shifted by a constant of 26 ms. The continuous data was re-referenced to average reference (excl. external electrodes), filtered with a 0.01 Hz second-order Butterworth filter, and the remaining 50-Hz line-noise was corrected with a function of the plugin "CleanLine" (v1.04, Mullen, 2012). Before performing independent component analysis (ICA), data was epoched from −500 ms to 1,000 ms around face-onset and the mean of the prestimulus baseline (−500 ms to 0 ms) was subtracted. Extended Infomax ICA was performed after a PCA reduction to 63 channels on a 1-Hz high-pass filtered copy of this dataset. The resulting ICA weights were transferred to the original 0.01-Hz filtered dataset. Independent components (ICs) were removed if labeled as muscle (>80%), eye (>90%), and channel-noise (>90%) components using "IClabel" (v1.2.4, Pion-Tonachini et al., 2017). Remaining diverging channels (>3 SD) were spherically interpolated. Then, epochs were trimmed to −200 to 800 ms and baseline-corrected (−200 ms to 0 ms). Trial-wise artifact rejection was performed: amplitudes exceeding −100/100 μV ($M = 42.15$; 4.84%), steep amplitude changes (> 100 μV within an epoch; $M = 3.80$; 0.44%), improbable activation (>3 SD of the mean distribution for every time point; $M = 108.33$; 12.4%) were excluded. Overall, the mean rejection rate was 15.29%. Eye blinks during baseline or face presentation were excluded in a separate step using pupil data. We extracted the following ERPs based on time windows and regions of interest (ROI) electrodes of a previous study (Ziereis and Schacht, in revision): mean and peak amplitudes for the P1 (80-120 ms) at an occipital electrode cluster (O1, O2, and Oz); mean (and peak)[1] amplitudes for the N170 (130-200 ms) at an occipitotemporal electrode cluster (P10,

---

[1] Not preregistered.

P9, PO8, PO7); mean amplitudes for the EPN (250-300 ms) at an occipitotemporal cluster (O1, O2, P9, P10, PO7, and PO8); and mean amplitudes for the LPC (400-600 ms) at an occipito-parietal electrode cluster (Pz, POz, PO3, and PO4). In addition, we explored ERP effects between familiar and novel identities in the old-new task. We analyzed a mid-frontal FN400 (300-500 ms, at Fc3, F3, Fc4, F4), which has been related to familiarity of faces (Curran & Hancock, 2007), and a later parietal old/new component LPON (500 800 ms at CP1, CP2, P3, and P4), which has been related to episodic memory and recollection (Proverbio et al., 2019). Notably, the later old/new component overlaps in time and topography with the LPC component. We also explored auditory processing of the refresher trials in voice-locked N1-P2 ERP complex with N1 (90-145 ms) and P2 (165-300 ms), both with the identical frontocentral electrode cluster: F3, F1, Fz, F2, F4, FC1, FC3, FC2, FC4, C3, C1, Cz, C2, C4, CP1, CP3, CPz, CP2, and CP4, which we have included in the Supplementary Information.

Statistical analysis

Tables with statistical models (incl. estimates, confidence intervals, stability measures, and likelihood ratio tests) are in the Supplementary information. All statistical analysis was conducted in R (v 4.0, R Core Team, 2020). All statistical models but the beta inflated distribution model (see below) were 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. The significance of the predictors was tested with likelihood ratio tests (LRT) of models including the predictor against reduced models and a null model. Post-hoc contrasts were used to test the difference between factor levels using "emmeans" (Lenth, 2020). We used the conventional significance level $\alpha = .05$ (two-sided) and for posthoc tests Šidák-correction to adjust for multiple comparisons. To estimate the parameters in the analyses, we used the maximum likelihood (ML) estimator. For the 95% confidence intervals we used nonparametric bootstrapping (nsim = 999) if not specified otherwise.

Ratings of the voices

Due to the nature of the slider response measure with lower and upper bounds, we used a beta inflated distribution model (GAMLSS family "BEINF"; Stasinopoulos & Rigby, 2007) for the ratings of the voice stimuli. The model allows zero and one as values for the response variable. The beta inflated distribution is given as

$$
\begin{aligned}
f(y) &= p_0, & & if\ (y = 0),\\
f(y) &= p_1, & & if\ (y = 1),\\
f(y|\alpha, \beta) &= \frac{1}{B(\alpha,\beta)} y^{\alpha-1}(1-y)^{\beta-1} & & otherwise
\end{aligned}
$$

for y = (0,1). The full model included the predictors emotion of the voice stimulus, type of the rating (valence of the voice vs. personal reaction to the voice), their interaction, and a random intercept for participant ID.

### Learning phase (learning speed)

We modelled accuracy of the learning check trials until the learning criterion was reached (for the first time) with a binomial mixed model (GLMM). Predictors of the binomial mixed model were valence, number of learning checks (per valence), and their interaction. We included random slopes of valence and number of checks and the random intercept participant ID.

Test session

Only correctly answered trials were included in the ERP analysis. The study had a 2 (task: old-new/valence-classification) × 3 (valence: negative/positive/neutral) within-subjects design. For all outcomes (P1, N170, EPN and LPC amplitudes), mixed models with the fixed effects valence (positive, negative, neutral), task (old-new and valence-classification), their interaction, and the random effect (intercept) participant ID were analyzed. Although we expected the associated effects to reflect valence rather than the individual emotion categories, we included models for all ERPs with the fixed effects task and emotion (6 levels) and their interaction. In addition to these ERPs, for the old-new task, we analyzed the FN400 and LPON in a separate models and added the level ("novel") to the predictor variables valence/emotion.

For response time data, only correct trials were selected. Separately for each participant, task, and condition, data were trimmed to a maximum cutoff of 5,000 ms after face onset and a skewness-adjusted boxplot method was used to exclude extreme values (function "adjbox" of the package "robustbase," Maechler et al., 2021; based on Hubert & Vandervieren, 2008). After averaging across participants and conditions, response time data still resulted in skewed residuals. By taking the natural log of the averaged response times, the distribution of residuals became less skewed. We reported all model parameters on the log scale. Our model included valence, task, the valence × task interaction as fixed effects, valence and task as random slopes and participant ID as random intercept. The model allowed random slopes and the random intercept to be correlated. Additionally, for the old-new task, we tested for response time differences between familiar and novel faces in a separate model by adding the level ("novel") to the predictor variable valence.

Valence and reaction ratings of the voice stimuli
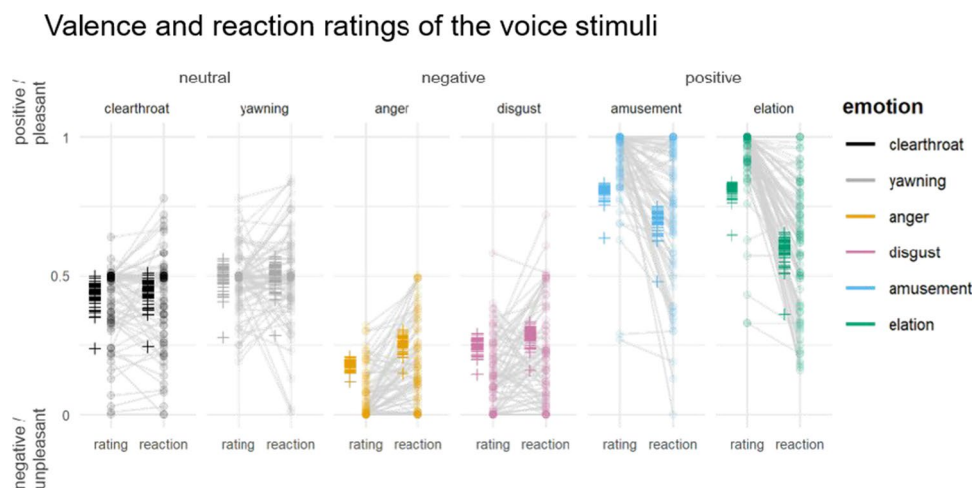
Fig. 2 Rating of the vocal bursts (pre-learning). Stimuli were rated regarding the speakers' emotional valence (positive/negative) and the personal reaction to the stimuli (unpleasant/pleasant) on separate slid-ers without an initial thumb. Dots represent the raw rating data per stimulus and participant. Crosses represent the predicted values per participant based on the zero-one-inflated GAMLSS model.

We ran mixed logistic regression models (binomial GLMMs) on the accuracy data of the test session. The predictor variables (UVs) were task (old-new and valence-classification), valence of the associated sounds (negative, neutral, positive), and their interaction. Because we detected overdispersion in the preregistered model (which included only participant ID as a random intercept), we maximized the random effects structure for the model including valence with a random intercept of participant ID, a random slope for valence and a random slope for task. Including a slope for the interaction between valence and task resulted in singularity issues and was dropped from the model.

For the likability rating, we ran ordinal-mixed-models with valence (including novel) or emotion as fixed effects and random intercepts for face and participant ID. Model estimates and their 95% confidence intervals are reported as odds ratios.

Results

Learning session

Valence rating of the voices

Before the first learning session, participants evaluated the individual voices along the dimensions "valence of the speaker's expression" and "reaction to the burst." The zero-one-inflated model showed a main effect of emotion ($\chi^2(0.70) = 6$, $p = .008$), rating type (valence rating vs. reaction: $\chi^2(23.10) = 828.21$, $p < .001$) and a significant emotion × rating type interaction ($\chi^2(6.34) = 137.78$, $p < .001$). Consistent with the prespecified valence categories,

participants rated the bursts as negative, neutral, and positive vocal expressions. However, their overall personal reaction to the stimuli was more homogeneous across emotion categories, with less positive reactions to elation ($\beta_{\text{elation\_reaction}} = -1.48$, CI $= [-1.85; -1.12]$) and amusement ($\beta_{\text{amusement\_reaction}} = -0.84$, CI $= [-1.20; -0.48]$) and less negative reactions to anger ($\beta_{\text{anger\_reaction}} = 0.67$, CI $= [0.30; 1.04]$); Fig. 2.

Repetitions of face-voice pairs

The number of repetitions of face-voice pairs varied across participants. The total number of repetitions ranging from 18 to 428 and, for a given valence group, from 4 to 7 to 125 to 157. When proportions were considered, positive face-voice pairs were repeated least frequently (median = 32%) but had the largest range (25-47%), followed by neutral (33%; 22-41%) and negative (35%; 26-45%) face-voice pairs.

Learning speed (accuracy) by valence

There was a main effect of number of learning checks ($\chi^2(1) = 51.09$, $p < .001$), no effect of valence ($\chi^2(2) = 1.9$, $p = .387$), but a valence × check number interaction ($\chi^2(2) = 7.95$, $p = .019$). Until the learning criterion was met, there were differences in learning speed between valence categories. Positive face-voice pairs were learned significantly faster than negative face-voice pairs at early check trials (predicted accuracies were outside 95% point-wise CI of the other valence category between the second and sixth learning checks per valence category). Differences between other valence categories over time were not significant (Fig. 3).

Accuracy by valence during learning
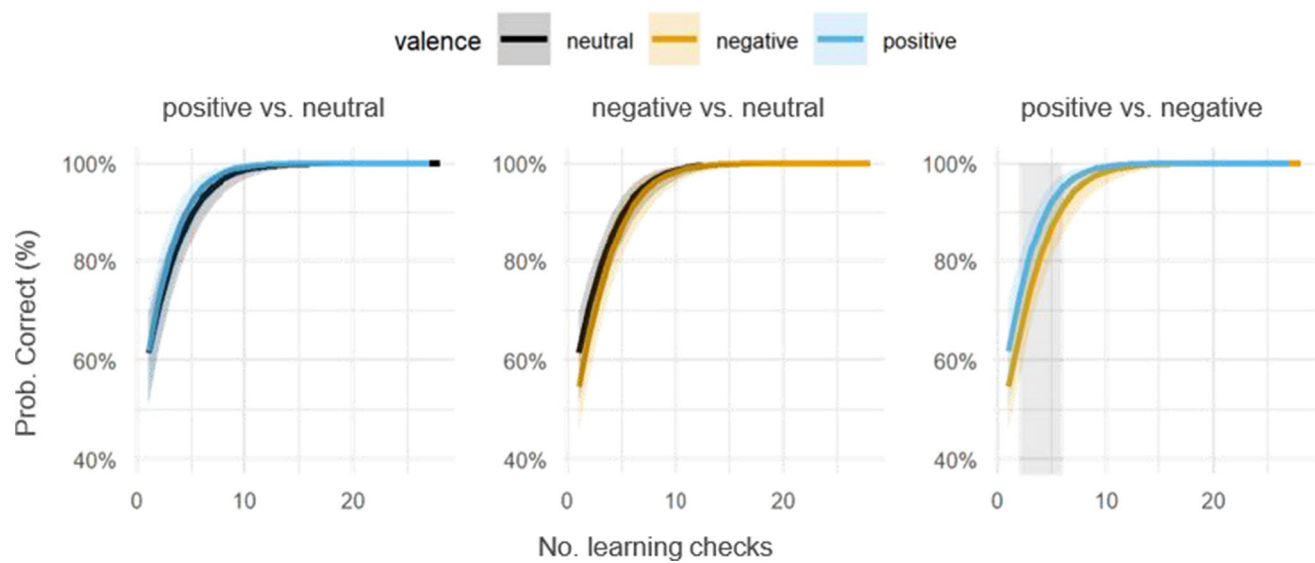
Fig. 3 Predicted accuracy for learning checks. The x-axis refers to the number of learning checks separately per valence category and the y-axis to the predicted probability of a correct answer in a learning check. The light-gray highlighted area reflects significant differences between valence categories (predicted accuracy of one curve outside the 95% CI of the other).

Learning strategies

Except for the mandatory learning checks, the learning phase could be organized flexibly by the participants. To gain more knowledge about how they experienced learning (i.e., perceived difficulty and subjective learning styles), we asked all participants to complete an online questionnaire the day before the lab session. Overall, participants varied in how difficult they rated the learning task. On a Likert scale ranging from 1 (very hard) to 5 (very easy), studying the face-voice pairs and reaching the learning criterion were rated on average as rather easy ($M = 3.76$, SD = 0.85). All participants indicated that certain face-voice pairs were more difficult to memorize. However, participants differed in what they specified as difficult: high similarity between faces ($n = 23$), lack of distinctive facial features ($n = 16$), gender of the face (female faces easier ($n = 3$), male faces easier ($n = 5$)), emotion (neutral more difficult ($n = 4$), anger and disgust within male faces more difficult ($n = 1$), and subjective mismatch between faces and voices ($n = 4$).

The majority of participants ($n = 33$) indicated that they used at least one specific strategy to study the face voice

pairs, of which mnemonic device[2] ($n = 28$) was mentioned most often, followed by focusing on specific distinctive facial features ($n = 26$) in order to be able to distinguish faces. Less frequently, they reported that they formed subgroups of stimuli (e.g., female pairs first) and learned them separately ($n = 5$). Most participants began by using the card deck ($n = 25$). However, after a while some ($n = 6$) preferred to use mainly the learning checks to look at the faces for a longer duration and to get feedback on which faces still needed practice. Only two participants initially used the spatial information of the preview cards but stopped because the positions were shuffled in each session.

Participants rated their everyday ability to memorize faces on a 5-point Likert scale from 1 (very hard) to 5 (very easy) as rather high ($M = 3.92$, SD = 1.10). This self-reported ability did not significantly correlate with the number of learning checks required to meet the learning criterion ($r(36) = -0.24$, $p = .814$).

Test session

Table 1 contains the averaged means and standard deviations of the behavioral measures of and ERPs of the testsession.
ERP results

P1: Associated valence P1 mean amplitudes were neither modulated by task ($\chi^2(1) = 0.46$, $p = .497$) nor valence ($\chi^2(2) = 0.14$, $p = .931$) nor their interaction ($\chi^2(2) = 0.54$, $p =$

---

[2] Mnemonic device entailed associating a particular facial feature with a character trait that matches the voice, making up stories about the person, comparing them to somebody they know or giving them names. In some cases, participants wrote their mnemonic devices on paper and used them as a learning aid until they could allocate the faces and voices without it.

.764). Similarly, P1 peak amplitudes were neither modulated by task ($\chi^2(1) = 0.67$, $p = .413$) nor valence ($\chi^2(2) = 0.22$, $p = .896$) nor their interaction ($\chi^2(2) = 0.91$, $p = .635$).

Associated emotion Replacing valence with the individual emotion categories did not change the results of P1 mean amplitudes (task:$\chi^2(1) = 0.26$, $p = .607$; emotion: $\chi^2(5) = 2.07$, $p = .839$; task × emotion: $\chi^2(5) = 0.96$, $p = .965$). Similarly, a model including emotion did not significantly explain P1 peak amplitudes (task: $\chi^2(1) = 0.99$, $p = .319$; emotion: $\chi^2(5) = 5.33$, $p = .377$; task × emotion: $\chi^2(5) = 1.41$, $p = .923$) (Table 1).

N170: Associated valence N170 mean amplitudes were not modulated by valence ($\chi^2(2) = 2.1$, $p = .350$), but there was a main effect of task ($\chi^2(1) = 28.72$, $p < .001$). Mean amplitudes averaged across valence conditions were significantly more negative in the valence-classification task ($-7.16\,\mu V$; $\beta_{valclass} = -0.23$, $SE = 0.04$, $t = -5.49$) than in the old-new task ($-6.69\,\mu V$). There was no interaction between valence and task ($\chi^2(2) = 1.69$, $p = .430$). N170 peak amplitudes were not modulated by valence ($\chi^2(2) = 1.65$, $p = .437$), task ($\chi^2(1) = 0.21$, $p = .648$) or the valence × task interaction ($\chi^2(2) = 0.77$, $p = .681$).

Associated emotion Looking at emotion categories separately, N170 mean amplitudes were significantly modulated by emotion ($\chi^2(5) = 13.67$, $p = .018$), with disgust showing an enhanced negative mean amplitude ($-7.21\,\mu V$; $\beta_{dis} = -0.28$, $SE = 0.09$, $t = -3.04$). Also, in this model, a main effect of task was present ($\chi^2(1) = 32.96$, $p < .001$) with more negative mean amplitudes for the valence-classification task ($-7.17\,\mu V$; $\beta_{valclass} = -0.23$, $SE = 0.04$, $t = -5.78$). However, there was no interaction between task and emotion present ($\chi^2(5) = 3.58$, $p = .612$). Similar to mean amplitudes, emotion significantly modulated peak amplitudes ($\chi^2(5) = 11.58$, $p = .041$) with enhanced peak amplitudes for disgust ($-11.42\,\mu V$; $\beta_{dis} = -0.31$, $SE = 0.10$, $t = -3.01$). There was no effect of task ($\chi^2(1) = 0.76$, $p = .383$) and no interaction between emotion and task ($\chi^2(5) = 1.66$, $p = .894$) (Fig. 4).

EPN: Associated valence There was a main effect of valence on EPN amplitudes ($\chi^2(2) = 10.86$, $p = .004$). This was due to enhanced negative amplitudes for negatively ($-2.54\,\mu V$; $\beta_{neg} = -0.28$, $SE = .09$, $t = -3.23$) compared to neutrally (diff$_{neu-neg} = 0.47$, $p = .006$) and positively (diff$_{pos-neg} = 0.37$, $p = .045$) associated faces. There was no main effect of task on EPN amplitudes ($\chi^2(1) = 0.01$, $p = .925$) and no interaction between valence and task ($\chi^2(2) = 0.13$, $p = .936$).

Associated emotion Looking at emotion categories separately, there was a main effect of emotion on EPN amplitudes

($\chi^2(5) = 21.61$, $p \leq .001$) due to enhanced negative amplitudes for disgust ($-2.73\,\mu V$; $\beta_{dis} = -0.46$, $SE = 0.12$, $t = -3.79$) compared with the neutral categories throat-clearing (diff$_{dis-clt} = -0.58$, $p = .031$) and yawning (diff$_{dis-yaw} = -0.76$, $p \leq .001$) and compared with the positive category elation (diff$_{dis-el} = -0.66$, $p = .008$), collapsed across tasks. Also in this model, task did not modulate EPN amplitudes ($\chi^2(1) = 0.04$, $p = .838$) and the emotion × task interaction ($\chi^2(5) = 1.47$, $p = .917$) was not significant (Fig. 5).

LPC: Associated valence LPC amplitudes showed no modulation by valence ($\chi^2(2) = 2.64$, $p = .268$) or task ($\chi^2(1) = 1.08$, $p = .298$). Although the valence × task interaction was not significant, ($\chi^2(2) = 4.46$, $p = .108$), looking at the time course of the component, affectively compared to neutrally associated faces appeared to show a different activation in the valence-classification task. Post-hoc tests showed trends toward a difference between positive and neutral (diff$_{pos-neu} = 0.55$, $p = .054$), and between negative and neutral categories (diff$_{neg-neu} = 0.49$, $p = .098$), which were present only in the valence-classification task.

Associated emotion LPC amplitudes were not significantly explained by individual emotion levels ($\chi^2(5) = 5.01$, $p = .414$), or task ($\chi^2(1) = 2.46$, $p = .117$) or an emotion × task interaction ($\chi^2(5) = 6.2$, $p = .287$). Descriptively, the neutral categories elicited lower amplitudes in the valence-classification task, but also here, none of the post-hoc contrasts were significant (Fig. 6).

LPON and FN400

LPON and FN400 mean amplitudes were analyzed only for faces presented in the old-new task. The predictors of the models include a level for novel faces in addition to the valence/emotion levels.

FN400[3]: Associated valence There was no effect of valence ($\chi^2(3) = 2.73$, $p = .436$), and thus, no significant difference between associated and novel faces.
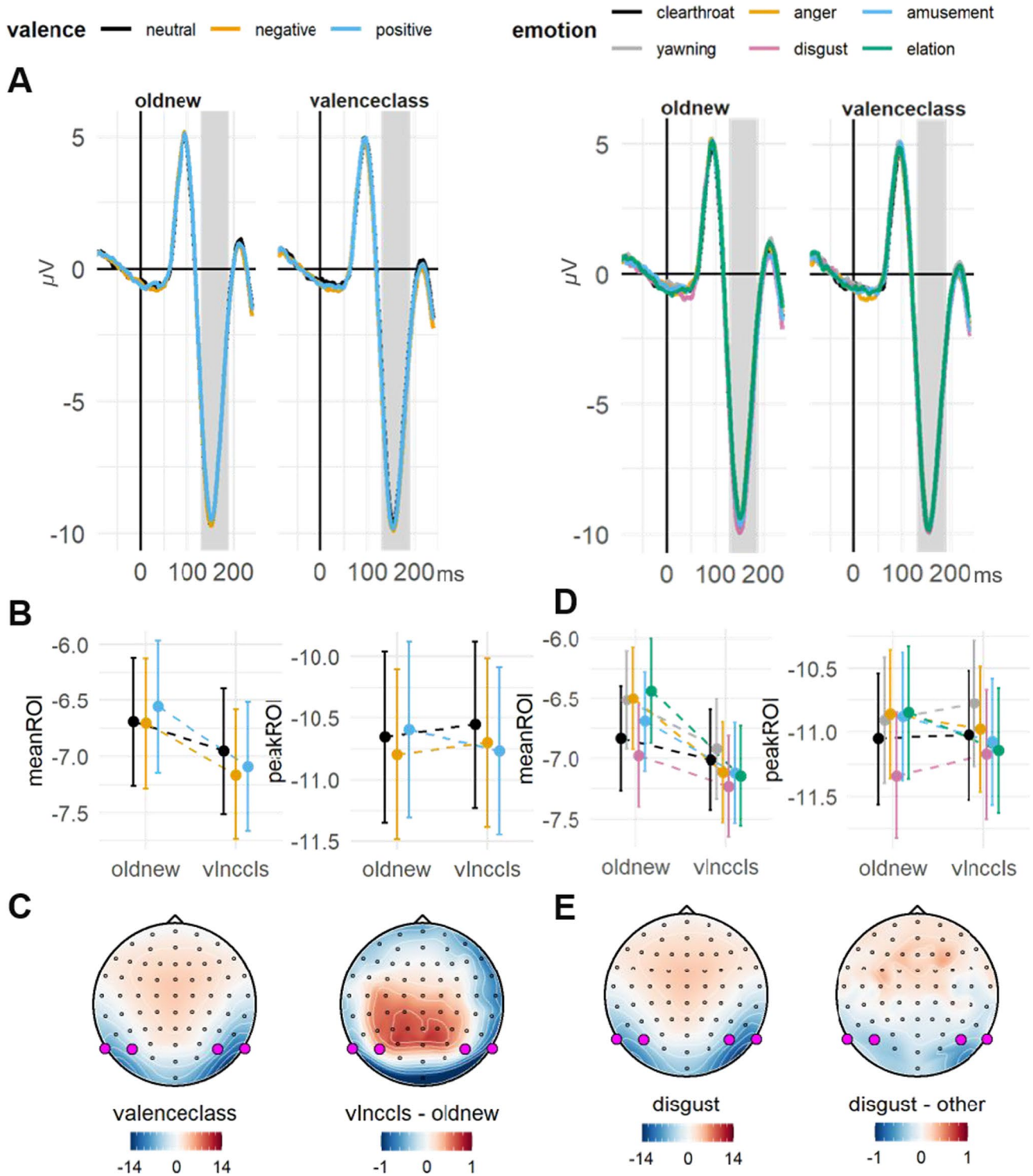
Associated emotion There was a trend for emotion ($\chi^2(6) = 10.85$, $p = .093$), but none of the post-hoc tests were significant. The largest difference between categories was between yawning and disgust (diff$_{yaw-dis} = -0.41$, $p = .137$).

---

[3] Two participants (IDs: 30, 7) were excluded from this analysis due to influential observations and Cook's distance >1 in the model with valence as a predictor. To compare results, they were also excluded from the model with separate emotion categories.

Table 1. Mean (SD) accuracy (%), RT (ms), ERP amplitudes in μV of the test session for valence and emotion by task

| Task	Measure	Neutral			Positive			Negative			Novel
		Avr. neu	Yawning	Clear throat	Avr. pos	Elation	Amusement	Avr. neg	Disgust	Anger	Novel
Old/new task	Acc.Test	98.6 (4.04)	99.1 (1)	98.1 (7.69)	98.83 (3.23)	99.45 (1)	98.2 (6.19)	98.35 (3.17)	98.25 (4.53)	98.45 (3.72)	98.45 (1.76)
	RT	725 (136)	726 (149)	723 (121)	722 (139)	721 (137)	723 (146)	733 (144)	743 (164)	724 (133)	756 (116)
	P1	4.05 (4.22)	4.02 (4.32)	4.04 (4.25)	4.08 (4.08)	3.98 (4.06)	4.17 (4.25)	4.08 (4.05)	4.16 (3.96)	3.93 (4.17)	
	P1 peak	7.04 (4.15)	7.2 (4.36)	7.51 (4.16)	6.93 (3.80)	7.06 (3.90)	7.4 (3.94)	6.88 (3.89)	7.32 (3.80)	7.1 (4.05)	
	N170	-6.73 (3.78)	-6.87 (4.01)	-6.55 (3.74)	-6.61 (3.90)	-6.75 (3.82)	-6.49 (4.04)	-6.74 (3.85)	-6.51 (3.91)	-7.05 (3.86)	
	N170 peak	-10.73 (4.57)	-11.14 (4.69)	-10.97 (4.55)	-10.7 (4.68)	-10.97 (4.62)	-10.96 (4.80)	-10.88 (4.56)	-10.93 (4.60)	-11 (4.54)	
	EPN	-2.08 (3.30)	-2.21 (3.48)	-1.91 (3.31)	-2.19 (3.25)	-2.38 (3.11)	-2.03 (3.50)	-2.51 (2.97)	-2.33 (3.23)	-2.73 (2.89)	
	LPC	5.32 (3.07)	5.25 (3.04)	5.36 (3.25)	5.25 (3.45)	5.21 (3.52)	5.28 (3.51)	5.24 (3.19)	5.39 (3.60)	5.1 (2.91)	
Valence-classi-fication task	Acc.Test	95.38 (9.50)	96.2 (9.11)	94.55 (11.84)	97.08 (4.58)	97.05 (6.23)	97.1 (5.51)	96.7 (6.09)	97 (8.66)	96.4 (8.46)	
	RT	957 (148)	943 (152)	974 (167)	930 (143)	919 (122)	944 (187)	934 (178)	928 (182)	945 (202)	
	P1	4.08 (4.41)	3.97 (4.43)	4.14 (4.46)	3.96 (4.18)	3.98 (3.92)	4 (4.60)	3.98 (4.01)	4.02 (3.96)	3.91 (4.20)	
	P1 peak	6.81 (4.26)	7.07 (4.24)	7.34 (4.26)	6.91 (3.99)	7.2 (3.67)	7.2 (4.46)	6.86 (4.08)	7.12 (4.02)	7.05 (4.35)	
	N170	-7.04 (3.70)	-7.14 (3.76)	-6.99 (3.75)	-7.15 (3.76)	-7.2 (3.79)	-7.12 (3.82)	-7.28 (3.80)	-7.21 (3.78)	-7.38 (3.84)	
	N170 peak	-10.71 (4.40)	-11.2 (4.51)	-10.9 (4.41)	-10.86 (4.46)	-11.18 (4.51)	-11.15 (4.51)	-10.86 (4.47)	-11.08 (4.48)	-11 (4.53)	
	EPN	-2.06 (3.21)	-2.1 (3.36)	-2.03 (3.30)	-2.16 (3.45)	-2.2 (3.75)	-2.12 (3.36)	-2.58 (3.10)	-2.52 (3.34)	-2.74 (3.09)	
	LPC	5.06 (3.42)	5.06 (3.38)	5.16 (3.57)	5.61 (3.80)	5.7 (3.96)	5.55 (3.78)	5.55 (3.46)	5.63 (3.56)	5.51 (3.54)	

Fig. 4 Face-locked N170 by valence and emotion. **A** Grand average ERP time series of the averaged ROI channels (left panel: valence, right panel: emotion). The highlighted area indicates the ROI time window. **B** (valence) and **D** (emotion): Grand-averages of the ROI mean amplitudes (left panel) and peak amplitudes (right panel), contrasted for the implicit and explicit task and all valence/emotion conditions. Error bars indicate ±1 SE of the mean. **C** Topographies of the ERP distribution for faces in the valence-classification task contrasted with the old-new task, averaged across the valence conditions. **E** Topographies of the ERP distribution for faces associated with disgust bursts contrasted with all other emotion conditions, averaged across the tasks. ROI channels are highlighted in pink

EPN by valence and emotion
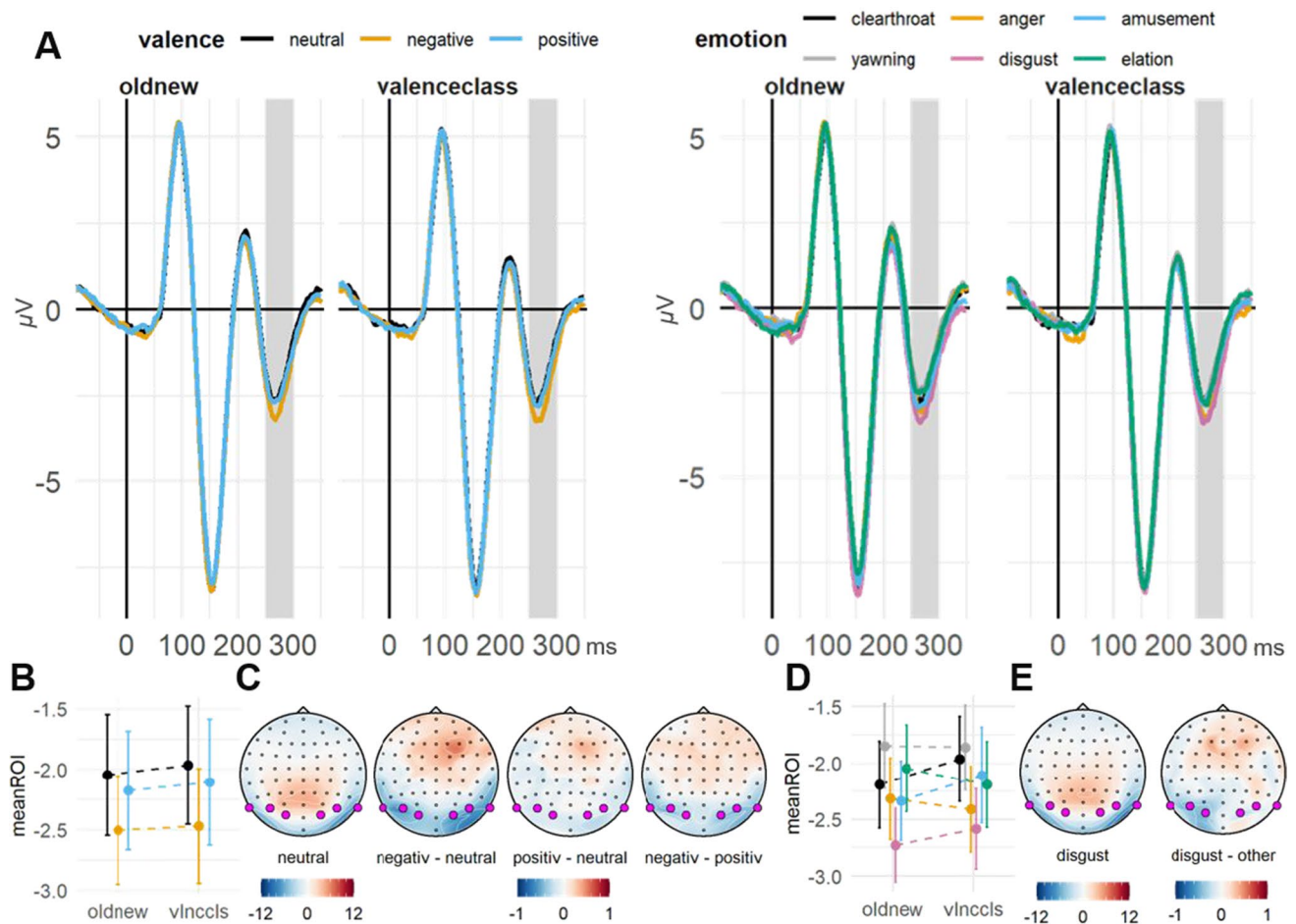
Fig. 5 Face-locked EPN by valence and emotion. **A** Grand average ERP time series of the averaged ROI channels for valence (left panel) and emotion (right panel). The highlighted area indicates the ROI time window. **B** (valence) and **D** (emotion): Grand-averages of the ROI mean amplitudes, contrasted for the implicit and explicit task and all valence/emotion conditi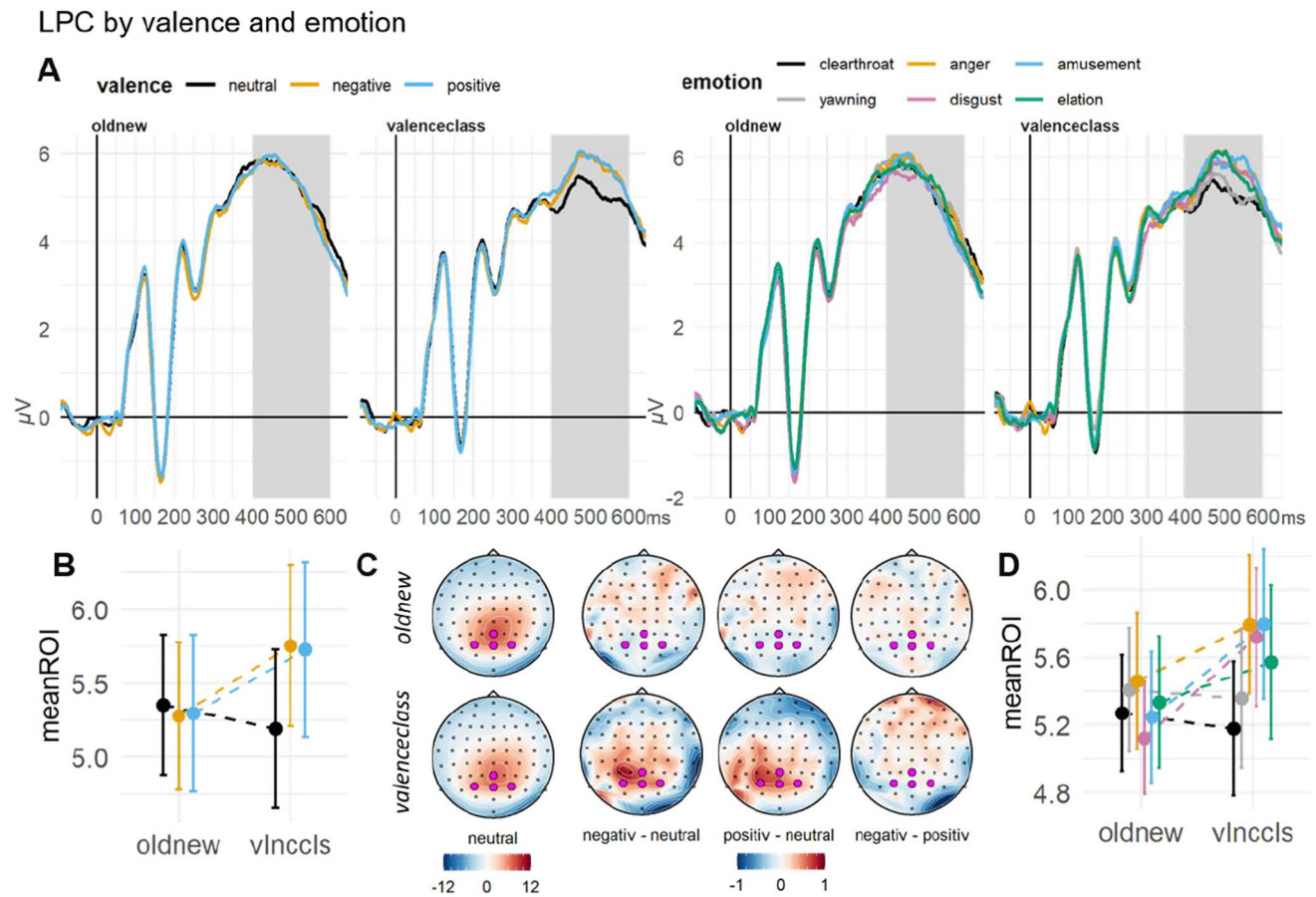ons. Error bars indicate ± 1 SE of the mean. **C** Topographies of the ERP distribution for faces between valence conditions, averaged across the implicit and explicit tasks. **E** Topographies of the ERP distribution for faces associated with disgust bursts contrasted with all other emotion conditions, averaged across the implicit and explicit tasks. ROI channels are highlighted in pink.

LPON: Associated valence There was a main effect of valence ($\chi^2(3) = 84.87$, $p < .001$) with a difference between novel faces and all associated faces (diff$_{nov-pos}$ = 1.36, $p < .001$; diff$_{nov-neg}$ = 1.30, $p < .001$; diff$_{nov-neu}$ = 1.17, $p < .001$). No other difference between valence levels was significant.

Associated emotion Analogously to the valence model, there was a main effect of emotion ($\chi^2(6) = 88.54$, $p < .001$) with a significant difference between novel faces and all emotion categories (all p-values $< .01$). None of the other post-hoc contrasts between emotion levels were significant (Fig. 7).

Behavioral Outcomes

Response times. Associated valence In line with our hypothesis, responses were slower in the valence-classification task than in the old-new task (diff$_{valclass-oldnew}$ = 212 ms; $\chi^2(1)$ = 56.42, $p < .001$). There was no main effect of valence ($\chi^2(2)$ = 2.14, $p = .343$), but an interaction between valence and task ($\chi^2(2) = 6.55$, $p = .038$). In the valence-classification task, neutral trials were descriptively slower than positive (and to a lesser extent also negative) trials, but post-hoc differences were not significant (all p-values $> .05$).

LPC by valence and emotion

Fig. 6 Face-locked LPC by valence and emotion. **A** Grand average ERP time series of the averaged ROI channels for valence (left panel) and emotion (right panel). The highlighted area displays the ROI time window. **B** (valence) and **D** (emotion): Grand averages of the ROI mean amplitudes, contrasted for the implicit and explicit task and all valence/emotion conditions. Error bars indicate ±1 SE of the mean. **C** Topographies of the ERP distribution for faces between valence conditions, separately for the implicit (old-new) and explicit (valence class.) tasks. ROI channels are highlighted in pink.

Associated emotion Due to singularity issues we reduced the model structure to a random intercept model with participant ID. Also in this model, there was a main effect of task ($\text{diff}_{\text{valclass-oldnew}} = 213$ ms; $\chi^2(1) = 406.07, p < .001$). However, neither emotion ($\chi^2(5) = 2.68, p = .749$), nor the interaction between emotion and task was significant ($\chi^2(5) = 4.64, p = .461$). *Old/new task comparison:* When comparing the valence categories and novel stimuli in the old new task, participants responded more slowly to novel faces than to faces known from the learning phase, irrespective of their valence ($\chi^2(3) = 19.67, p < .001$). The largest difference was between positive and novel faces ($\text{diff}_{\text{pos-nov}} = -37$ ms, $p < .001$).

Accuracy. Associated valence As hypothesized, accuracy was lower in the valence-classification task compared with the old-new task ($\text{OR}_{\text{valclass/oldnew}} = 0.32, \chi^2(1) = 13.85, p < .001$). Valence was not significant ($\chi^2(2) = 0.13, p = .938$), and there was no significant interaction between task and valence ($\chi^2(2) = 0.56, p = .756$).

Associated emotion A model with single emotion levels resulted again in a main effect of task ($\chi^2(1) = 14.16, p < .001$). There was no main effect of emotion ($\chi^2(5) = 2.74, p = .740$) but a significant interaction between task and emotion ($\chi^2(5) = 12.72, p = .026$). Post-hoc tests showed that for all emotion categories but anger ($\text{OR} = 0.61, p = .198$) and elation ($\text{OR} = 0.48, p = .051$) the valence-classification task had a significantly lower accuracy compared to the old-new task (all p-values $\leq .05$).

Likability rating

We ran two (one for valence and one for emotion levels) cumulative linked mixed models to account for the ordinal scale of the likability ratings. Both models included random intercepts for participant and face stimulus. Likelihood ratio tests of both models and a model without a fixed effect indicated that valence significantly explained the variance of the

LPON by valence and emotion (incl. novel faces)
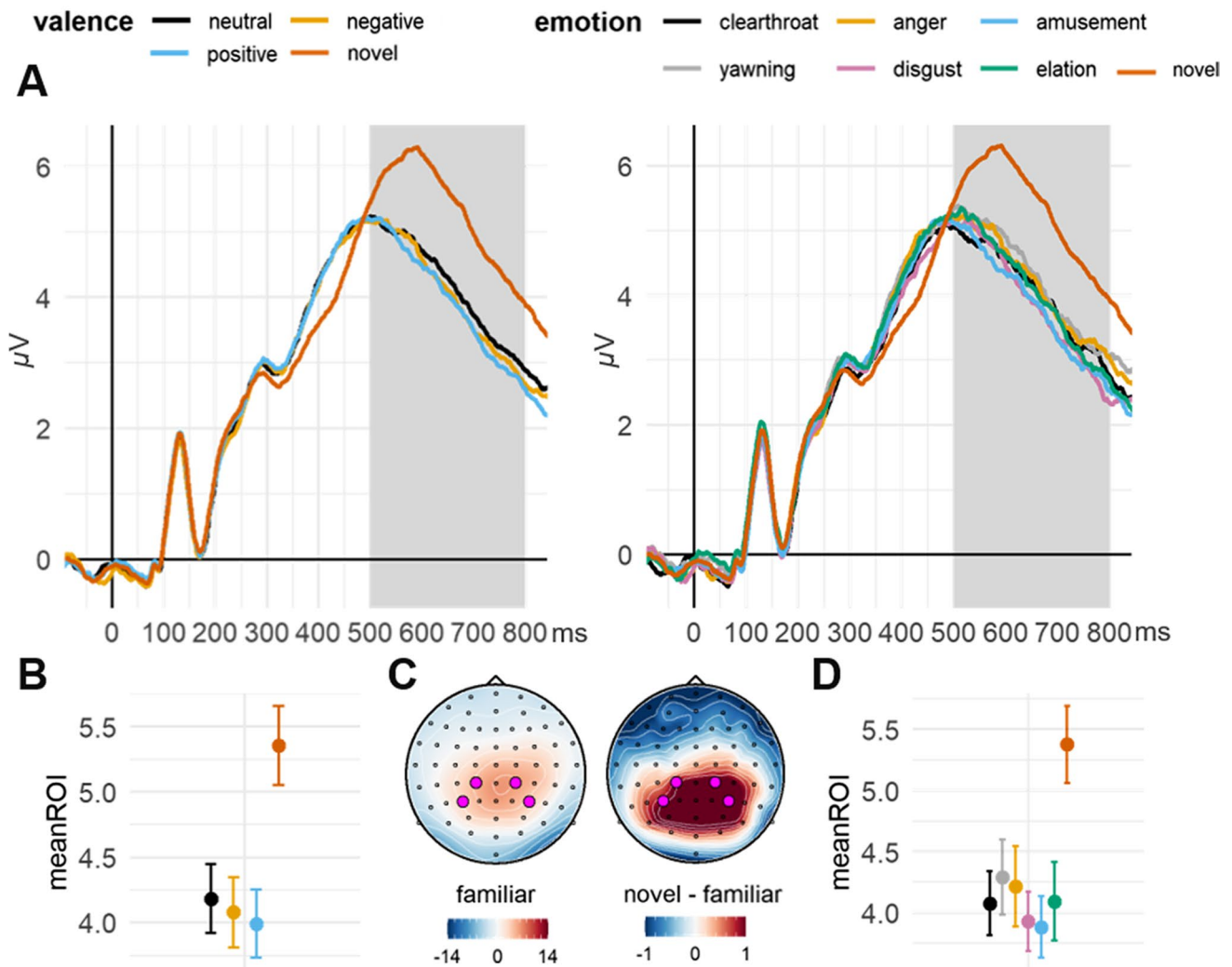
Fig. 7 Face-locked LPON by valence and emotion in the old/new task. **A** Grand average ERP time series of the averaged ROI channels for valence (left panel) and emotion (right panel). The highlighted area displays the ROI time window. **B** (valence) and **D** (emotion): Grand-averages of the ROI mean amplitudes, contrasted for all conditions including novel faces. Error bars indicate ±1 SE of the mean. **C** Topographies of the ERP distribution for familiar (i.e., associated) and novel faces. ROI channels are highlighted in pink.

rating data ($\chi^2(3) = 96.19$, $p \leq .001$). However, separating emotion categories did not explain the data better than the valence categories ($\chi^2(3) = 0.79$, $p = .851$). The odds ratios and 95% CI of both models are reported in Table A20 of the Supplementary Information. Mean ratings and model predictions are shown in Fig. 8.

Associated valence Associated valence modulated the likability ratings in line with our hypothesis, with positively associated faces being rated the most likable and negatively associated faces the least likable. With the exception of novel and negatively associated faces, all pairwise differences between valence categories were significant (all p-values < .01).

Associated emotion The ordinal model including emotion categories showed a grouping of levels according to the pre-specified valence categories. There were no significant differences within valence categories (neutral: throat-clearing and yawning; negative: anger and disgust; positive: amusement and elation). Pairwise comparisons of emotion categories of different valences were significant (all p-values < .05), except for throat-clearing and all positive categories, yawning and amusement, and novel and all negative categories.
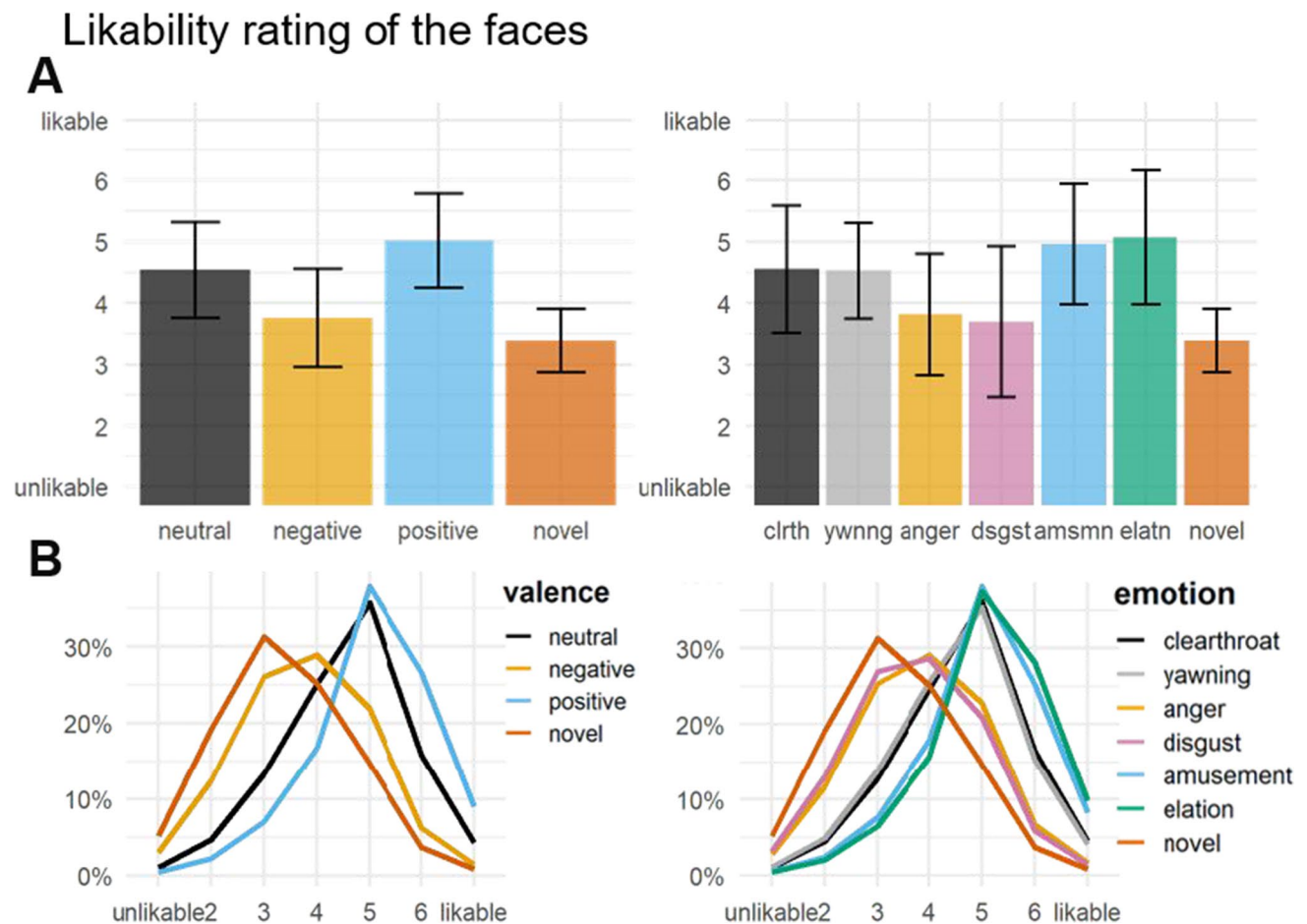
Fig. 8 Likability rating. (left panels: by valence, right panels: by emotion). **A** Barplots represent the likability ratings per condition, averaged within and across subjects. Error bars show ±1 SD. **B** Fitted values are the predicted probabilities of the ordinal models.

Discussion

The present study aimed to investigate memory-based attention effects on the retrieval of valence-based associations in face perception. After having faces associated with affect bursts in an online learning paradigm, we measured short-, mid-, and long-latency ERPs to faces associated with positive, neutral, or negative valence in a valence-implicit and a valence-explicit task. Consistent with our hypotheses and previous research, we found that faces previously associated with affect bursts were not only rated according to the valence of the context but also elicited differential neural responses from faces associated with a neutral context. Moreover, associated effects in late components were strongly affected by task requirements, suggesting that goal-directed attention on specific associated features affected especially later, more elaborate processing of the faces.

The first associated effect was present in the N170. Although the averaged associated valence did not moderate

N170 amplitudes, there were differences between individual emotion levels, with an enhanced negative amplitude for disgust-associated faces. A number of studies reported N170 effects for valence-associations (Aguado et al., 2012; Bruchmann et al., 2021; Camfield et al., 2016; Luo et al., 2016; Schellhaas et al., 2020; Schindler et al., 2021; Sperl et al., 2021). Due to its measured spatial overlap with the EPN, the N170 has been suggested to represent a mixture of configural face processing and relevance encoding (Rellecke et al., 2012b). In addition, there was an independent effect of task starting in the N170 time window and extending to a positive-going deflection over the lateral occipito-temporal areas peaking around 200 ms (similar to findings by Itier & Neath-Tavares, 2017, and Schindler et al., 2021). The interpretation of this effect is not straightforward: the visually evoked P2 component has been linked to higher-order configural processing (Latinus & Taylor, 2006), differences in task difficulty (Philiastides, 2006), tasks requiring expertise on subgroups of faces (Stahl et al., 2008), and face typicality (Pell et al., 2022), all of which could be roughly

related to deeper processing demands (Banko et al., 2011) of faces in the valence-classification task and differences in processing depth (Itier and Neath-Tavares, 2017). Remarkably, these early task differences did not extend to the EPN time window.

EPN amplitudes were modulated by associated valence, with enhanced amplitudes for the negative compared to the positive and neutral conditions. Several studies reported enhanced neural processing of negatively but not positively associated faces (Luo et al., 2016; Suess et al., 2014; Wieser et al., 2014). This negativity bias also has been shown for threatening facial expressions (Schupp et al., 2004; for a review, see Schindler & Bublatzky, 2020). That negatively associated faces were preferentially processed in our study is remarkable given that affect bursts resemble rather low-intense stimuli. In addition, task neither modulated EPN amplitudes nor did it moderate valence effects, suggesting a rapid and automatic allocation of attention toward negative information related to faces (similar to Baum & Rahman, 2021; Bruchmann et al., 2021; cf. Schindler et al., 2021). Similar to the N170, EPN amplitudes were particularly pronounced for disgust-related faces. Typically, expressions of disgust serve to detect and reject objects that are potentially offensive, toxic, or contaminating to keep oneself safe and healthy (e.g., spoiled food or open wounds). Expressions of disgust directed at us also could serve as social-communicative signals and be interpreted as a risk of social exclusion (Amir et al., 2005; Gan et al., 2022; Judah et al., 2015). Although facial expressions of disapproval, disgust, and anger have been shown to trigger different neural processes (Burklund et al., 2007), auditory expressions of disgust may be perceived as more ambiguous and provide more room for interpretation of social disapproval.

We hypothesized that the attentional focus of the task would particularly affect later processing. Consistent with this hypothesis and previous research on the processing of faces with emotional expressions (for a review, see Schindler & Bublatzky, 2020) and associated faces (Schindler et al., 2021; Bruchmann et al., 2021), associated valence modulations of the LPC were only descriptively present in the valence-explicit task. While early ERPs in the test session showed only effects of negative associations and specifically of disgust-related faces, later processing was modulated by both negatively and positively associated faces, with less strong differences between individual emotion categories. Moreover, positively associated valence was not extinguished but instead triggered by goal-directed memory retrieval, although this was not evident in the valence-implicit task. In our study, LPC modulations were related to the task-relevant goals, while at the same time discriminating between affective and neutral, but not between positive and negative associations. Our results add to findings of previous research reporting LPC effects of positively

associated faces (Baum & Rahman, 2021; Hammerschmidt et al., 2018a, b) and other kinds of visual stimuli (Schacht et al., 2012) and also show that positive affect bursts can be cross-modally associated to faces.

P1 amplitudes were not modulated by task and, contrary to our predictions, were not modulated by associated valence. The P1 has been related to the processing of lower-level stimulus properties, and selective attention through sensory gain mechanisms (Hillyard and Anllo-Vento, 1998; Russo, 2003). Several studies have reported a sensitivity of the P1 to valence-based associations (Aguado et al., 2012; Hammerschmidt et al., 2017; Muench et al., 2016; Schacht et al., 2012; Schindler et al., 2022) and even earlier processing (Rehbein et al., 2014; Sperl et al., 2021; Steinberg et al., 2012). However, other studies on associated faces have reported no modulations of the P1 (Hammerschmidt et al., 2018b; Schindler et al., 2021) or have not examined early ERPs (Baum & Rahman, 2021). It is possible that the association with affective vocal stimuli of lower intensity in our study was not sufficient to elicit a differential activation of the P1. Although associated emotional expressions of the face have been shown to modulate P1 amplitudes (Aguado et al., 2012), and comparable effect sizes of cross-modal and within-modal associations have been reported (Hofmann et al., 2010), the variability in learning might have played a more important role. As other studies reported stable associations after very few conditioning trials (Rehbein et al., 2014; Steinberg et al., 2013; Ventura-Bort et al., 2016), it is unclear what drives neural changes at the early processing of conditioned stimuli (e.g., the number of CS-US couplings, the (dis-)similarity between CS, the intensity of the US, the stimulus duration, or the consolidation period, etc.). By including a learning criterion in our study, we ensured associations between the faces and affective bursts. In addition, we included refresher trials between the valence-implicit and valence-explicit tasks to counteract extinction. However, the number of face-voice conditioning trials in the learning phase varied between participants and thus differed from typical conditioning studies. Some participants developed their own strategies and preferred studying the pairs by doing learning checks, which allowed them to see the faces longer and to get feedback on their answer. However, in these trials, not one face but five faces and the voice were presented simultaneously. It is possible that the association of the face and the voice occurred here at a more explicit level and was rather defined by attending to specific facial features rather than by a gradual tuning of sensory discrimination through associative learning.

We included valence ratings of the voices prior to any association with faces. Overall, participants rated the vocal expressions according to our pre-specified valence categories. Interestingly, ratings on their reaction towards the bursts were less extreme than the expression ratings and

showed larger interindividual variation. Behavioral performance between valence categories differed only in the learning phase of our study, in which faces with positive bursts were learned faster (similar to reward-associated faces in Hammerschmidt et al., 2017, 2018b; and reward-associated words or symbols, e.g., Bayer et al., 2018; Kulke et al., 2019; Rossi et al., 2017). In contrast, during testing, there was no clear evidence that accuracy and reaction times were affected by associated valence, although descriptively, in the valence-explicit task, responses were slower for the neutral condition than for the positive and negative conditions. Nevertheless, as expected, RTs were shorter and accuracy higher in the old-new task than in the valence-classification task, probably due to the number of choices (two vs. three) and to the required depth of processing (recognition vs. explicit recall). The old-new task might have become more difficult over time due to the repetition of the novel faces. However, the behavioral results suggest that, overall, the valence-classification task was more difficult than the old-new task. If cognitive load alone suppressed ERP effects of associated valence, we would have expected it to occur in the more difficult, i.e., in the valence-classification task. The repeated presentation of non-associated faces, the expectation that they would be repeated, and the relative difficulty of discriminating faces by their inner parts might have prevented typical old/new ERP effects, such as the FN400 and the later parietal effects (Curran & Hancock, 2007; Guillaume & Tiberghien, 2013; Proverbio et al., 2019) in the old-new task.

Likability ratings at the end of the test session were affected by associations with voices expressing positive and negative emotions during the learning (similar to Suess et al., 2014). More specifically, and as we hypothesized, ratings were made according to our pre-specified valence categories, whereas emotion within valence categories did not differ. It is possible that valence, rather than the specific emotion category, altered decisions on likability, although we cannot rule out the possibility that the preceding valence-classification task increased homogenization within valence conditions. Although likability ratings supported the ERP results, in our opinion, the ratings may more closely resemble contingency awareness than true changes in likability and may be biased by the focus on valence differences in the preceding task.

Our novel learning paradigm allowed participants to study the face-voice pairs in a flexible manner, and participants took advantage of this as documented by the learning strategy questionnaire. Despite some variation, participants followed similar self-chosen strategies to memorize the pairs, although we did not provide any hints or recommendations on how to study the face-voice pairs. Participants actively searched for distinct facial features to combine them with what they thought would match the emotional valence of the voice (e.g., the man with tired eyes yawned; the woman with warm brown eyes giggled). As the pairing of the faces and voices was randomized, participants reported taking the features that best distinguished the faces and voices, and some participants even took notes to study. Hence, the type of learning was very different from classical Pavlovian conditioning or instrumental learning, where associations might form more gradually. Remarkably, faces associated with moderately negative bursts elicited distinct neural activation regardless of the task requirements and despite this variability in learning.

One limitation of the study might be that, although we deliberately chose this option, we fixed the order of the tasks in the test session (refresher I, old-new task, refresher II, and the valence-classification task). To ensure that only the valence-classification task would elicit explicit attention to the valence-based associations and to avoid spill-over effects to the valence-implicit task, we set the valence-explicit task at the last position of the experimental part, in which we recorded ERPs. Nevertheless, early effects were similar between tasks, and the valence effects in the LPC occurred only in the valence-classification, i.e., the second task, which should have been more prone to be affected by the extinction of the associations or simply by fatigue.

Conclusions

The present study provides new evidence that faces cross-modally associated with affective stimuli of both positive and negative valence have the potential to elicit neurophysiological responses similar to those of inherent affective stimuli. During testing, task demands affected later, more effortful processing, whereas earlier processing indicated an automatic discrimination of negative from other information across both tasks. We demonstrated that associations with even mildly negative stimuli, flexibly acquired through our novel learning paradigm, could influence face processing even in a valence-implicit task, suggesting a rapid prioritization of learned negative context as a protection against potential threats (Lundqvist & Öhman, 2005; Öhman et al., 2001), largely independent of goal-directed attention. In addition, positive associations were learned faster and affected later processing, but only in the presence of goal-directed attention toward valence.

Declarations

Ethics approval The study was conducted in accordance with the Declaration of Helsinki and approved by the local Ethics committee of the Institute of Psychology at the University of Göttingen.

Consent to participate Informed consent was obtained from all individual participants included in the study.

Consent to publish The manuscript does not contain identifying information.

Conflict of interest The authors declare no conflict of interest.

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

Amir, N., Klumpp, H., Elias, J., Bedwell, J. S., Yanasak, N., & Miller, L. S. (2005). Increased activation of the anterior cingulate cortex during processing of disgust faces in individuals with social phobia. *Biological Psychiatry, 57*(9), 975–981. https://doi.org/10.1016/j.biopsych.2005.01.044

Anderson, B. A., Kim, H., Kim, A. J., Liao, M.-R., Mrkonja, L., Clement, A., & Grégoire, L. (2021). The past, present, and future of selection history. *Neuroscience & Biobehavioral Reviews, 130*, 326–350. https://doi.org/10.1016/j.neubiorev.2021.09.004

Anikin, A. (2019). Soundgen: An open-source tool for synthesizing nonverbal vocalizations. *Behavior Research Methods, 51*(2), 778–792. https://doi.org/10.3758/s13428-018-1095-7

Antov, M. I., Plog, E., Bierwirth, P., Keil, A., & Stockhorst, U. (2020). Visuocortical tuning to a threat-related feature persists after extinction and consolidation of conditioned fear. *Scientific Reports, 10*(1), 3926. https://doi.org/10.1038/s41598-020-60597-z

Apergis-Schoute, A. M., Schiller, D., LeDoux, J. E., & Phelps, E. A. (2014). Extinction resistant changes in the human auditory association cortex following threat learning. *Neurobiology of Learning and Memory, 113*, 109–114. https://doi.org/10.1016/j.nlm.2014.01.016

Audacity Team. (2021). Audacity(R): Free Audio Editor and Recorder [Computer application]. Version 2.4.2 retrieved April 28th 2021 from https://audacityteam.org/

Banko, E. M., Gal, V., Kortvelyes, J., Kovacs, G., & Vidnyanszky, Z. (2011). Dissociating the effect of noise on sensory processing and overall decision difficulty. *Journal of Neuroscience, 31*(7), 2663–2674. https://doi.org/10.1523/JNEUROSCI.2725-10.2011

Baum, J., & Rahman, R. A. (2021). Negative news dominates fast and slow brain responses and social judgments even after source credibility evaluation. *NeuroImage, 244*, 118572. https://doi.org/10.1016/j.neuroimage.2021.118572

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., Grass, A., & Schacht, A. (2018). Associated valence impacts early visual processing of letter strings: Evidence from ERPs in a cross-modal learning paradigm. *Cognitive, Affective and Behavioral Neuroscience, 19*(1), 98–108. https://doi.org/10.3758/s13415-018-00647-2

Blechert, J., Testa, G., Georgii, C., Klimesch, W., & Wilhelm, F. H. (2016). The pavlovian craver: Neural and experiential correlates of single trial naturalistic food conditioning in humans. *Physiology & Behavior, 158*, 18–25. https://doi.org/10.1016/j.physbeh.2016.02.028

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

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

Burklund, L. J., Eisenberger, N. I., & Lieberman, M. D. (2007). The face of rejection: Rejection sensitivity moderates dorsal anterior cingulate activity to disapproving facial expressions. *Social Neuroscience, 2*(3–4), 238–253. https://doi.org/10.1080/17470910701391711

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

Cowen, A. S., Elfenbein, H. A., Laukka, P., & Keltner, D. (2019). Mapping 24 emotions conveyed by brief human vocalization. *American Psychologist, 74*(6), 698–712. https://doi.org/10.1037/amp0000399

Curran, T., & Hancock, J. (2007). The FN400 indexes familiarity-based recognition of faces. *NeuroImage, 36*(2), 464–471. https://doi.org/10.1016/j.neuroimage.2006.12.016

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

Davis, F. C., Johnstone, T., Mazzulla, E. C., Oler, J. A., & Whalen, P. J. (2009). Regional response differences across the human amygdaloid complex during social conditioning. *Cerebral Cortex, 20*(3), 612–621. https://doi.org/10.1093/cercor/bhp126

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 Leeuw, J. R. (2014). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods, 47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

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

Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review, 118*(2), 247–279. https://doi.org/10.1037/a0022327

Gan, X., Zhou, X., Li, J., Jiao, G., Jiang, X., Biswal, B., … Becker, B. (2022). Common and distinct neurofunctional representations of core and social disgust in the brain: Coordinate-based and network meta-analyses. *Neuroscience & Biobehavioral Reviews, 135*, 104553. https://doi.org/10.1016/j.neubiorev.2022.104553

Gao, C., Conte, S., Richards, J. E., Xie, W., & Hanayik, T. (2019). The neural sources of N170: Understanding timing of activation in face-selective areas. *Psychophysiology, 56*(6), e13336. https://doi.org/10.1111/psyp.13336

Gottfried, J. A., O'Doherty, J., & Dolan, R. J. (2002). Appetitive and aversive olfactory learning in humans studied using event-related functional magnetic resonance imaging. *The Journal of Neuroscience, 22*(24), 10829–10837. https://doi.org/10.1523/JNEUROSCI.22-24-10829.2002

Guillaume, F., & Tiberghien, G. (2013). Impact of intention on the ERP correlates of face recognition. *Brain and Cognition, 81*(1), 73–81. https://doi.org/10.1016/j.bandc.2012.10.007

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

Hammerschmidt, W., Kagan, I., Kulke, L., & Schacht, A. (2018a). 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. (2018b). 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

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

Heisz, J. J., & Shedden, J. M. (2009). Semantic learning modifies perceptual face processing. *Journal of Cognitive Neuroscience, 21*(6), 1127–1134. https://doi.org/10.1162/jocn.2009.21104

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

Hofmann, W., Houwer, J. D., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin, 136*(3), 390–421. https://doi.org/10.1037/a0018916

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

Hudson, A., Durston, A. J., McCrackin, S. D., & Itier, R. J. (2021). Emotion, gender and gaze discrimination tasks do not differentially impact the neural processing of angry or happy facial expressions a mass univariate ERP analysis. *Brain Topography, 34*(6), 813–833. https://doi.org/10.1007/s10548-021-00873-x

Itier, R. J., & Neath-Tavares, K. N. (2017). Effects of task demands on the early neural processing of fearful and happy facial expressions. *Brain Research, 1663*, 38–50. https://doi.org/10.1016/j.brainres.2017.03.013

Judah, M. R., Grant, D. M., & Carlisle, N. B. (2015). The effects of self-focus on attentional biases in social anxiety: An ERP study. *Cognitive, Affective, & Behavioral Neuroscience, 16*(3), 393–405. https://doi.org/10.3758/s13415-015-0398-8

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

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

Kulke, L., Bayer, M., Grimm, A.-M., & Schacht, A. (2019). Differential effects of learned associations with words and pseudowords on event-related brain potentials. *Neuropsychologia, 124*, 182–191. https://doi.org/10.1016/j.neuropsychologia.2018.12.012

Kyllingsbæk, S., Van Lommel, S., Sørensen, T. A., & Bundesen, C. (2014). Automatic attraction of visual attention by supraletter features of former target strings. *Frontiers in Psychology, 5*, 1383. https://doi.org/10.3389/fpsyg.2014.01383

Lange, K., Kühn, S., & Filevich, E. (2015). "Just another tool for online studies" (JATOS): An easy solution for setup and management of web servers supporting online studies. *PLOS ONE, 10*(6), e0130834. https://doi.org/10.1371/journal.pone.0130834

Latinus, M., & Taylor, M. J. (2006). Face processing stages: Impact of difficulty and the separation of effects. *Brain Research, 1123*(1), 179–187. https://doi.org/10.1016/j.brainres.2006.09.031

Lenth, R. (2020). *emmeans: Estimated marginal means, aka least-squares means*. Version 2.4.2 retrieved July 8th 2021 from https://CRAN.R-project.org/package=emmeans

Lundqvist, D., & Öhman, A. (2005). Emotion regulates attention: The relation between facial configurations, facial emotion, and visual attention. *Visual Cognition, 12*(1), 51–84. https://doi.org/10.1080/13506280444000085

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., … Anna di Palma, M. (2021). *robustbase: Basic robust statistics*. Version 0.93-8 retrieved June 1st 2021 from http://robustbase.r-forge.r-project.org/

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

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

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

Mullen, T. (2012). *NITRC: CleanLine: Tool/Resource Info*. Version 1.04 retrieved May 16th 2019 from http://www.nitrc.org/projects/cleanline

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

Olofsson, J. K., & Polich, J. (2008). Affective visual event-related potentials: Arousal, repetition, and time-on-task. *Biological Psychology, 75*(1), 101–108. https://doi.org/10.1016/j.biopsycho.2006.12.006

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

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., Sethi, S., Rigoulot, S., Rothermich, K., Liu, P., & Jiang, X. (2022). Emotional voices modulate perception and predictions about an upcoming face. *Cortex, 149*, 148–164. https://doi.org/10.1016/j.cortex.2021.12.017

Philiastides, M. G. (2006). Neural representation of task difficulty and decision making during perceptual categorization: A timing diagram. *Journal of Neuroscience, 26*(35), 8965–8975. https://doi.org/10.1523/jneurosci.1655-06.2006

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

Proverbio, A. M., Vanutelli, M. E., & Viganò, S. (2019). Remembering faces: The effects of emotional valence and temporal recency. *Brain and Cognition, 135*, 103584. https://doi.org/10.1016/j.bandc.2019.103584

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Version 4.0 retrieved June 8th 2020 from https://www.R-project.org/

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

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

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

Röer, J. P., & Cowan, N. (2021). A preregistered replication and extension of the cocktail party phenomenon: One's name captures attention, unexpected words do not. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 47*(2), 234–242. https://doi.org/10.1037/xlm0000874

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

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

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

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

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., Bruchmann, M., Steinweg, A.-L., Moeck, R., & Straube, T. (2020). Attentional conditions differentially affect early, intermediate and late neural responses to fearful and neutral faces. *Social Cognitive and Affective Neuroscience, 15*(7), 765–774. https://doi.org/10.1093/scan/nsaa098

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*(8), 1311–1324. https://doi.org/10.1177/0956797621996667

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 and Behavioral Neuroscience.* https://doi.org/10.3758/s13415-022-00998-x

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., 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

Shin, Y. S., Masís-Obando, R., Keshavarzian, N., Dáve, R., & Norman, K. A. (2020). Context-dependent memory effects in two immersive virtual reality environments: On mars and underwater. *Psychonomic Bulletin & Review, 28*(2), 574–582. https://doi.org/10.3758/s13423-020-01835-3

Shinners, P. (2011). *PyGame*. Retrieved April 24th 2018 from http://pygame.org/

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

Stahl, J., Wiese, H., & Schweinberger, S. R. (2008). Expertise and own-race bias in face processing: An event-related potential study.

NeuroReport, 19(5), 583–587. https://doi.org/10.1097/wnr.0b013e3282f97b4d

Stasinopoulos, D. M., & Rigby, R. A. (2007). Generalized Additive Models for Location Scale and Shape (GAMLSS) in R. *Journal of Statistical Software*, *23*(7), 1–46. https://doi.org/10.18637/jss.v023.i07

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

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

Straube, T., Mothes-Lasch, M., & Miltner, W. H. R. (2011). Neural mechanisms of the automatic processing of emotional information from faces and voices. *British Journal of Psychology, 102*(4), 830–848. https://doi.org/10.1111/j.2044-8295.2011.02056.x

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 Kesteren, M. T. R., Rijpkema, M., Ruiter, D. J., & Fernandez, G. (2010). Retrieval of associative information congruent with prior knowledge is related to increased medial prefrontal activity and connectivity. *Journal of Neuroscience, 30*(47), 15888–15894. https://doi.org/10.1523/jneurosci.2674-10.2010

Valdés-Conroy, B., Aguado, L., Fernández-Cahill, M., Romero-Ferreiro, V., & Diéguez-Risco, T. (2014). Following the time course of face gender and expression processing: A task-dependent ERP study. *International Journal of Psychophysiology, 92*(2), 59–66. https://doi.org/10.1016/j.ijpsycho.2014.02.005

Ventura-Bort, C., Löw, A., Wendt, J., Dolcos, F., Hamm, A. O., & Weymar, M. (2016). When neutral turns significant: Brain dynamics of rapidly formed associations between neutral stimuli and emotional contexts. *European Journal of Neuroscience, 44*(5), 2176–2183. https://doi.org/10.1111/ejn.13319

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

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

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., 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

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, 07*, 1258. https://doi.org/10.3389/fpsyg.2016.01258